Brain Tumor Classification from MRI Radiomic and Image Features

Ahmad Alashoury
Department of Computer Science
Western University
London, Ontario
aalashou@uwo.ca

Christopher Betancur
Department of Computer Science
Western University
London, Ontario
cbetancu@uwo.ca

Otilia Pasculescu
Department of Computer Science
Western University
London, Ontario
opascule@uwo.ca

Karim Abousamra
Department of Computer Science
Western University
London, Ontario
kabousam@uwo.ca

***Abstract*—Early and accurate detection of brain tumors is essential for improving patient outcomes. Yet, the traditionally used methods of identifying brain tumors can be spotty. In this study, we examine how different types of features, derived from MRI scans of the brain, contribute to automated tumor classification. Using the Kaggle Brain Tumor dataset containing 3,762 MRI slices, we compare logistic regression and RBF-kernel models trained on first-order features, second-order features, image-only inputs, and a combination of these features. Our findings show that combining tabular radiomic features with image data produces the strongest overall performance. It also shows that the logistic regression model remained competitive with SVM in several feature-rich settings.**

***Index Terms*—brain tumor classification, MRI, logistic regression, support vector machine, SVM, medical imaging.**

I. Introduction

There are many life-threatening forms of cancer in the world but brain cancer is considered to be one of the most severe. According to the Global Cancer Observatory, approximately 320,000 people died from brain and central nervous system cancer in 2022 alone [8]. In Canada, it is estimated that around 3300 people will be diagnosed with brain cancer, and of which 78% will die as a result of that cancer in 2025 [9].

To improve the patient's chance of living, early detection can be the difference between a full recovery and death. Research indicates that earlier diagnosis and intervention significantly improve survival, with one study noting that a diagnosis within two weeks of symptom appearance and surgical intervention within three weeks is associated with prolonging the patient's survival [10].

Typically, brain tumors are found by radiologists who examine MRI scans for abnormalities. Studies also show that radiologist overburden is a major barrier to timely diagnosis, with nearly 29.7% of delayed MRI and CT reports caused specifically by radiologists having too many scans to interpret [11]. Another thing to consider is that tumors are not all the same sizes, contrasts, textures and etc. They tend to be unique, making it harder to identify. While some tumors can be easy to identify, others can be challenging to detect.

Challenges like this have led to a rapidly growing interest in utilizing machine learning to help improve tumor detection and improve the patient's chance of living. A well-designed model can be trained to learn subtle patterns not obvious to humans, which will improve early detection, increase consistency, and reduce the workload placed on radiologists, which may in turn lower the risk of human error. Despite substantial work in this area, most studies rely exclusively on image-only inputs or deep learning pipelines. Our work is different as it directly compares first-order and second-order features alongside tabular features and tests how well they can predict tumors, while also checking whether adding the images improves performance. It offers a simpler approach than most existing studies.

Due to these issues, there has been a growing interest in finding ways to help with tumor detection, including using machine learning. A well-designed model may be able to pick up patterns in MRI imaging that are not immediately obvious to the human eye. This will help improve early detection, increase consistency, and reduce the workload placed on radiologists. Despite substantial work in this area, most studies rely exclusively on image-only inputs or deep learning pipelines. Our work has a different approach: we directly compare different tabular feature sets and image-plus-tabular combinations to see how each dataset configuration affects brain-tumor classification performance the most. It offers a simpler approach than most existing studies.

In this paper, we explore how different types of features can be used to predict whether an MRI scan contains a tumor. We will do this by comparing logistic regression and SVM (Support Vector Machine) models trained on the first-order features, second-order features, and the image data. The goal is to determine which combination provides the best representation and provides the most reliable performance for automated brain classification.

II. Background & Related Work

For supervised classification studies for Brain MRI slices, it is common to use publicly defined datasets, and in this paper, we use the Kaggle Brain tumor dataset. This dataset provides 2D MRI slices with corresponding labels and precomputed first- and second-order radiomic features [1].

Most of the recent literature on the topic of brain tumor classification from MRI scans utilises deep convolutional neural networks (CNNs) applied directly to 2D slices. Abiwinanda et al. train a CNN on grayscale brain MRIs that are resized to small spatial resolutions (e.g., 64×64), which shows that even smaller representations of the image produce good performance [2]. Ait Amou et al. propose a deep-learning MRI diagnosis pipeline for brain tumors and explicitly adopt a similar resizing strategy, citing Abiwinanda et al. as motivation [3]. Ait Amou et al. introduce a CNN-based MRI diagnosis method

for brain tumor classification, where it too uses resized images to feed the CNN, citing Abiwinanda et al. as motivation [3]. These papers demonstrate that resized MRI slices can be used to achieve a strong performance while reducing computational cost, but note that they focus only on image-only inputs and do not examine any extra precomputed features based on the MRI scan.

Along with resizing, there are other preprocessing techniques and methods that are used in MRI-based studies. Shinohara et al. argue that intensity across many MRI scanners and acquisition protocols may vary in intensity and are never consistent, and due to that, normalization is an important step to reduce variation before modeling [4]. This supports our use of converting the images to grayscale, resizing, and intensity normalization before feeding the compressed image into the model, alongside standardization of tabular features.

Cortes and Vapnik introduced SVMs as margin-based classifiers that can learn non-linear decision boundaries via kernel functions such as the radial basis function (RBF) kernel [5].

To apply our models, both SVMs use scikit-learn library, which provides standard implementations of classifiers, feature scaling, and cross-validation procedures [6].

First and second order features are crucial for improving diagnostic, prognostic, and predictive accuracies [12].

Since imaging is routinely medically practiced, adding radiomic features could result in much better classifications, indicating that radiomic features are useful for detecting the presence of tumors [13].

III. Methods

A. Research Objectives

In this paper, we test the following hypothesis:

> Across the feature subsets we consider, models that combine image-based MRI features with tabular features achieve higher performance in classifying tumor vs non-tumor MRI scans than models that use a single feature type (tabular-only or image-only), and for any given feature subset, an RBF-kernel support vector machine (SVM) performs better than the logistic regression model.

To investigate this hypothesis, we define the following objectives:

O1: Construct seven datasets from the Kaggle Brain Tumor data [1] representing different feature subsets: tabular-only (all 13 features, first-order only, second-order only), image-only features, image + first-order features, image + second-order features, and image + all tabular features.

O2: Train and tune logistic regression and RBF-kernel SVM classifiers on each subset using scikit-learn pipelines with standardization, splitting the data into training and test sets and optimizing hyperparameters via cross-validation with grid search.

O3: Compare classification performance across feature subsets and models using held-out test accuracy and macro-averaged F1-score, and analyze which feature subset and which model (logistic vs SVM) performs best.

B. Research Methodology

1) Dataset and Preprocessing: We use the Brain Tumor dataset from Kaggle [1], which provides a CSV file with the precomputed radiomic features and corresponding MRI slice filenames that those features belong to, along with the raw brain MRI images themselves.

The raw CSV contains $N = 3762$ samples. We check for and drop any rows with features taking infinite values, but in this dataset, no rows are removed. All feature vectors in the dataset have one to one correspondence to a 2D MRI slice, along with a binary label `Class` where label 1 means that the MRI slice contains a tumor and 0 means the MRI slice does not contain a tumor. For each row in the CSV, it includes an `Image` identifier (to link the row of features to the image), the class (tumor or no tumor), and 13 tabular features where the features themselves are divided into two types:

a) First-order features.:

- Mean
- Variance
- Standard deviation
- Skewness
- Kurtosis

b) Second-order (texture) features.:

- Contrast
- Energy
- ASM (angular second moment)
- Entropy
- Homogeneity
- Dissimilarity
- Correlation
- Coarseness

So we now construct seven aligned datasets from the same N samples:

1) **D1: First-order only (tabular)**
   Utilize the 5 first-order intensity statistics: Mean, Variance, Standard Deviation, Skewness, Kurtosis.
2) **D2: Second-order only (tabular)**
   Utilize the 8 second-order texture features: Contrast, Energy, ASM, Entropy, Homogeneity, Dissimilarity, Correlation, Coarseness.
3) **D3: All tabular features**
   Utilize all 13 tabular features (first-order + second-order).
4) **D4: Image only**
   Utilize only the grayscale MRI slice, resized to 64×64 and normalized.
5) **D5: Image + first-order**
   Concatenate image pixels with the 5 first-order features.
6) **D6: Image + second-order**
   Concatenate image pixels with the 8 second-order features.
7) **D7: Image + all tabular features**
   Concatenate image pixels with all 13 tabular features.

Each MRI slice is loaded in grayscale, resized to a resolution of 64×64, and then normalized to the range $[0, 1]$. By

reducing the size of the original MRI slice from 240×240 to 64×64 we reduce the dimensionality (the number of features we pass into the model) while still retaining the overall structural information needed for tumor classification, following prior work that successfully used small resized MRIs for CNN-based brain tumor classification [2], [3]. In MRI analysis, it is common to normalize the intensities so that MRI machine settings have comparable brightness and contrast before we fit any models [4].

Let $\mathbf{x}_{\text{img}} \in \mathbb{R}^{4096}$ be the flattened representation (of length 64×64) of the compressed and normalized grayscale MRI slice in pixels and $\mathbf{x}_{\text{tab}} \in \mathbb{R}^{13}$ represent the full tabular feature vector for a given MRI slice.

For each dataset variant defined above, we construct a single feature vector $\mathbf{x}$ from these components are as follows:

- tabular-only variants (D1–D3): $\mathbf{x} \subseteq \mathbf{x}_{\text{tab}}$ where $\mathbf{x}$ can contain all tabular features or a subset.
- image-only variant (D4): $\mathbf{x} = \mathbf{x}_{\text{img}}$.
- Combined variants (D5–D7):

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{\text{img}} \\ \mathbf{x}' \end{bmatrix},$$

  where $\mathbf{x}'$ as a subvector of $\mathbf{x}_{\text{tab}}$ containing either the 5 first-order features, the 8 second-order features, or all 13 tabular features.

The goal is to learn a function (model) that, given a feature vector $\mathbf{x}$, outputs a binary label (a standard supervised binary classification task):

$$f : \mathcal{X} \to \{0, 1\},$$

where $\mathcal{X}$ denotes the space of all feature vectors constructed as above.

2) Logistic Regression Model: Logistic regression is essentially a modified form of linear regression that is designed for binary outcomes. Instead of predicting raw continuous values, it models the log-odds of the positive class as a linear function of the inputs. In our case, it models the log-odds of tumor presence and then passes that value through the sigmoid function (a sideways S-shaped curve) to convert it into a probability between 0 and 1. [7]

3) Support Vector Machine Model: SVMs are margin-based classifiers that find the decision boundary which maximizes the distance between classes [5]. This model can represent non-linear decision boundaries by implicitly mapping the input features into a higher-dimensional space through the kernel function. In this paper, we utilize an SVM model with a radial basis function (RBF) kernel [5] as our main model, implemented using the machine learning library scikit-learn [6].

For each of the seven datasets D1–D7 we have defined above, each with a unique feature configuration, we then train a separate RBF SVM model on the corresponding feature vectors $\mathbf{x}$. To avoid features with large values dominating the kernel, we standardize all input features to zero mean and unit variance using a `StandardScaler` inside a scikit-learn `Pipeline`.

The RBF SVM has two key hyperparameters:

1) Regularization parameter C which controls the trade-off between having a wider margin with more training errors versus fitting the training data more strictly with fewer margin violations.
2) Kernel width parameter γ controls the influence of a single training point.

4) Training Procedure and Experimental Design: For each dataset variant D1–D7, we follow the same training and evaluation protocol for both logistic regression and SVM models. We split each dataset into 80% training and 20% testing using `train_test_split(..., test_size=0.2, stratify=y, random_state=42)` to preserve the tumor / non-tumor class proportions and ensure reproducibility.

To stabilise optimisation and make features comparable in scale, we implement each classifier as a scikit-learn pipeline [6]. For logistic regression, the pipeline uses a median imputer, a `StandardScaler`, and a `LogisticRegression` model with ℓ_1 or ℓ_2 regularisation. For the SVM, the pipeline uses a `StandardScaler` and an RBF-kernel `SVC` classifier. In both models, the scaler operates on the full feature vector $\mathbf{x}$ (tabular features and, when present, flattened image pixels), so that each feature dimension has approximately zero mean and unit variance.

Hyperparameters are tuned on the training set using stratified k-fold cross-validation with $k = 3$, shuffling, and a fixed `random_state`. For logistic regression, the grid search explores regularisation strengths C, penalties (ℓ_1 vs. ℓ_2), and compatible solvers (`liblinear`, `lbfgs`, `saga`). For the SVM, the grid search varies the regularisation parameter C and the RBF kernel width γ over a small logarithmic grid, with the kernel type fixed to RBF.

In both models, we select the hyperparameter combination that maximises the mean cross-validation accuracy on the training data, retrain the best model on the full training set, and evaluate it once on the held-out test set. For every model (logistic vs. SVM) and dataset D1–D7, we report test accuracy and macro-averaged F1-score to compare performance across feature subsets and classifiers.

5) Evaluation Metrics: For a given model (logistic regression or RBF SVM), we train and tune hyperparameters and then evaluate performance using three quantities:

1) Mean cross-validated accuracy on the training split (used only for hyperparameter selection),
2) Overall accuracy on the held-out test set,
3) Macro-averaged F1-score on the held-out test set.

To select hyperparameters, we use mean cross-validated accuracy: for each hyperparameter configuration, we run k-fold cross-validation on the training data and compute the average accuracy over the k train/validation splits. This provides a more stable estimate of generalization than a single train/validation split and helps reduce overfitting of hyperparameters.

After choosing the best hyperparameters, we retrain the model on the full training split and evaluate it once on the held-out test set. On this test set, we report overall accuracy and macro-averaged F1-score to compare how well each model performs across the different datasets.

We evaluate test performance by comparing the predicted label $\hat{y}$ with the true label y for every MRI slice. We treat the tumor class as positive ($y = 1$) and the no-tumor class as negative ($y = 0$), and define:

- TP (true positives): slices where $y = 1$ and $\hat{y} = 1$ (tumor correctly predicted),
- TN (true negatives): slices where $y = 0$ and $\hat{y} = 0$ (no tumor correctly predicted),
- FP (false positives): slices where $y = 0$ but $\hat{y} = 1$ (model predicts a tumor where there is none),
- FN (false negatives): slices where $y = 1$ but $\hat{y} = 0$ (model misses a true tumor).

The overall *accuracy* is the fraction of correct predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

For each class $c \in \{\text{tumor}, \text{no-tumor}\}$ we compute precision, recall, and F1-score. For the positive (tumor) class these are

$$\text{Precision}_+ = \frac{\text{TP}}{\text{TP} + \text{FP}}, \qquad \text{Recall}_+ = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{F1}_+ = 2 \cdot \frac{\text{Precision}_+ \cdot \text{Recall}_+}{\text{Precision}_+ + \text{Recall}_+}.$$

Analogous quantities Precision_-, Recall_-, and F1_- are defined for the no-tumor class. The macro-averaged F1-score reported in our experiments corresponds to scikit-learn's `f1_score(y_test, y_pred, average="macro")`, which simply computes

$$\text{F1}_{\text{macro}} = \tfrac{1}{2}\big(\text{F1}_+ + \text{F1}_-\big),$$

the unweighted mean of the two class-wise F1-scores.

6) Threats to Validity and Rationale: Our study has several limitations that affect how broadly the results can be interpreted. First, all experiments use only a single Kaggle dataset [1] with precomputed radiomic features and 2D slices. How that data was computed and what was used are not fully documented, so performance estimations in this paper may not be able to generalize to other scanners, institutions, or clinical populations.

Second, we work at the slice level and do not have patient identifiers. If multiple MRI slices come from the same patient and appear in both the training and test splits, this may inflate reported accuracy and F1-scores because the model effectively sees very similar images during training and testing. Since, there is no provided documentation to verify whether a given MRI slice has a tumor or not, we simply trust that the provided labels are true.

Third, our models are restricted to logistic regression and RBF-kernel SVMs with relatively small hyperparameter grids. More complex architectures (e.g., CNNs) or richer hyperparameter searches could change the absolute performance and possibly the relative ranking of feature configurations. In addition, resizing MRIs to 64×64 and flattening them into vectors may discard fine-grained spatial information.

Despite these limitations, our design is appropriate for O1–O3 because it lets us compare the various feature subsets we have defined above and compare logistic regression versus SVM models. We are able to apply an 80/20 stratified split with cross-validated model selection, and two classifiers provide a clear test of our hypothesis, although the resulting metrics should not be viewed as definitive clinical performance.

IV. Results

A. System and Design Requirements

For all code implementation, see the attached zip file. This file includes all information needed to run the experiment and acquire the results. The experimental pipeline was designed with the following requirements:

Consistency across datasets: All models must use the same aligned sample ordering to allow direct dataset-to-dataset comparisons. The implementation meets this requirement by constructing all feature matrices (D1–D7) from a shared cleaned DataFrame and applying a single stratified train/test split.

Scalability: Image preprocessing is performed once and cached. Subsequent runs reuse the stored feature matrix, ensuring reproducible and efficient execution across experiments.

Model robustness: Logistic regression must remain numerically stable even in high-dimensional settings (e.g., D7 with 4109 features). Our implementation using the `lbfgs` and `liblinear` solvers satisfies this requirement, with all models converging without numerical instability.

B. Implementation, Testing, and Verification

During implementation, several aspects of the system were verified:

Image–tabular alignment: Each image feature vector was correctly matched with its corresponding radiomic features. Consistency checks using sampled rows confirmed correct alignment across all 3762 samples.

Feature normalization: Standardization was applied jointly over image and tabular dimensions. Logistic regression performed substantially worse without scaling, confirming the necessity of normalization for stable optimization.

Cross-validation behavior: Grid search using stratified 3-fold CV produced consistent scores across folds, indicating reliable model behavior and low sensitivity to sampling variation.

C. Logistic Regression Results Analysis

Logistic regression was trained and tuned independently on each dataset variant (D1–D7), representing different combinations of radiomic features and image-based features. Across all experiments, D7 (Image + All Tabular Features) achieved the best performance, with a test accuracy of 0.993 and a macro

F1-score of 0.993. Conversely, the weakest performance was observed in D1 (First-order only) and D4 (Image-only), both of which lack higher-level texture information that appears essential for robust tumor classification.

Fig. 1 summarizes the performance across D1–D7.

Fig. 1. Logistic regression test accuracy and macro F1-score across feature subsets D1–D7. Performance improves as additional radiomic or image features are incorporated, with the combined image + tabular feature set (D7) performing best.

D. SVM Results Analysis

SVM was trained and tested on each dataset variant (D1–D7), with each dataset representing different combinations of radiomic features and image-based features. The best test accuracy result comes from the D3 data, with a score of 0.980, and also a very high CV score as well (0.986). The best F1 score is also from the D3 data, with a score of 0.991. Across all experiments, D3 achieved the best overall performance. The weakest performance was observed in D4 (image only) with a test accuracy of 0.896, along with a low CV score and a low F1 score. However, the lowest F1 score was reported in D1 with a score of 0.885.

Fig. 2 summarizes the performance across D1–D7.

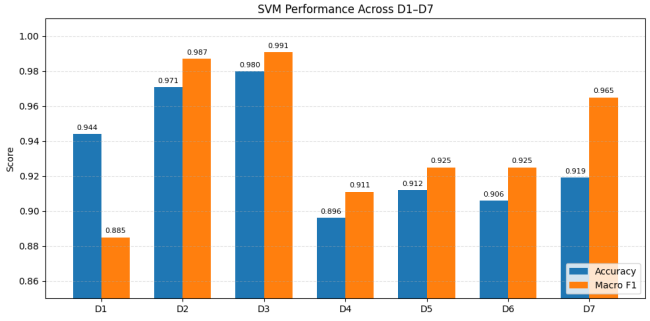

Fig. 2. SVM test accuracy and macro F1-score across feature subsets D1–D7. Performance is best when focusing solely on tabular data (D3), and decreases once images are introduced.

E. Tabulated Results

Tables I and II list the cross-validated accuracy, test accuracy, and macro F1-score obtained for each dataset using logistic regression and SVM, respectively.

TABLE I
LOGISTIC REGRESSION PERFORMANCE ACROSS D1–D7

Dataset	CV Accuracy	Test Accuracy	Macro F1
D1 – First-order only	0.885	0.885	0.883
D2 – Second-order only	0.977	0.984	0.984
D3 – All tabular features	0.981	0.991	0.991
D4 – Image only (64×64)	0.883	0.888	0.887
D5 – Image + first-order	0.955	0.975	0.975
D6 – Image + second-order	0.976	0.985	0.985
D7 – Image + all tabular	0.979	0.993	0.993

TABLE II
SVM PERFORMANCE ACROSS D1–D7

Dataset	CV Accuracy	Test Accuracy	Macro F1
D1 – First-order only	0.946	0.944	0.885
D2 – Second-order only	0.982	0.971	0.987
D3 – All tabular features	0.986	0.980	0.991
D4 – Image only (64×64)	0.839	0.896	0.911
D5 – Image + first-order	0.845	0.912	0.925
D6 – Image + second-order	0.847	0.906	0.925
D7 – Image + all tabular	0.850	0.919	0.965

F. Result Interpretations - Logistic Regression

After examining the results from our logistic regression model findings, these are the key observations we noticed and the conclusions we drew:

First-order features alone are limited. From the results, we can see that first-order features have a low accuracy, and lack higher-level information.

Second-order texture features substantially improve performance. Texture descriptors such as contrast, entropy, and homogeneity encode spatial irregularities that are highly characteristic of tumor presence [12].

Image-only inputs perform poorly with linear models. Because logistic regression is a linear classifier, it cannot efficiently model complex spatial patterns present in raw pixel data when the image is flattened.

Combining image and tabular features yields the strongest results. The combination provides both high-level radiomic summaries and fine-grained pixel cues, leading to the highest-performing model for logistic regression [13].

G. Result Interpretations - SVM

After examining the results from our SVM model findings, these are the key observations we noticed and the conclusions we drew:

Tabular features are strong: All tabular datasets outperform image datasets in test accuracy, which could indicate that SVM performs better with tabular data than it does with image data with this specific model and dataset. D3 (all tabular features) is also the best performer, which strengthens the indication that tabular features outperform non-tabular features in this evaluation [13].

Image data performs poorly: D4 (image only) is observed as the weakest performer in test accuracy, which indicates that images alone are not strong. When tabular data is added, performance rises, which shows us that tabular data strengthens the model.

First-order features are not enough: D1 is also observed as a weak performer and it only takes into account first-order features. This is an indication that first-order features are not enough to detect brain tumors and the data itself may not descriptive enough for the model to be accurate in this assessment. We believe that this leads to a smaller accuracy. When adding first-order and second-order features, the accuracy improved [12].

H. Comparison of Logistic Regression and SVM

Table III directly compares logistic regression and RBF SVM on each dataset D1–D7 using test accuracy and macro F1-score.

Logistic regression generally achieves higher test accuracy on the tabular-only and fused image+tabular datasets (D2, D3, D5–D7), while SVM is competitive or slightly better on some macro F1-scores (e.g., D3 and D7) and on the image-only configuration (D4). In particular, the best overall test accuracy is obtained by logistic regression on D7 (image + all tabular features), whereas the best SVM configuration is D3 (all tabular features). This means that our original hypothesis that an RBF SVM would consistently outperform logistic regression on every feature subset is *not* supported by the empirical results: for many radiomic-rich settings, a well-regularized linear model is at least as strong as the non-linear SVM.

TABLE III
COMPARISON OF LOGISTIC REGRESSION AND SVM ACROSS D1–D7

Dataset	LR Acc	LR Macro F1	SVM Acc	SVM Macro F1
D1 – First-order only	0.885	0.883	0.944	0.885
D2 – Second-order only	0.984	0.984	0.971	0.987
D3 – All tabular features	0.991	0.991	0.980	0.991
D4 – Image only (64×64)	0.888	0.887	0.896	0.911
D5 – Image + first-order	0.975	0.975	0.912	0.925
D6 – Image + second-order	0.985	0.985	0.906	0.925
D7 – Image + all tabular	0.993	0.993	0.919	0.965

I. Novelty of Our Approach

These are the main aspects of novelty that characterize our methodology:

1) **A unified evaluation of seven radiomic–image feature subsets.** While prior work often focuses solely on image-based deep learning models, our experiments demonstrate that radiomic features alone can achieve near-perfect performance with both linear regression and SVM implementation.
2) **Fusion of high-dimensional pixel data with engineered features.** Few studies evaluate logistic regression on a combined 4096-pixel + 13-feature input space. Our results show that with proper standardization and regularization, logistic regression can remain stable and highly accurate even in this setting.
3) **Certain SVMs work better with tabular only data.** Our findings show that the strongest SVM model uses tabular, radiomic features only. This challenges the assumption that adding features improves performance.
4) **A well-regularized linear model can be competitive with a non-linear SVM.** Our results suggest that a linear model can be superior to and outperform a non linear model.

V. CONCLUSIONS & FUTURE WORK

This project set out to investigate three objectives (O1–O3): (1) construct seven aligned datasets combining radiomic and image features (D1–D7), (2) train and tune logistic regression and RBF SVM models on each subset using a common pipeline, and (3) compare performance across feature configurations and models using held-out test accuracy and macro F1-score.

With respect to O1, we successfully built seven consistent feature configurations from the Kaggle Brain Tumor dataset, ranging from simple first-order radiomic features (D1) to a fused representation of 64×64 image pixels and all 13 tabular features (D7). For O2, both logistic regression and SVM were implemented in standardized scikit-learn pipelines with stratified train/test splits and 3-fold cross-validated grid search. The resulting models converged reliably on all datasets, including the high-dimensional fused representations.

Regarding O3 and our hypothesis, the results show that combining image and tabular features does indeed produce the strongest models overall: for both classifiers, the best-performing configurations involve either all tabular features (D3) or image+tabular fusion (D5–D7). However, the second part of our hypothesis, that an RBF SVM would uniformly outperform logistic regression on each feature subset, is not supported. Logistic regression matches or exceeds SVM in test accuracy on most tabular-rich datasets, including the best overall configuration D7, while SVM is only clearly superior for the image-only case and some macro F1-scores. A key conclusion is that carefully regularized linear models remain highly competitive for radiomic-based brain tumor classification, especially when informative texture features are available.

Several avenues for future work arise from this study. First, we worked at the slice level without patient identifiers; a natural extension is to reconstruct patient-level splits to avoid potential information leakage and to evaluate per-patient rather than per-slice performance. Second, our models are limited to logistic regression and RBF SVMs; exploring modern deep learning architectures (e.g., CNNs for the images plus MLP heads for tabular features) could reveal whether more complex models still gain significant accuracy beyond the radiomic baselines. Third, our hyperparameter grids were deliberately small; more extensive searches, calibration analysis, and class-imbalance handling (e.g., focal losses or cost-sensitive learning) could further refine model performance. Finally, explainability methods such as feature importance analysis or SHAP values could be applied to better understand which radiomic descriptors drive classification decisions.

From an implementation perspective, this project also produced several practical lessons. We found that strict alignment between image files and CSV rows is critical; even

small indexing mistakes can silently corrupt supervised labels. Normalization and standardization are indispensable for both logistic regression and SVM, particularly when fusing pixel values and radiomic features in a single high-dimensional vector. Caching preprocessed image features made repeated experimentation feasible and highlights the importance of engineering the pipeline, not just tuning the model. More broadly, the work reinforced that simple, well-validated baselines are essential: they provide strong performance, expose bugs early, and offer a clear reference point for any future, more complex models to beat.

VI. Acknowledgements

We used Grammarly for grammar suggestions and spelling corrections.

References

[1] J. Bohaju, "Brain Tumor," Kaggle, Jul. 2020, ver. 3. [Online]. Available: https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor. doi: 10.34740/KAGGLE/DSV/955413. [Accessed: Dec. 6, 2025].

[2] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, "Brain tumor classification using convolutional neural network," in *World Congress on Medical Physics and Biomedical Engineering 2018*. Singapore: Springer, 2019, pp. 183–189. doi: 10.1007/978-981-10-9035-6_33.

[3] M. Ait Amou, H. Xia, A. Larhmam, and M. Alami El Filali, "A novel MRI diagnosis method for brain tumor classification based on deep learning," *Healthcare*, vol. 10, no. 3, p. 494, 2022. doi: 10.3390/healthcare10030494.

[4] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage: Clinical*, vol. 6, pp. 9–19, 2014. doi: 10.1016/j.nicl.2014.08.008.

[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. doi: 10.1007/BF00994018.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://www.jmlr.org/papers/v12/pedregosa11a.html.

[7] A. DeMaris and S. H. Selman, "Logistic regression," in *Converting Data into Evidence: A Statistics Primer for the Medical Practitioner*. New York, NY: Springer, 2013, pp. 115–136. doi: 10.1007/978-1-4614-7792-1_7.

[8] Global Cancer Observatory, "Cancer Today." [Online]. Available: https://gco.iarc.fr/today/en/dataviz/pie?mode=population&group_populations=0. [Accessed: Dec. 7, 2025].

[9] Canadian Cancer Society / Société canadienne du cancer, "Brain and other nervous system cancer statistics," *Canadian Cancer Society*. [Online]. Available: https://cancer.ca/en/cancer-information/cancer-types/brain-and-spinal-cord/statistics. [Accessed: Dec. 7, 2025].

[10] D. Kawauchi *et al.*, "Early diagnosis and surgical intervention within 3 weeks from symptom onset are associated with prolonged survival of patients with glioblastoma," *Neurosurgery*, vol. 91, no. 5, pp. 741–748, Nov. 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9531976/. [Accessed: Dec. 8, 2025].

[11] G. Wahid, A. Haroon, M. Samad, and N. Tamkeen, "Causes of delay in radiological reporting and ways to reduce them," *J. Saidu Med. Coll.* [Online], vol. 12, no. 3, pp. 133–137, Sep. 26, 2022. Available: https://jsmc.pk/index.php/jsmc/article/view/697. [Accessed: Dec. 8, 2025].

[12] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016. doi: 10.1148/radiol.2015151169.

[13] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, Art. no. 4006, 2014. doi: 10.1038/ncomms5006.