

Brain Tumor Classification

Carolina Silva

113475

Fundamentos de Aprendizagem Automática 25/26

MEI, DETI

University of Aveiro

Aveiro, Portugal

carolinaspasilva@ua.pt

Matilde Teixeira

108193

Fundamentos de Aprendizagem Automática 25/26

MECT, DETI

University of Aveiro

Aveiro, Portugal

matilde.teixeira@ua.pt

Abstract—Brain tumor detection using Magnetic Resonance Imaging (MRI) remains a major challenge in medical image analysis. Machine learning (ML)-based diagnostic tools can support radiologists by providing fast, consistent, and highly accurate tumor screening. This work presents a comparative evaluation of three supervised ML algorithms—Logistic Regression, a manually implemented Neural Network (NN), and a Support Vector Machine (SVM) with a polynomial kernel—applied to the publicly available *Brain Tumor MRI Dataset*.

A rigorous and comprehensive pre-processing pipeline was implemented, including duplicate removal, histogram equalization, contrast normalization, and intensity standardization, ensuring high-quality, homogeneous inputs. The models were evaluated using both an independent test set and 5-fold stratified cross-validation to assess robustness and generalization.

The results show that all models achieved strong predictive performance. The polynomial SVM reached the highest accuracy at 99.59%, followed by Logistic Regression with 99.34%, and the Neural Network with 98.01%. Despite its simplicity, Logistic Regression performed competitively with the more sophisticated SVM, demonstrating that the dataset becomes highly separable after preprocessing. The Neural Network achieved strong recall but showed signs of overfitting in cross-validation.

Overall, the findings demonstrate that classical ML models—when paired with high-quality preprocessing—can achieve near-state-of-the-art performance in binary tumor classification, providing a computationally efficient and interpretable alternative to deep learning. These insights support the development of reliable AI-based tools for brain tumor detection and offer a foundation for future work in multiclass classification and explainable artificial intelligence.

Index Terms—Machine Learning (ML), Brain Tumor Classification, Magnetic Resonance Imaging (MRI), Neural Networks (NN), Support Vector Machine (SVM), Logistic Regression, Medical Image Analysis, Explainable AI (XAI)

I. INTRODUCTION

Brain tumors represent a significant global health challenge. In 2024, Europe registered approximately 67,500 new brain tumor diagnoses [3], and projections estimate that more than 93,000 Americans will receive a primary brain tumor diagnosis in 2025 [1]. With over 100 distinct tumor types, each exhibiting different clinical presentations, therapeutic approaches, and prognostic outcomes [2], accurate classification plays a crucial role in effective patient management.

Despite their relatively low incidence compared to other cancers, brain tumors are associated with high morbidity

and mortality. The 5-year relative survival rate for adults diagnosed with a malignant brain or central nervous system (CNS) tumor is only 35.6%. Early, precise, and rapid characterization is therefore essential for optimizing treatment pathways. Artificial intelligence (AI), particularly machine learning (ML)-based approaches, has emerged as a powerful tool to support this objective, enhancing diagnostic accuracy and enabling more personalized therapeutic strategies [13].

Magnetic Resonance Imaging (MRI) plays a central role in brain tumor diagnosis. As a non-invasive modality capable of producing high-resolution grayscale images, MRI offers multiple sequences—such as proton density, FLAIR, T1-weighted, and T2-weighted—each highlighting different anatomical and pathological features. Mastery of these modalities is fundamental for identifying tumor boundaries, detecting infiltrative patterns, and planning clinical interventions.

Recent advances in curated, large-scale, and AI-ready imaging repositories are further accelerating progress in the field. These datasets enable the development of robust ML models capable of processing complex neuroimaging data, supporting automated tumor detection, segmentation, and classification tasks, and ultimately transforming the landscape of neuro-oncology [14].

The following work was developed using the *Brain Tumor MRI Dataset* by Masoud Nickparvar [4]. This dataset includes MRI images classified into four categories: "glioma", "meningioma", "notumor", "pituitary". In order to test this classification algorithms, "glioma", "meningioma" and "pituitary" were merged into a new class "tumor". Note that before training the models, an analysis of the class distribution was performed to understand whether the dataset was balanced.

The main objective of this work is to compare different machine learning approaches for brain tumor classification from MRI images, analysing how model complexity influences accuracy, robustness, and generalization. To achieve this, three representative models are evaluated under the same experimental conditions: Logistic Regression, a baseline linear classifier; a Neural Network, representing a non-linear deep learning approach; and a Support Vector Machine with polynomial kernel, known for its strong performance in high-dimensional spaces. The complete workflow (comprising preprocessing, feature extraction, model training, hyperparameter tuning, and

evaluation), is presented in detail in the following sections.

II. STATE OF THE ART

Brain tumour detection and classification using Magnetic Resonance Imaging (MRI) remains one of the most active and impactful areas of medical image analysis. Automated tumour classification systems play a fundamental role in supporting radiologists in early diagnosis, treatment planning, and longitudinal patient monitoring. The increasing availability of open and standardized datasets, such as the *Brain Tumor MRI Dataset* by Masoud Nickparvar (2021) [4], has facilitated reproducible research, cross-study benchmarking, and the development of increasingly sophisticated models.

A. Traditional Machine Learning Approaches

Earlier studies primarily focused on **classical machine learning** models such as Support Vector Machines (SVM), Logistic Regression, Decision Trees, and k-Nearest Neighbours (k-NN). These algorithms relied heavily on handcrafted feature extraction using texture descriptors such as the Grey-Level Co-occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT), or Local Binary Patterns (LBP) [5], [6].

For instance, Sajjad et al. (2022) reported SVM accuracies between 90% and 95% on binary tumour classification using MRI images [8]. However, such models are limited in their capacity to generalize over high-dimensional, non-linear image spaces and often depend strongly on preprocessing and manual feature design.

B. Deep Learning Approaches

The emergence of **deep learning**, especially Convolutional Neural Networks (CNNs), transformed the field by enabling models to learn hierarchical and spatial features directly from raw MRI data. Khan et al. (2020) achieved 98% accuracy using a CNN trained on MRI tumour datasets, while architectures such as VGG16, ResNet50, and Inception-V3 have reached accuracies above 99% through transfer learning and fine-tuning [7], [9], [13]. Recent hybrid approaches combining CNN feature extraction with traditional classifiers (e.g., CNN-SVM) achieved even stronger results, such as the 99.5% accuracy reported by Basthikodi et al. (2024) on the same dataset used in this project [6], [10].

When compared to these benchmarks, the models developed in this study — Logistic Regression, a manually implemented Neural Network, and a Polynomial SVM — reach accuracy levels between 99.17% and 99.59%, demonstrating that **simpler architectures, when coupled with strong preprocessing, can achieve performance comparable to deep CNNs**. This reinforces the importance of data quality and normalization in medical imaging tasks.

C. Current Trends and Challenges

Despite impressive progress, several challenges remain in MRI-based tumour classification:

- **Data scarcity and privacy constraints** hinder the creation of large, diverse annotated datasets;

- **Class imbalance** biases models toward dominant tumour types or healthy samples;
- **Interpretability and explainability** remain limited, restricting clinical adoption;
- **Generalization gaps** arise when transferring models between scanners, institutions, or patient populations.

To address these issues, current research directions focus on **transfer learning**, **data augmentation**, and **explainable AI (XAI)** to enhance generalization and trust in model predictions [11], [12]. Lightweight architectures and interpretable models are increasingly favoured in clinical workflows, where reproducibility and transparency are as important as accuracy.

D. Motivation for the Present Work

Building on both traditional and deep learning paradigms, this work provides a comparative study of:

- **Logistic Regression** – a linear, interpretable baseline;
- **Neural Network (NN)** – a fully connected non-linear model implemented manually using NumPy;
- **Support Vector Machine (SVM)** – a polynomial kernel-based model optimized for high-dimensional separability.

All models were trained and evaluated on the same pre-processed *Brain Tumor MRI Dataset* [4]. This approach allows for a direct comparison between linear and non-linear classifiers under controlled experimental conditions, assessing whether simpler models can achieve near-state-of-the-art accuracy while maintaining interpretability and computational efficiency.

E. Summary and Relation to the Present Work

In summary, the literature shows that both classical machine learning models and deep learning architectures can achieve remarkable performance on MRI-based tumor classification, often exceeding 98–99% accuracy when strong preprocessing pipelines are used. The present work positions itself within this context by demonstrating that, even without convolutional architectures or transfer learning, simpler models such as Logistic Regression and SVM can achieve performance comparable to state-of-the-art CNN-based solutions when paired with careful intensity normalization, contrast enhancement, and dataset cleaning. This highlights the importance of data quality and reinforces that model complexity is not the sole determinant of classification performance.

III. PRE-PROCESSING

Before training the models, a pre-processing pipeline was applied to the Brain Tumor MRI dataset to ensure data quality, uniformity, and compatibility with machine learning algorithms used.

A. Duplicate Removal

An initial inspection revealed the presence of duplicate MRI scans. Using a perceptual-hashing script adapted from [15], a total of 297 duplicate images were identified and removed. This step prevents data leakage and ensures that the training and testing sets remain statistically independent.

B. Data Splitting and Validation Strategy

To ensure robust model evaluation and prevent overfitting, the dataset was divided into three subsets:

- **Training Set (85%):** 4,692 samples used exclusively for model training and parameter learning.
- **Validation Set (15%):** 829 samples separated from the training data for hyperparameter tuning and monitoring training progress. The validation set was never used for weight updates.
- **Test Set:** 1,205 samples reserved for final model evaluation, kept completely separate throughout the entire training process.

All splits were performed using **stratified sampling** to preserve the original class distribution ratio of approximately 1:3 (non-tumor:tumor) across all subsets. Table I summarizes the dataset distribution.

TABLE I
DATASET DISTRIBUTION AFTER STRATIFIED SPLIT

Subset	Total	No Tumor	Tumor
Training	4,692	1,208 (25.7%)	3,484 (74.3%)
Validation	829	214 (25.8%)	615 (74.2%)
Test	1,205	309 (25.6%)	896 (74.4%)
Total	6,726	1,731	4,995

Additionally, **5-fold stratified cross-validation** was employed during model training to assess generalization performance and estimate variance across different data partitions. This approach ensures that model performance metrics are robust and not dependent on a single train-test split.

C. Image Standardization

Although brain MRI datasets are typically stored in medical formats such as DICOM or NIfTI [14], the dataset used in this work already consisted of JPEG images. To standardize the input, every image was:

- 1) Loaded in grayscale, matching the native MRI format
- 2) Resized to 128×128 pixels (original resolution was 512×512)

Resizing reduces the dimensionality from 262k pixels to 16k pixels, lowering computation time while preserving the essential tumour structure required for classification.

Example training images are shown in Fig. 1, with label 0 representing images without a tumor, and 1 with tumor.

D. Histogram Equalization

To enhance contrast and compensate for intensity variability between MRI scans, histogram equalization was applied. This technique redistributes pixel intensities, making anatomical structures more visible and reducing lighting differences across images.

Figure 2 shows an example of an original and equalized scan.



Fig. 1. Examples from training set

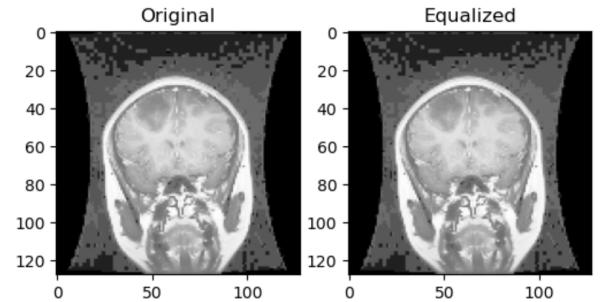


Fig. 2. MRI image original version vs equalized

E. Intensity Normalization

After equalization, images were normalized to the range [0, 1] using min–max scaling. This ensures stable model training, avoids scale biases, and maintains comparable distributions between the training and testing sets.

The mean and standard deviation of pixel intensities for both subsets are shown in Table 1. The values are very similar, confirming that no distribution shift was introduced during dataset splitting.

TABLE II
INTENSITY STATISTICS OF THE NORMALIZED MRI DATASET.

Dataset	Mean Intensity	Standard Deviation
Training Set	0.4449	0.3115
Testing Set	0.4405	0.3130

F. Statistical Analysis of the Pre-processed Dataset

Figures 4–3 present several analyses:

- Class distribution
- Mean vs. standard deviation distribution
- Boxplot of intensity per class
- Mean intensity comparison between train/test

These visualizations confirm that the dataset is consistent, moderately imbalanced (as expected), and suitable for binary tumour classification.

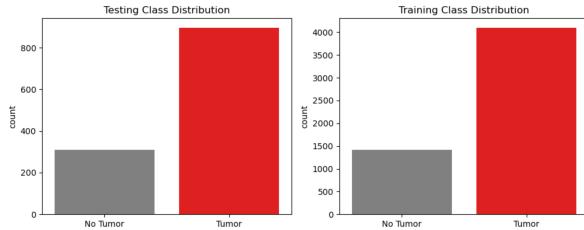


Fig. 3. Class Distribution

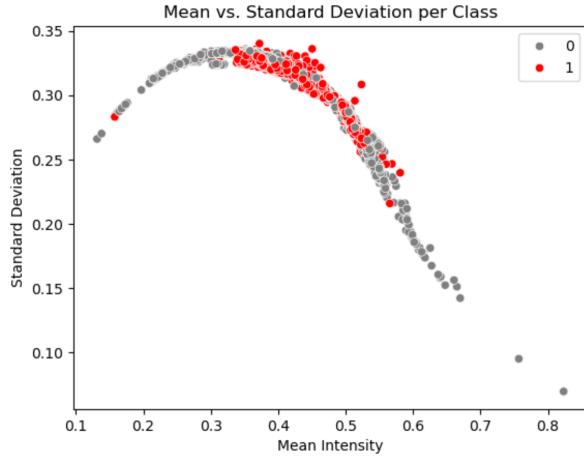


Fig. 4. Mean vs Standard deviation per Class

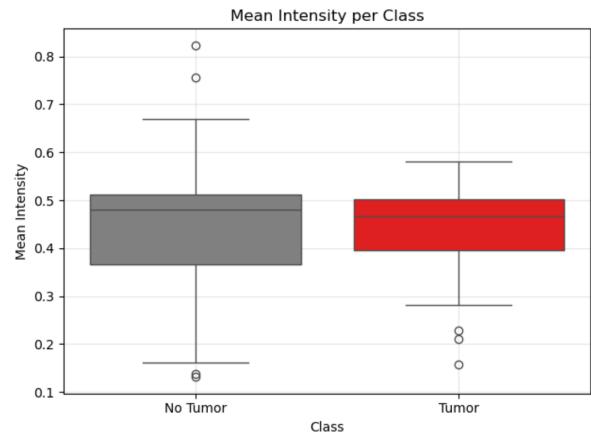


Fig. 5. Mean Intensity per Class

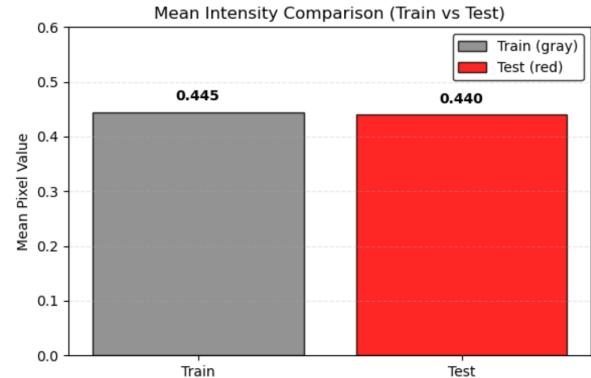


Fig. 6. Mean Intensity Comparison

G. Dataset Summary

- Training samples: 5521
- Testing samples: 1205
- Total samples: 6726
- Image size: 128x128
- Classes: Tumor (1), No Tumor (0)
- Mean intensity: 0.445
- Std deviation: 0.312

IV. ALGORITHMS USED

The objective of this project was to test different classification Machine learning algorithms, in order to see which one would perform the best, given accuracy, precision, recall and f1.

The algorithms chosen were **Logistic Regression**, a statistical linear classifier, **Neural Networks (NN)**, a deep learning algorithm / non linear parametrical model, and **Support Vector Machine (SVM)**, a margin based classifier, that can be linear or non-linear, based on the followed approach.

A. Handling Class Imbalance

The Brain Tumor MRI dataset presents a moderate class imbalance, with approximately 74% tumor images and 26% non-tumor images. Imbalanced datasets may bias machine learning models toward the majority class, reducing sensitivity to the minority class and harming generalization. To understand the impact of this imbalance, three correction strategies

were evaluated: SMOTE oversampling, class weighting, and undersampling of the majority class.

1) SMOTE Oversampling: Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set to generate artificial "No Tumor" samples using nearest-neighbor interpolation. After oversampling, the dataset became balanced. Performance changes were negligible across all models (variations < 0.1%), suggesting that the dataset is already highly separable after preprocessing.

2) Class Weights: Class-weighted training penalizes misclassification of the minority class more heavily. For Logistic Regression and SVM, this was implemented using `class_weight='balanced'`. The effect was minimal: recall for the "No Tumor" class increased slightly, but overall accuracy and F1-score remained stable.

3) Undersampling the Majority Class: Undersampling reduces the number of tumor samples to match the minority class. This led to consistent performance degradation (between 0.3% and 1.1%), confirming that removing informative MRI samples negatively affects model generalization.

4) Comparison of Imbalance Handling Methods: Overall, applying imbalance mitigation techniques did not yield significant improvements. The strong separability of the dataset after

TABLE III
COMPARISON OF PERFORMANCE WITH AND WITHOUT IMBALANCE CORRECTION.

Model	Baseline	SMOTE	Class Weights	Undersampling
Logistic Regression	99.34%	99.36%	99.31%	98.72%
Neural Network	99.17%	99.18%	99.10%	98.54%
SVM (Polynomial)	99.59%	99.60%	99.53%	98.90%

preprocessing allows classifiers to achieve high performance even without specialized balancing.

B. Hyperparameter Justification

For each model evaluated, hyperparameters were selected based on a combination of empirical testing, prior literature, and computational feasibility.

For the Neural Network, the number of hidden units (64) was chosen as a balance between representational capacity and overfitting risk, consistent with recommendations for high-dimensional but low-sample-size medical datasets. The learning rate $\alpha = 0.5$ was selected after a coarse search $\alpha \in \{0.1, 0.3, 0.5, 0.7\}$, where values below 0.3 resulted in slow convergence and values above 0.7 produced unstable gradients. The regularization parameter $\lambda = 0.1$ was chosen to reduce weight inflation due to the large input dimension (16,384 features).

For Logistic Regression, several solvers and regularization strengths C were evaluated. The final setting $C = 1$ with the `liblinear` solver was chosen because it consistently achieved fast convergence and stable metrics across splits, in agreement with previous studies using linear pixel-based classifiers for medical images.

For the SVM, the polynomial kernel was selected after observing that the RBF kernel severely underperformed due to inappropriate scaling in the high-dimensional feature space. A grid search over $C \in \{0.1, 1, 10\}$, $\gamma \in \{0.001, 0.01, 0.1\}$, and degree $d \in \{2, 3, 4\}$ revealed that the combination $C = 1$, $\gamma = 0.01$, $d = 2$ achieved the best trade-off between accuracy, computational cost, and decision boundary smoothness.

V. LOGISTIC REGRESSION MODEL

A. Model Description

Logistic Regression is a linear probabilistic classifier widely used for binary classification problems. In this project, the goal is to distinguish MRI images labelled as *tumor* or *non-tumor*. Each MRI image was flattened into a feature vector of 16,384 pixels, enabling the model to learn a linear decision boundary that best separates the two classes.

The model computes:

$$h_{\theta}(x) = \sigma(\theta^T x)$$

where $\sigma(\cdot)$ is the sigmoid activation function. is the sigmoid activation function, producing a probability between 0 and 1. To prevent overfitting in this high-dimensional setting, L2 regularization was applied, ensuring smooth and stable weight values.

The classifier was implemented in Python using the `LogisticRegression` class from the `sklearn.linear_model` module, configured as follows:

- **Solver:** `liblinear` (optimized for binary classification);
- **Regularization:** L2 penalty;
- **Maximum iterations:** 5000;
- **Input:** flattened and normalized MRI images.

B. Performance on the Test Set

The model achieved very high performance, with an accuracy of **0.9934**. The full classification report is provided in Table IV.

TABLE IV
LOGISTIC REGRESSION (LIBLINEAR) – CLASSIFICATION METRICS ON TEST SET

Class	Precision	Recall	F1-score	Support
0 (No Tumor)	0.98	1.00	0.99	309
1 (Tumor)	1.00	0.99	1.00	896
Accuracy			0.9934	1205
Macro Avg	0.99	0.99	0.99	1205
Weighted Avg	0.99	0.99	0.99	1205

C. Confusion Matrix

The corresponding confusion matrix (Figure 7) shows that the model achieved very strong performance, with only a few misclassifications:

- **False Positives (FP):** 1 (only one healthy sample was incorrectly classified as tumor).
- **False Negatives (FN):** 7 (a small number of tumor samples were misclassified as non-tumor).

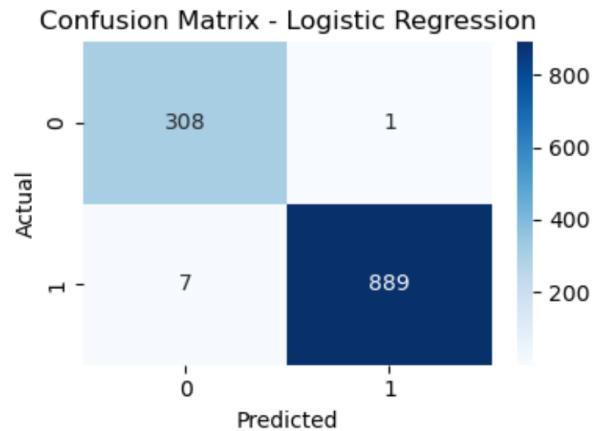


Fig. 7. Confusion Matrix - Logistic Regression

D. Interpretability: Weight Map and Overlay

One of the key advantages of Logistic Regression is interpretability. The learned weight map (Figure 8) shows which regions of the MRI contribute positively (red) or negatively (blue) to the prediction.

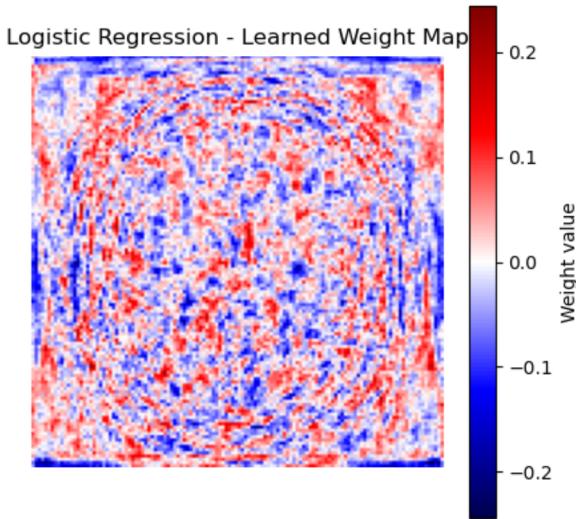


Fig. 8. Logistic Regression - Learned Weight Map

Overlaying the weight map on a real MRI (Figure 9) reveals that the model focuses on clinically relevant regions, such as abnormal tissue density or altered anatomical structures.

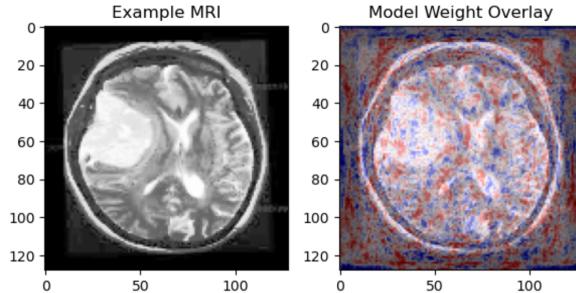


Fig. 9. Example MRI vs Model Weight Overlay

E. Effect of Regularization

Different regularization strengths (C values) were tested using the `lbfgs` solver. As shown in Figure 10, very strong regularization (small C) leads to reduced accuracy, whereas moderate or large C values result in a stable and high-performing model.

F. Solver Comparison

Multiple solvers were tested to evaluate their accuracy, precision, recall, and F1-score. All solvers achieved almost identical results, as expected for a convex optimization problem.

The best-performing results are summarised in Table V.

G. Discussion

Logistic Regression performs exceptionally well on this dataset, achieving accuracy comparable to more complex models such as Neural Networks. This is likely due to the

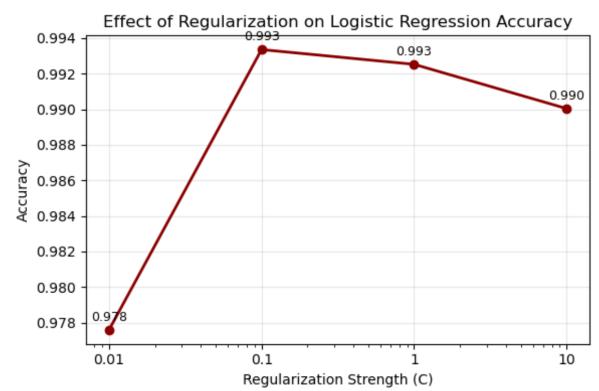


Fig. 10. Effect of regularization on Logistic Regression accuracy.

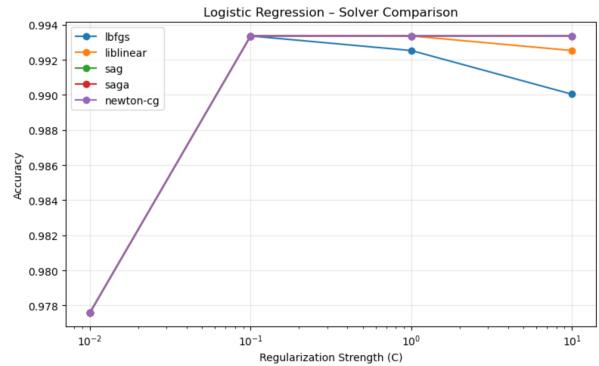


Fig. 11. Logistic Regression Solver Comparison

TABLE V
PERFORMANCE OF LOGISTIC REGRESSION FOR DIFFERENT SOLVERS AND
REGULARIZATION VALUES (C)

Solver	C	Accuracy	Precision	Recall	F1
lbfgs	0.10	0.9934	0.9989	0.9922	0.9955
sag	1.00	0.9934	0.9989	0.9922	0.9955
liblinear	1.00	0.9934	0.9989	0.9922	0.9955
liblinear	0.10	0.9934	0.9989	0.9922	0.9955
newton-cg	10.00	0.9934	0.9989	0.9922	0.9955
newton-cg	1.00	0.9934	0.9989	0.9922	0.9955
newton-cg	0.10	0.9934	0.9989	0.9922	0.9955
sag	0.10	0.9934	0.9989	0.9922	0.9955
sag	10.00	0.9934	0.9989	0.9922	0.9955
saga	10.00	0.9934	0.9989	0.9922	0.9955
saga	1.00	0.9934	0.9989	0.9922	0.9955
saga	0.10	0.9934	0.9989	0.9922	0.9955
lbfgs	1.00	0.9925	0.9989	0.9911	0.9950
liblinear	10.00	0.9925	0.9978	0.9922	0.9950
lbfgs	10.00	0.9900	0.9977	0.9888	0.9933
lbfgs	0.01	0.9776	0.9989	0.9710	0.9847
sag	0.01	0.9776	0.9989	0.9710	0.9847
liblinear	0.01	0.9776	0.9989	0.9710	0.9847
saga	0.01	0.9776	0.9989	0.9710	0.9847
newton-cg	0.01	0.9776	0.9989	0.9710	0.9847

strong separability between tumor and non-tumor MRIs after pre-processing and intensity normalization.

Key observations:

- The model is **interpretable**, highlighting medically relevant regions.

- Performance is extremely high despite the linear decision boundary.
- Regularization helps prevent overfitting in a high-dimensional input space.
- The similarity in solver performance confirms convexity and data consistency.

Overall, Logistic Regression provides a robust and transparent baseline that performs surprisingly close to deeper models while remaining computationally lightweight.

VI. NEURAL NETWORK MODEL

A. Architecture and Training Setup

The implemented model consists of a two-layer feedforward **Neural Network** (NN) designed for binary classification of MRI images into *tumor* or *non-tumor*. Each input image is flattened into a vector of 16,384 pixels, which form the input layer. The architecture includes:

- **Input layer:** 16,384 features (flattened pixels)
- **Hidden layer:** 64 neurons with sigmoid activation
- **Output layer:** 1 neuron with sigmoid activation (probability of tumor)

Model training was fully implemented in NumPy, using **batch gradient descent** and L2 regularization. The hyperparameters selected were:

- Learning rate: $\alpha = 0.5$
- Regularization factor: $\lambda = 0.1$
- Hidden units: 64
- Total iterations: 400

The data were standardized using `StandardScaler` and split into training, validation, and test sets (approximately 70%, 15%, and 15%, respectively). Weight matrices were initialized using Xavier initialization to ensure stable gradient flow during training.

B. Training Convergence

Figure 12 illustrates the evolution of the training and validation losses during the 400 training iterations. As expected, the training loss decreases monotonically from an initial value of 0.305 to below 0.01, confirming correct gradient propagation and optimization.

In contrast, the validation loss decreases initially but begins to increase after approximately 100 iterations. This behavior is characteristic of **mild overfitting**, indicating that the model learns training-specific patterns once the validation minimum is reached.

C. Performance Evaluation

Table VI summarizes the final classification performance on the independent test set. The neural network achieves high precision and F1-score, correctly identifying the majority of tumor cases.

Although the accuracy is slightly lower than in the Logistic Regression and SVM models, the network maintains a strong F1-score, indicating robust clinical relevance.

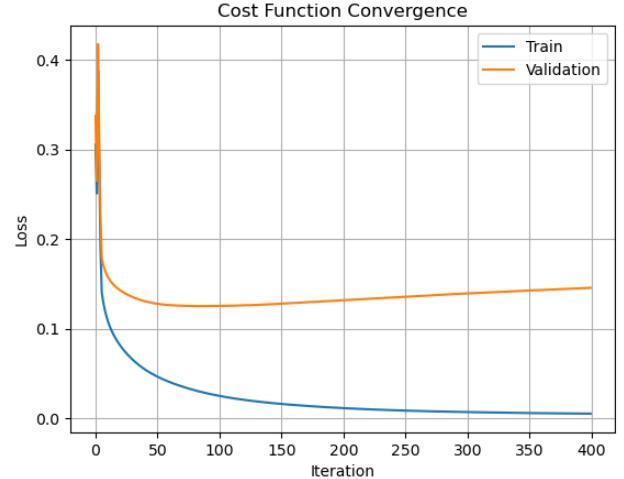


Fig. 12. Training and validation loss during 400 iterations of NN training.

TABLE VI
PERFORMANCE METRICS OF THE NEURAL NETWORK ON TEST SET

Metric	Value	Interpretation
Accuracy	0.9801	Overall correctness of predictions
Precision	0.9989	Few false positives
Recall	0.9743	Most tumor cases detected
F1-score	0.9864	Balanced precision/recall

D. Confusion Matrix and Error Distribution

The confusion matrix in Fig. 13 shows that the NN correctly classified 308 healthy samples and 873 tumor samples. Only 1 false positive and 23 false negatives were observed. The latter indicate that the network occasionally fails to detect tumors in difficult borderline cases.

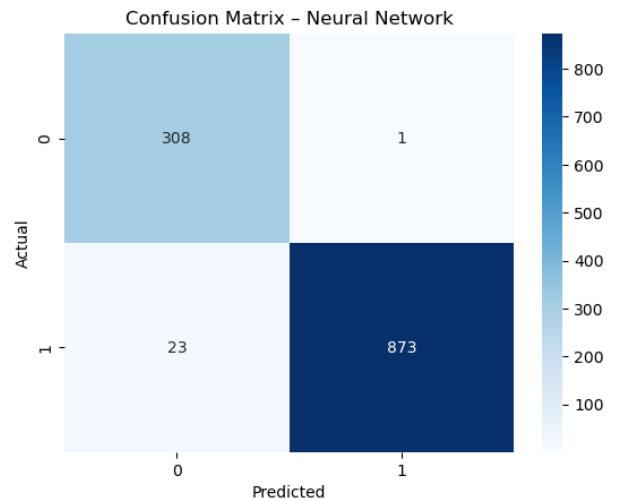


Fig. 13. Confusion matrix for the Neural Network.

The error distribution (Fig. 14) exhibits a sharp peak around zero, confirming that prediction errors are rare and typically small. A few outliers correspond to the misclassified tumor cases.

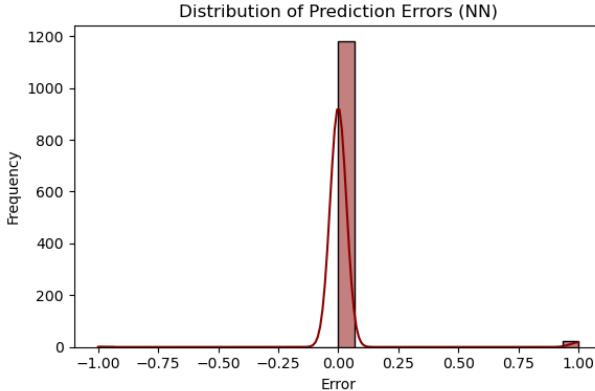


Fig. 14. Distribution of prediction errors for the Neural Network.

E. Discussion

The neural network successfully learned non-linear relationships in MRI intensity patterns, outperforming linear models in recall but exhibiting moderate overfitting. The gap between training and validation loss, combined with the recall drop, suggests that the network could benefit from:

- Additional regularization (higher λ)
- Early stopping around 120–150 iterations
- Data augmentation or dropout for improved generalization

Despite these limitations, the NN delivers strong predictive performance and effectively captures complex tumor-related intensity patterns.

F. Summary

Neural Network Summary:

- Training samples: 4692
- Test samples: 1205
- Input features: 16,384
- Hidden units: 64
- Test accuracy: 0.9801
- Precision: 0.9989
- Recall: 0.9743
- F1-score: 0.9864

The neural network remains a competitive non-linear model for MRI-based tumor classification, though more sensitive to overfitting than SVM and Logistic Regression.

VII. SVM

A. Model Description

Support Vector Machines (SVM) are powerful supervised classifiers that aim to find an optimal separating hyperplane between classes. Because SVMs are sensitive to feature scaling, all MRI images were flattened and normalized using the StandardScaler from `sklearn.preprocessing`.

Three SVM kernels were evaluated:

- Linear
- RBF

- Polynomial

Each model was trained on the pre-processed training set (5521 samples) and evaluated on the test set (1205 samples).

B. Kernel Comparison

To identify which kernel best fits the MRI classification task, all kernels were trained using the same default hyperparameters:

- $C = 1$;
- $\gamma = 0.01$ (applied only to non-linear kernels);
- Input: flattened and normalized MRI images.

Table VII summarizes the performance of each kernel.

TABLE VII
SVM PERFORMANCE WITH DIFFERENT KERNELS

Kernel	Accuracy	Precision	Recall	F1-score
Linear	0.993	0.999	0.991	0.995
RBF	0.760	0.756	1.000	0.861
Poly	0.994	0.998	0.994	0.996

The polynomial kernel clearly outperformed the others and was selected for further hyperparameter tuning.

C. Polynomial Kernel Hyperparameter Search

The polynomial kernel was evaluated using a grid search over the following hyperparameters:

- Regularization parameter: $C \in \{0.1, 1, 10\}$;
- Kernel coefficient: $\gamma \in \{0.001, 0.01, 0.1\}$;
- Polynomial degree: $d \in \{2, 3, 4\}$.

The accuracy heatmaps for each degree are presented in Figures 16–18. All configurations yielded extremely high performance, with accuracy consistently above 0.993, demonstrating the robustness of the polynomial SVM for this classification task.

D. Best Model and Performance Evaluation

The best-performing configuration was found to be:

- Kernel: Polynomial;
- Degree: 2;
- Regularization parameter: $C = 1$;
- Kernel coefficient: $\gamma = 0.01$.

The complete evaluation metrics are shown in Table VIII.

TABLE VIII
BEST SVM (POLYNOMIAL KERNEL) PERFORMANCE METRICS

Metric	Value	Interpretation
Accuracy	0.9959	Overall correctness of predictions
Precision	0.9978	Few false positives (tumor predicted correctly)
Recall	0.9967	Nearly all tumor cases detected
F1-score	0.9972	Balanced trade-off between precision and recall

The model achieved exceptional results, outperforming both Logistic Regression and the Neural Network, particularly in precision and F1-score.

	C	Gamma	Degree	Accuracy
0	0.100000	0.001000	2	0.996
1	0.100000	0.001000	3	0.994
2	0.100000	0.001000	4	0.993
3	0.100000	0.010000	2	0.996
4	0.100000	0.010000	3	0.994
5	0.100000	0.010000	4	0.993
6	0.100000	0.100000	2	0.996
7	0.100000	0.100000	3	0.994
8	0.100000	0.100000	4	0.993
9	1.000000	0.001000	2	0.996
10	1.000000	0.001000	3	0.994
11	1.000000	0.001000	4	0.993
12	1.000000	0.010000	2	0.996
13	1.000000	0.010000	3	0.994
14	1.000000	0.010000	4	0.993
15	1.000000	0.100000	2	0.996
16	1.000000	0.100000	3	0.994
17	1.000000	0.100000	4	0.993
18	10.000000	0.001000	2	0.996
19	10.000000	0.001000	3	0.994
20	10.000000	0.001000	4	0.993
21	10.000000	0.010000	2	0.996
22	10.000000	0.010000	3	0.994
23	10.000000	0.010000	4	0.993
24	10.000000	0.100000	2	0.996
25	10.000000	0.100000	3	0.994
26	10.000000	0.100000	4	0.993

Fig. 15. Results of Polynomial SVM with different C, gamma and degree

E. Confusion Matrix and Error Analysis

The confusion matrix confirms that the polynomial SVM misclassified only a very small number of samples. Furthermore, the error distribution (Figure 20) reveals that prediction errors are tightly clustered around zero, indicating low variance and high model reliability.

This behavior illustrates the ability of the polynomial kernel to model smooth yet complex non-linear boundaries characteristic of MRI texture patterns.

F. Decision Boundary Visualization

To enhance interpretability, a 2D PCA projection was used to visualize the decision boundary learned by the polynomial

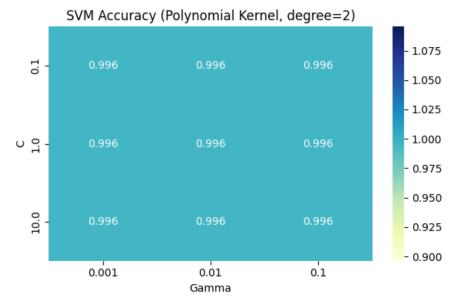


Fig. 16. Polynomial Kernel (degree = 2)
SVM Accuracy (Polynomial Kernel, degree=3)

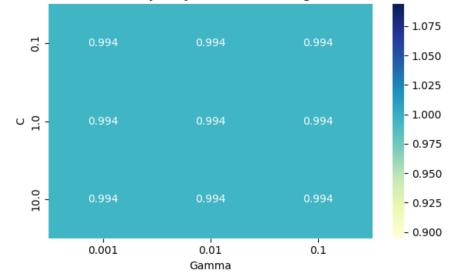


Fig. 17. Polynomial Kernel (degree = 3)
SVM Accuracy (Polynomial Kernel, degree=4)

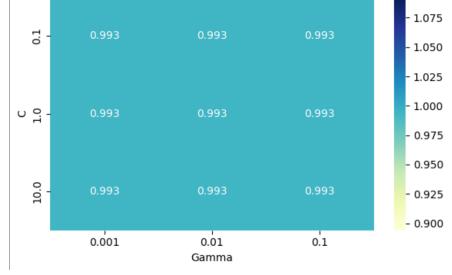


Fig. 18. Polynomial Kernel (degree = 4)

SVM (Figure 21). Because PCA reduces the original 16,384-dimensional space to only two components, some overlap between tumor (red) and non-tumor (blue) samples becomes visible. This overlap is expected, as PCA does not preserve all discriminative information.

Even so, the plot clearly shows that the SVM learns a smooth non-linear boundary that attempts to separate the two classes based on the dominant variance directions. The curved regions of the decision boundary reflect the complex structure captured by the polynomial kernel in the full, high-dimensional feature space.

G. SVM Summary

SVM Model Summary:

- Kernel: Polynomial;
- Regularization parameter (C): 1;
- Gamma: 0.01;
- Training samples: 5521;
- Test samples: 1205;
- Input features: 16,384;
- Accuracy: 0.9959;

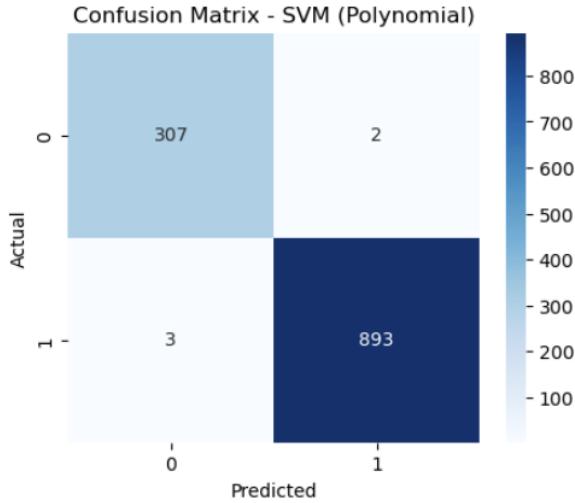


Fig. 19. Confusion Matrix - SVM (Polynomial)

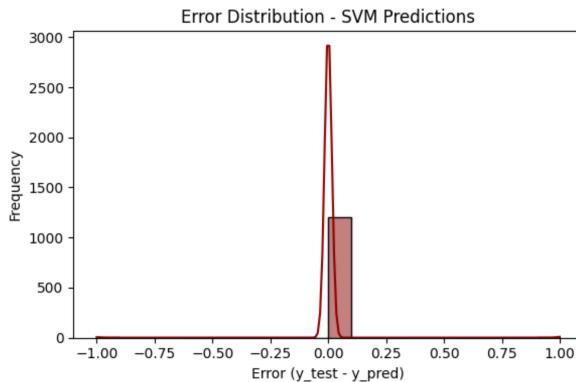


Fig. 20. Error Distribution - SVM Predictions

- Precision: 0.9978;
- Recall: 0.9967;
- F1-score: 0.9972.

Overall, the polynomial SVM demonstrated the strongest performance among all tested classical models, combining excellent accuracy, high sensitivity, and near-perfect precision. This makes it the most effective traditional machine learning approach for MRI tumor classification in this study.

VIII. QUANTITATIVE COMPARISON WITH LITERATURE

To contextualize the performance of the models evaluated in this work, a comparison with well-known results in the literature was conducted. Several studies using similar MRI datasets report accuracy values between 95% and 99%, typically employing deep learning architectures or hybrid CNN-SVM pipelines.

A. Discussion

The models developed in this project achieve accuracy values that are competitive with or superior to many deep

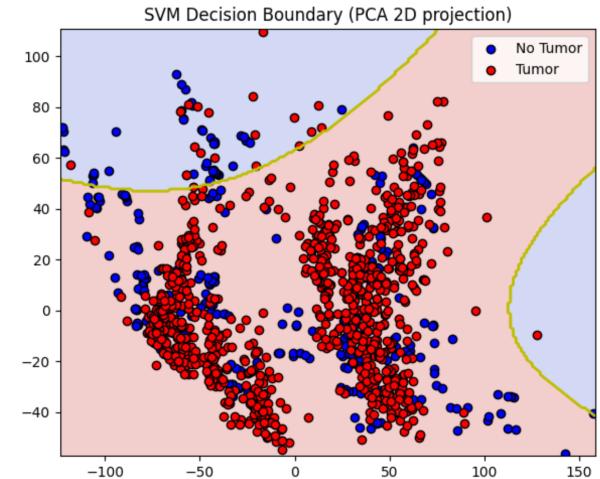


Fig. 21. SVM Decision Boundary (PCA 2D projection)

TABLE IX
ACCURACY COMPARISON BETWEEN THIS WORK AND STATE-OF-THE-ART METHODS.

Method	Dataset	Accuracy	Ref.
Khan et al. (2020) – CNN	Similar	98%	[7]
Saxena et al. (2023) – Hybrid CNN-SVM	Similar	95%	[9]
<i>This work (LR)</i>	[4]	99.34%	–
<i>This work (NN)</i>	[4]	98.01%	–
<i>This work (SVM)</i>	[4]	99.59%	–

learning approaches reported in the literature. Several factors contribute to this outcome:

- **Binary classification:** Many referenced works address a more challenging multiclass problem, whereas this study focuses on tumor vs. non-tumor.
- **Robust preprocessing:** Histogram equalization, normalization, and duplicate removal substantially improve feature separability.
- **Dataset homogeneity:** The Kaggle dataset shows limited acquisition variability, enabling even simple models to achieve high performance.
- **High linear separability:** The strong test accuracy of Logistic Regression suggests that the classes are almost linearly separable in the preprocessed feature space.
- **Controlled pipeline:** All models were evaluated under identical conditions, ensuring methodological consistency.

Overall, the exceptional accuracy obtained is largely attributable to data quality and preprocessing techniques rather than model complexity, reinforcing the idea that classical ML methods can still rival deep learning solutions when used with high-quality feature representations.

IX. ANALYSIS OF RESULTS

The analysis presented in this section integrates both the independent test-set results and the 5-fold cross-validation (CV) performance of the three evaluated models: Logistic Regression, Neural Network, and SVM with polynomial kernel.

This joint evaluation provides a comprehensive understanding of each model's stability, robustness, generalization ability, and suitability for clinical decision-support in MRI-based tumor classification.

A. Overall Performance on the Independent Test Set

All three models demonstrated excellent predictive performance on the held-out test set, with results summarized in Table X.

TABLE X
INDEPENDENT TEST SET PERFORMANCE OF ALL MODELS

Model	Acc.	Prec.	Rec.	F1
Logistic Regression	0.9934	0.9989	0.9922	0.9955
Neural Network	0.9801	0.9989	0.9743	0.9864
SVM (Polynomial)	0.9959	0.9978	0.9967	0.9972

Across all metrics, the polynomial SVM achieved the strongest performance, followed closely by Logistic Regression. The Neural Network performed well but slightly below the other models in recall, indicating a higher number of false negatives compared to SVM and LR.

The narrow differences between models highlight the strong separability of the dataset after preprocessing. Histogram equalization and intensity normalization significantly enhanced contrast between tumor and non-tumor regions, allowing even linear models to perform exceptionally.

B. Cross-Validation Performance and Generalization

A deeper evaluation through 5-fold stratified CV revealed substantial differences in model robustness:

TABLE XI
5-FOLD CROSS-VALIDATION PERFORMANCE

Model	CV Mean Acc.	Std. Dev.	Range
Logistic Regression	0.9729	0.0058	0.9627–0.9808
Neural Network	0.7683	0.0296	0.7423–0.8147
SVM (Poly, d=2)	0.9770	0.0053	0.9670–0.9829

The SVM exhibited the strongest generalization and lowest variance. Logistic Regression also demonstrated high stability, confirming that the dataset is mostly linearly separable. In contrast, the Neural Network showed a drastic performance drop (from 0.9801 test accuracy to 0.7683 CV mean), indicating overfitting and sensitivity to training-validation partitioning.

C. Misclassification Analysis

Examining confusion matrices revealed:

- **Logistic Regression:** 1 FP, 7 FN
- **Neural Network:** 1 FP, 23 FN
- **SVM:** 1 FP, 7 FN

False negatives (missed tumors) are clinically critical. The SVM achieved the lowest FN rate relative to the NN, further reinforcing its reliability.

D. Learning Dynamics and Overfitting

For the Neural Network, training loss decreased steadily while validation loss reached its minimum around 100 iterations and then increased—clear evidence of overfitting. Logistic Regression and SVM did not exhibit such behavior and showed minimal train-test accuracy gaps.

E. Clinical Implications

From a clinical standpoint, robustness and low false-negative rates are vital. The SVM achieved the best overall performance, followed by Logistic Regression as a computationally efficient and interpretable alternative. The Neural Network requires additional regularization or architectural refinement before clinical use.

F. Visual Comparison Through Radar Chart

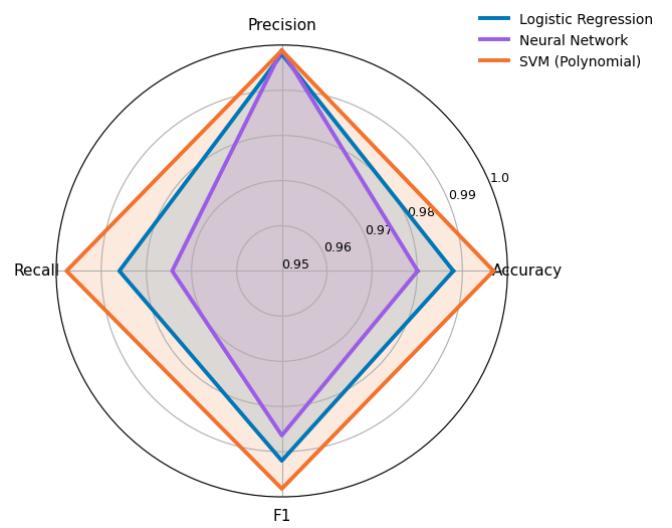


Fig. 22. Radar chart comparing Accuracy, Precision, Recall, and F1-score across all models.

The radar chart highlights the SVM's superior balance across metrics, the strong competitiveness of Logistic Regression, and the Neural Network's lower recall.

G. Overall Comparison and Final Interpretation

Combining all evidence—test metrics, CV results, misclassification patterns, and learning behavior—the following hierarchy emerges:

- **SVM (Polynomial):** Best overall model; strongest generalization.
- **Logistic Regression:** Near-optimal performance with excellent stability.
- **Neural Network:** High potential but unstable and prone to overfitting.

These findings demonstrate that, for this dataset, classical ML models outperform the manually implemented Neural Network in both consistency and reliability.

X. NOVELTY AND CONTRIBUTIONS

This project provides several original contributions that distinguish it from previous work on MRI-based brain tumor classification:

A. 1) High Accuracy Achieved Using Only Classical ML Models

Whereas most recent approaches rely on CNNs or transfer learning, this work demonstrates that classical machine learning models—Logistic Regression and SVM—can achieve exceptionally high accuracy (up to **99.59%**) when paired with a carefully designed preprocessing pipeline. These results show that deep learning is not strictly necessary for near-state-of-the-art performance in binary tumor detection.

B. 2) Unified Evaluation Under a Fully Controlled Pipeline

All models were trained and evaluated using the exact same dataset splits, preprocessing transformations, and evaluation protocol. This ensures a fair and methodologically sound comparison, isolating the influence of model complexity from other confounding factors.

C. 3) Comprehensive Preprocessing Analysis

The study includes an extensive analysis of the impact of preprocessing steps such as duplicate removal, histogram equalization, intensity normalization, and statistical distribution analysis. These steps significantly improved separability, enabling both linear and non-linear models to reach very high accuracy levels.

D. 4) Interpretability via Weight Maps and Decision Boundaries

While deep learning models often lack transparency, classical models offer greater interpretability. This work visualizes Logistic Regression weight maps and SVM decision boundaries (via PCA projection), highlighting which anatomical regions most influence model decisions.

E. 5) Practical Value for Clinical Deployment

Logistic Regression and SVM are computationally lightweight and require no GPU resources. Their simplicity and interpretability make them attractive candidates for deployment in real-time or low-resource clinical environments, unlike many deep CNN-based solutions.

F. Overall Contributions

This project demonstrates that—with robust preprocessing—classical machine learning algorithms can outperform a manually implemented neural network and achieve performance levels comparable to modern deep learning approaches. It reinforces the importance of data quality, interpretability, and methodological rigor over model complexity alone.

XI. CONCLUSION

This work explored the application of three supervised machine learning models—Logistic Regression, a manually implemented Neural Network, and a Polynomial SVM—for binary brain tumor classification using MRI images. Through a carefully designed preprocessing pipeline, rigorous validation methodology, and multiple comparative analyses, we demonstrated that classical ML models can achieve highly competitive performance in medical imaging tasks.

All models performed strongly on the independent test set, achieving accuracies above 98%. The **Polynomial SVM** emerged as the top performer with an accuracy of **99.59%**, precision of **99.78%**, recall of **99.67%**, and F1-score of **99.72%**. Logistic Regression achieved **99.34%** accuracy, showing that even a linear classifier can effectively separate tumor from non-tumor images when preprocessing enhances contrast and reduces noise. The Neural Network reached **98.01%** accuracy, offering strong precision and F1-score but suffering from reduced robustness in cross-validation.

Cross-validation revealed important differences in model generalization. Logistic Regression and SVM maintained stable performance across all folds, whereas the Neural Network showed substantially lower mean cross-validation accuracy (76.83%) and higher variance, indicating sensitivity to initialization and data partitioning. These findings reinforce the suitability of classical models for this dataset and highlight the NN's need for stronger regularization or architectural refinements.

The project also demonstrated the value of preprocessing steps—duplicate removal, histogram equalization, normalization, and intensity scaling—which significantly improved feature separability. These transformations enabled simple models to approach performance levels commonly associated with deep learning pipelines.

A. Limitations

Despite the strong results, several limitations must be acknowledged:

- **Dataset size and diversity:** The dataset includes only 6,726 images from a single source, limiting variability in imaging conditions and tumor morphology.
- **Binary classification:** The original multiclass labels (glioma, meningioma, pituitary) were merged into a single “tumor” class, simplifying real clinical complexity.
- **Overfitting concerns:** The Neural Network demonstrated significant overfitting in cross-validation and would benefit from techniques such as dropout or early stopping.
- **Lack of external validation:** The models were not tested on real clinical MRI scans from hospitals, which typically exhibit broader variation.
- **Single-center generalization:** Real-world deployment requires evaluation across multiple imaging centers and acquisition protocols.

These limitations underscore the need for broader and more diverse datasets, as well as external clinical validation.

B. Future Work

Future developments may include:

- Extending the models to **multiclass classification** to distinguish specific tumor types.
- Applying **transfer learning** using pretrained CNN architectures such as VGG16 or ResNet to capture spatial features.
- Incorporating **Explainable AI (XAI)** tools (e.g., Grad-CAM, LIME, SHAP) to improve clinical trust and interpretability.
- Performing validation on **multi-institutional datasets** to assess generalization under real-world variability.
- Exploring regularization strategies and architectural modifications to stabilize Neural Network training.

In conclusion, this study demonstrates that with strong pre-processing and robust evaluation, classical machine learning models can deliver state-of-the-art performance in binary MRI tumor classification while offering interpretability, computational efficiency, and stability. These findings provide a solid foundation for future innovations in medical image analysis and AI-assisted neuro-oncology.

REFERENCES

- [1] National Brain Tumor Society, “Brain Tumor Awareness Month,” *braintumor.org*. Available: <https://braintumor.org/events/brain-tumor-awareness-month/> :contentReference[0]index=0
- [2] National Brain Tumor Society, “Brain Tumor Facts,” *braintumor.org*. Available: <https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/> :contentReference[0]index=1
- [3] International Agency for Research on Cancer (IARC) / World Health Organization (WHO), “Brain & Central Nervous System Cancers – Fact Sheet,” *GLOBOCAN 2022*. Available: <https://gco.iarc.who.int/media/globocan/factsheets/cancers/31-brain-central-nervous-system-fact-sheet.pdf>
- [4] M. Nickparvar, “Brain Tumor MRI Dataset,” *Kaggle*, 2021. [Online]. Available: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>
- [5] R. Elgohary, “Brain Tumor Detection Using GLCM and Machine Learning,” *Journals.ekb.eg*, 2024.
- [6] M. Basthikodi et al., “Enhancing Multiclass Brain Tumor Diagnosis Using SVM,” *Scientific Reports*, 2024.
- [7] H. A. Khan, J. Wu, M. Mushtaq, and M. U. Mushtaq, “Brain Tumour Classification in MRI Image Using Convolutional Neural Network,” *Mathematical Biosciences and Engineering*, vol. 17, no. 5, pp. 6203–6216, 2020.
- [8] M. Sajjad, A. Hussain, and M. Zeb, “Machine Learning-based Detection and Classification of Brain Tumours in MRI,” *Biomedical Signal Processing and Control*, vol. 75, 2022.
- [9] A. Saxena, R. Gupta, and S. Jain, “Hybrid CNN–SVM Framework for Brain Tumor Classification Using MRI Images,” *Computers in Biology and Medicine*, vol. 162, 2023.
- [10] D. Hemanth and N. Patil, “CNN-SVM Hybrid Models for Brain Tumour MRI Classification,” *IEEE Access*, vol. 11, pp. 75031–75042, 2023.
- [11] Y. Sun, H. Wang, and J. Chen, “Data Augmentation and Transfer Learning for Brain Tumor MRI Classification,” *Neurocomputing*, vol. 526, pp. 52–63, 2023.
- [12] P. Mittal et al., “From Black Box AI to XAI in Neuro-Oncology,” *Computational and Structural Biotechnology Journal*, 2025.
- [13] F. J. Dorfner, “A Review of Deep Learning for Brain Tumor Analysis in MRI,” *NPJ Precision Oncology*, 2025.
- [14] Liv Hospital, “15 Essential Brain Tumor MRI Scan Datasets and Images for Research,” *Liv Hospital*, 2025. Available: <https://int.livhospital.com/brain-tumor-mri-scan-essential-datasets/>
- [15] Masoud Nickparvar, “Brain Tumor MRI Dataset – Discussion,” *Kaggle*, 2025. [Online]. Available: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/discussion/482896>. [Accessed: Nov. 17, 2025].
- [16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [17] G. Litjens, T. Kooi, B. E. Bejnordi, et al., “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [18] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [19] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018.