HELLENIC MEDITERRANEAN UNIVERSITY

ΠΑΝΤΕΣ ΑΝΘΡΩΠΟΙ ΤΟΥ ΕΙΔΕΝΑΙ
ΟΡΕΓΟΝΤΑΙ ΦΥΣΕΙ

Glioma-meningioma tumor classification on MRI scans

Written By: Nikolaos Mouzakitis

Date Last Edited: May 20, 2025

1 Introduction

Brain tumors are among the most complex and dangerous diseases affecting the central nervous system. Early and accurate diagnosis is crucial in order to achieve effective treatment on time. Magnetic Resonance Imaging (MRI) is a widely used modality for detecting and characterizing brain tumors due to its high resolution and contrast.

Meningiomas and gliomas are two of the common types of primary brain tumors with different prognoses and treatment strategies. Manual differentiation by radiologists can be time-consuming and subject to variability. One solution for supporting clinical decision making, is an automated classification system. A system like this, can reduce diagnostic workload, increase consistency, and potentially improve early detection rates. It also provides a framework for future research into AI-assisted diagnostics in radiology or related application domains.

In this project, a machine learning-based system that automatically classifies MRIs into meningioma or glioma categories using custom handcrafted, frequency and radiomic features is developed and evaluated. The two machine learning models employed for the classification task of tumors in the two categories are a Random Forest classifier and a Neural Network (Multi-Layer Perceptron).

2 Related Work

In [4], authors review the usage of AI based radiomics and radiogenomics in glioma, while emphasizing in their roles in diagnosis, treatment response prediction and understanding tumor heterogeneity. Also the challenges in standardizing feature extraction and analysis methodologies are addressed.

Duron et al. [3], in their research proposed a radiomics-based classification model for distinguishing gliomas from meningiomas by utilizing T1-weighted MRI scans. The author's approach involved extraction of a vast set of radiomic features which was then followed by the application of machine learning techniques for training and validating a classifier. Results from this particular study demonstrated high diagnostic performance, showing the potential of data-driven radiomics for the support of non invasive tumor characterization. This work can serve as a foundational reference for MRI based tumor classification and can support development of automated systems that may offer assistance in clinical decision-making, similar to the goals of the current project.

Li and co-authors[5], in their work created radiomic models for the prediction of meningioma grade and Ki-67 index, by integrating clinical and radiological features. The models used, demonstrated the potential of radiomics in assessing the biological behavior of meningiomas.

3 Hw and Sw Requirements

The software stack used to implement the classification system consists of Python libraries such as *SimpleITK, OpenCV, scikit-learn, pandas, PyRadiomics, matplotlib, seaborn.* The code and the MRIs used for this project report are available in the following repository [2].

3.1 Data Details

MRIs were sourced from [1], which contains 7023 MRIs of human brain, divided in 4 categories: *glioma - meningioma - no tumor and pituitary.* For this report's purpose the first two categories are utilized, and selected 1000 MRI scans from both *glioma* and *meningioma* classes.

3.2 Method

Related to the methodology of the classification system, for the preprocessing step all images are loaded using SimpleITK and converted into NumPy arrays and their respective pixel intensities are normalized into the range of grayscale images ([0, 255]). In the next step a 10-pixel masking takes place, in order to reduce the black surrounding areas appearing in every MRI. This border mask is applied and excludes irrelevant regions at the edges of the images. The number of the extracted features can be divided into three subcategories:

1)*Features acquired from Pyradiomics*: by utilization of PyRadiomics first-order statistics (mean, variance, entropy) and texture features (GLCM, GLRLM, GLSZM) are extracted.
2)*Custom Features*: have been designed and implemented in order to intuitively help detecting a tumor alike object on an MRI. (*intensity_skewness, intensity_outlier_score, high_intensity_area, max_circularity, top3_circularity_mean, solidity_outlier, abnormal_area_ratio, circular_area_score, asymmetry_score, asymmetry_outlier, boundary_sharpness_mean, boundary_sharpness_max, boundary_sharpness_outlier*).

3)*Frequency Domain Features*: Energy, entropy, mean, standard deviation and skewness in low, mid, and high-frequency bands using FFT.

The frequency feature extraction module analyzes an image's frequency content by decomposing its Fourier spectrum in three frequency bands: low, mid, and high frequencies. In the first place, the image is converted to grayscale (in the case it was an RGB one), and its 2D Fast Fourier Transform (FFT) is computed. The frequency spectrum is then partitioned into concentric circular bands centered around the zero frequency, with radius scaled adaptively to the image dimensions (20%, 20–60%, and 60–100% of the maximum frequency). For each one of the bands, five statistical features have been extracted: energy (total squared magnitude), entropy (spectral uniformity), mean intensity, standard deviation and skewness (asymmetry of the distribution). These utilization of these features characterize the image's texture and structural patterns This way is handling possible variations of the image sizes due to the adaptive band scaling.

In total, 139 features were examined, which is the combination of the three previous described categories. By conducting the feature extraction process, *min-max* normalization is performed in all the generated feature values mapping them on the [0, 1] range as we can observe in Figure 1. This have been performed as a necessary step before training machine learning models to help the models generazibility and for models skewness avoidance.

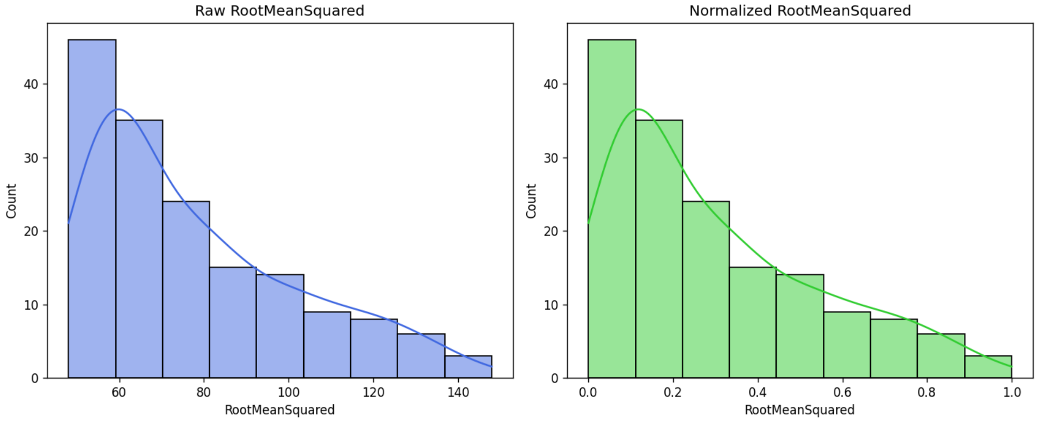


Figure 1: Comparisson of feature distribution prior and after the *min-max* normalization.

4 Evaluation

In the classification stage, two different classifiers have been trained and evaluated, a Random Forest classifier and a MultiLayer Perceptron Neural Network. Classification evaluation examined using different configurations of features in order to be able to compare the contribution of each of the feature sets in the classification accuracy and estimate the degree of their contribution.

The Random Forest model, an ensemble based method known for its robustness and interpretability, was trained using 100 decision trees and a fixed random seed for ensuring reproducibility. For the second classifier, the neural network model was configured having two hidden layers comprising 100 and 50 neurons respectively, trained for up to 500 iterations.

4.1 Classification with texture, shape and statistical features

Classification evaluation with texture, shape and statistical features results are presented below.

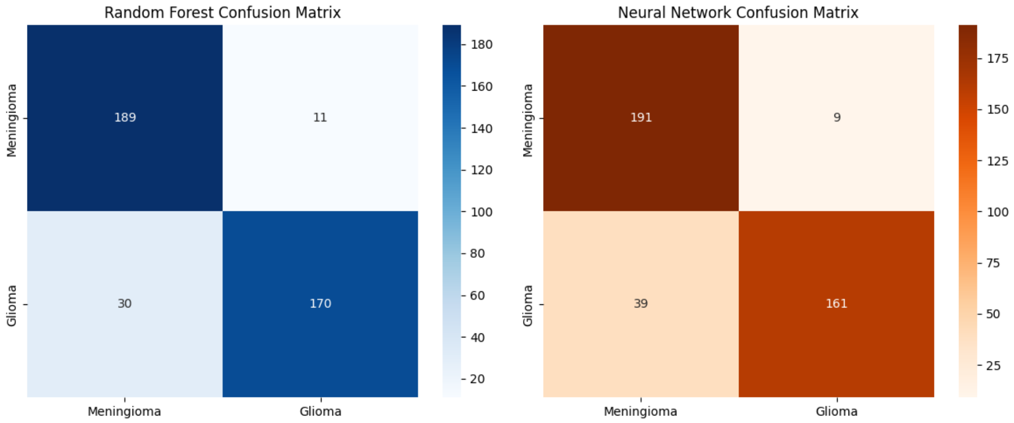


Figure 2: Metric results.

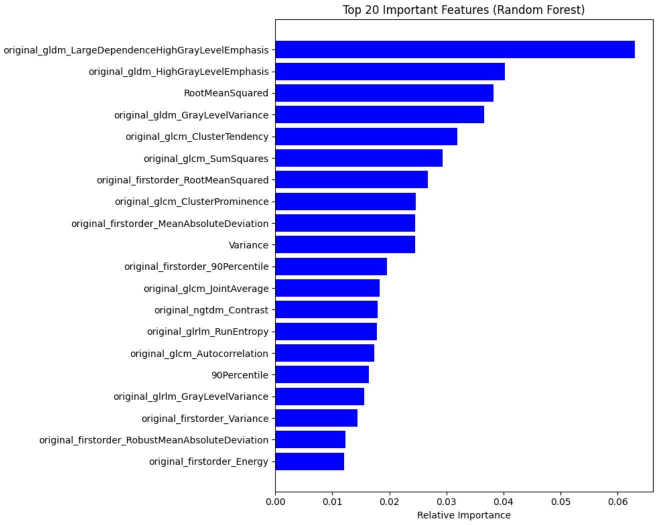


Figure 3: Top20 features of RF.

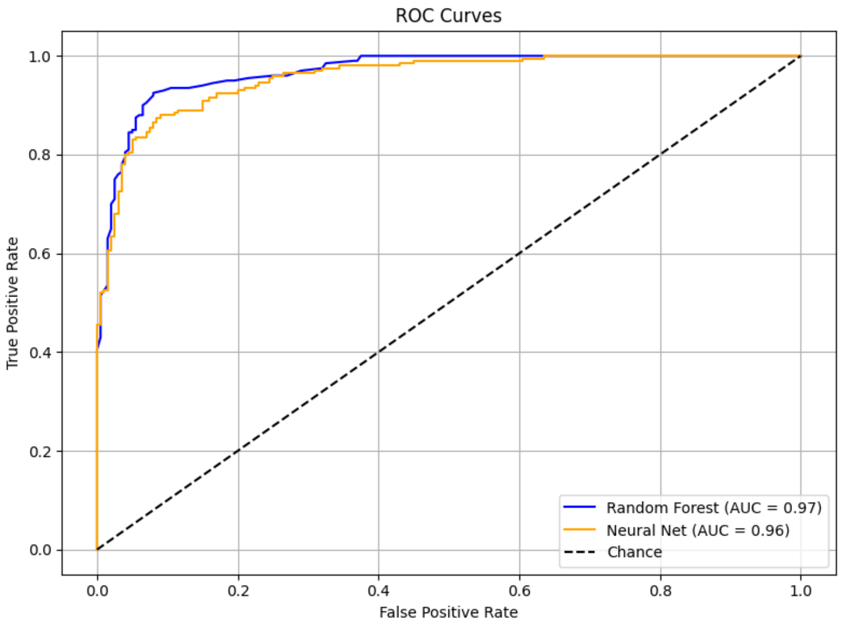
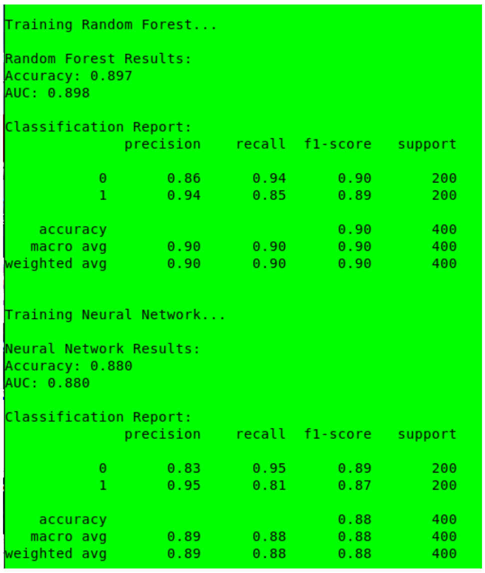


Figure 4: RoC curves for RF and MLP-nn.



```
Training Random Forest...

Random Forest Results:
Accuracy: 0.897
AUC: 0.898

Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.94      0.90       200
           1       0.94      0.85      0.89       200

    accuracy                           0.90       400
   macro avg       0.90      0.90      0.90       400
weighted avg       0.90      0.90      0.90       400


Training Neural Network...

Neural Network Results:
Accuracy: 0.880
AUC: 0.880

Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.95      0.89       200
           1       0.95      0.81      0.87       200

    accuracy                           0.88       400
   macro avg       0.89      0.88      0.88       400
weighted avg       0.89      0.88      0.88       400
```

Figure 5: Classification report

4.2 Classification with frequency features

Classification evaluation's results using the extracted frequency features are presented in the next figures.

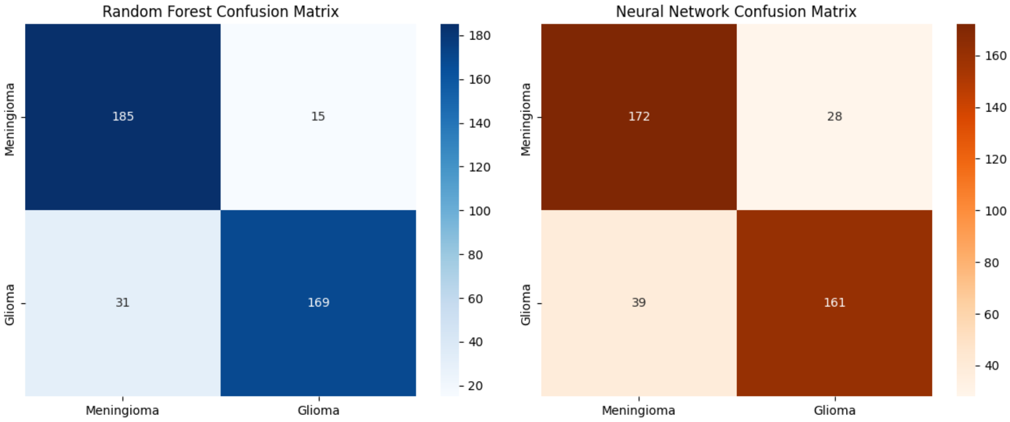


Figure 6: Metric results.

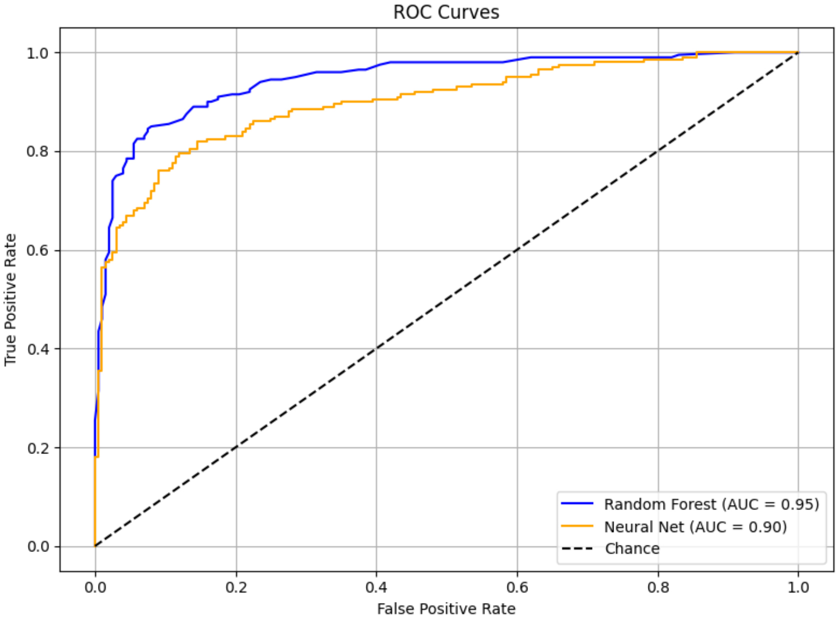


Figure 8: RoC curves for RF and MLP-nn.

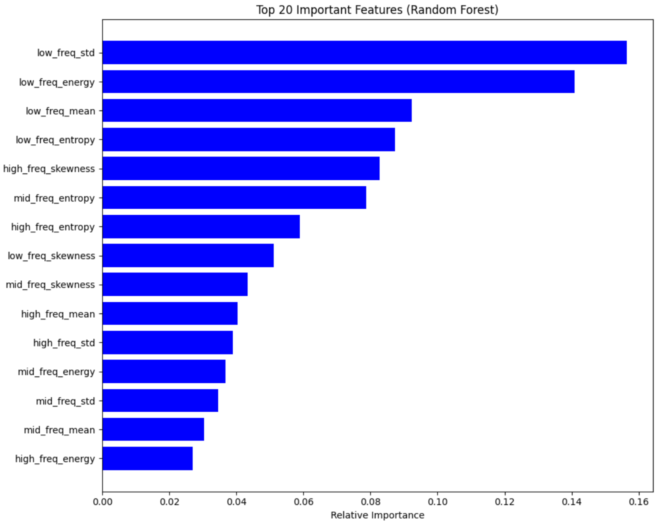
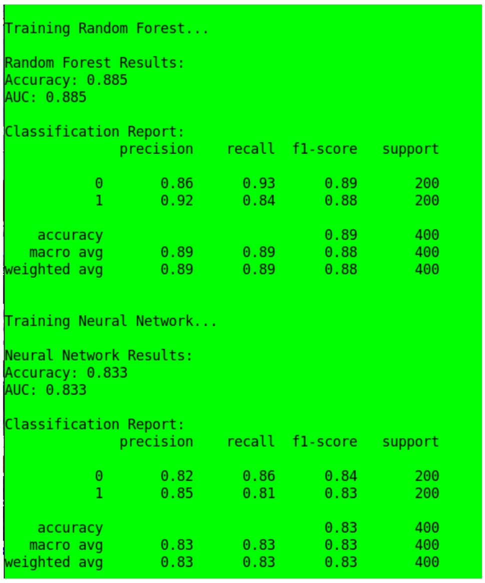


Figure 7: Top20 features of RF.



```
Training Random Forest...

Random Forest Results:
Accuracy: 0.885
AUC: 0.885

Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.93      0.89       200
           1       0.92      0.84      0.88       200

    accuracy                           0.89       400
   macro avg       0.89      0.89      0.88       400
weighted avg       0.89      0.89      0.88       400


Training Neural Network...

Neural Network Results:
Accuracy: 0.833
AUC: 0.833

Classification Report:
              precision    recall  f1-score   support

           0       0.82      0.86      0.84       200
           1       0.85      0.81      0.83       200

    accuracy                           0.83       400
   macro avg       0.83      0.83      0.83       400
weighted avg       0.83      0.83      0.83       400
```

Figure 9: Classification report

4.3 Classification utilizing all features

Classification results utilizing all the available features extracted are presented in following figures.

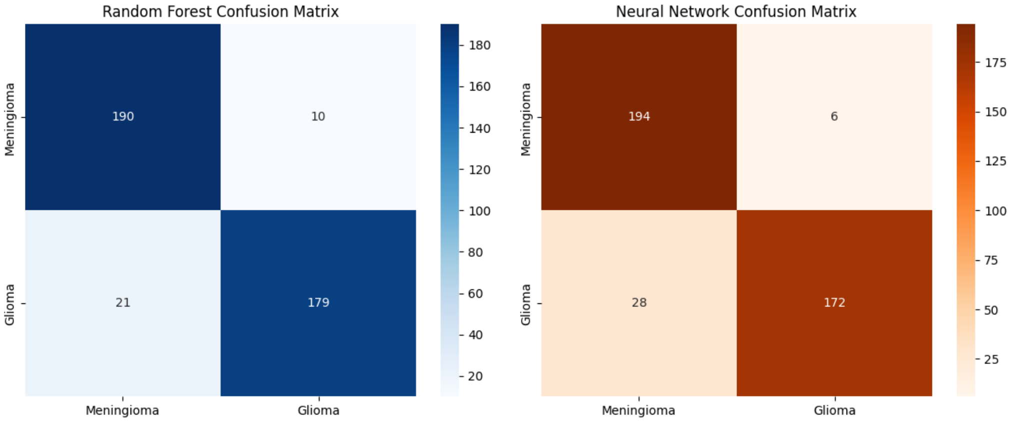


Figure 10: Metric results.

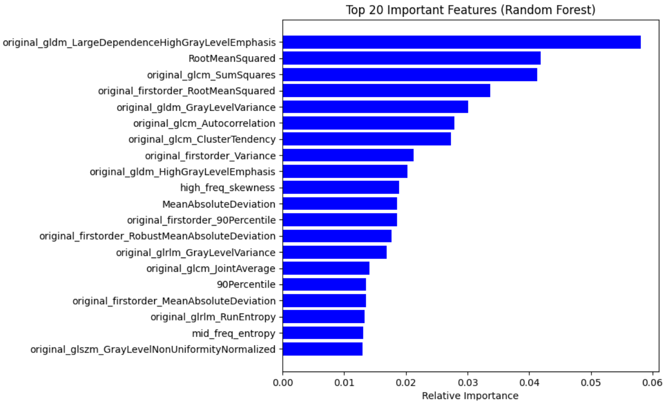


Figure 11: Top20 features of RF.

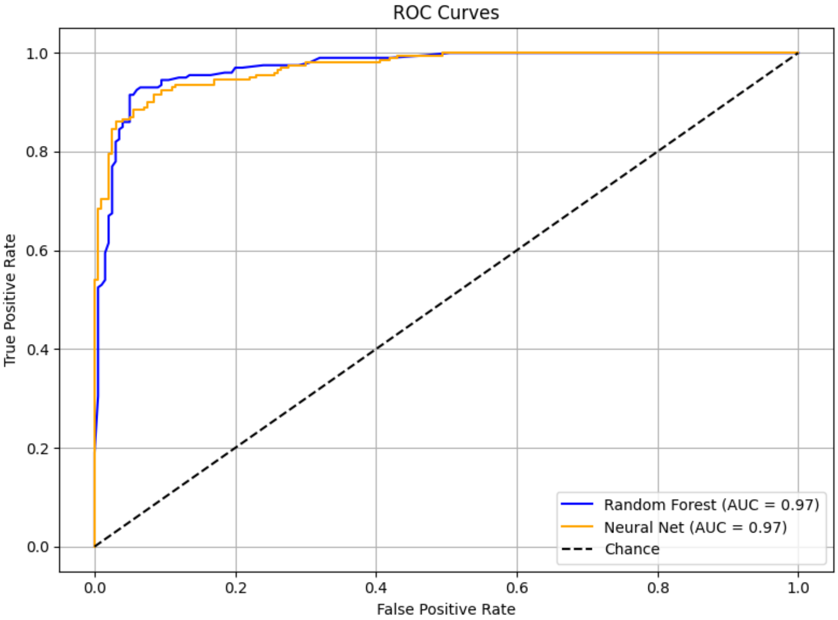
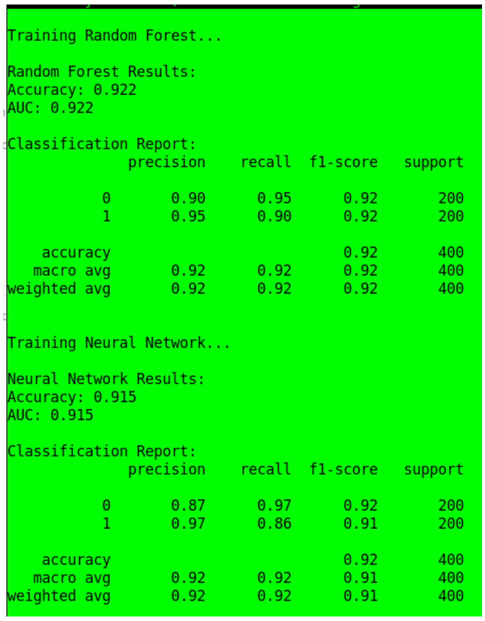


Figure 12: RoC curves for RF and MLP-nn.



```
Training Random Forest...

Random Forest Results:
Accuracy: 0.922
AUC: 0.922

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.95      0.92       200
           1       0.95      0.90      0.92       200

    accuracy                           0.92       400
   macro avg       0.92      0.92      0.92       400
weighted avg       0.92      0.92      0.92       400


Training Neural Network...

Neural Network Results:
Accuracy: 0.915
AUC: 0.915

Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.97      0.92       200
           1       0.97      0.86      0.91       200

    accuracy                           0.92       400
   macro avg       0.92      0.92      0.91       400
weighted avg       0.92      0.92      0.91       400
```

Figure 13: Classification report

4.4 Evaluation Findings

The results from the three feature configurations demonstrate that combining all available features (texture/shape/statistical and frequency domain features) yields the highest classification performance. Key observations include:

- **Feature set comparison:**
  - *All features* achieved the best results, underscoring the complementary nature of radiomic, custom-designed, and frequency-domain features.
  - *Texture/shape/statistical features* alone provided competitive performance, suggesting and confirming the strong discriminative power they yield for tumor classification.
  - *Frequency features* were slightly less effective in isolation but contributed meaningfully when integrated with other features.
- **Glioma classification improvement:** Misclassification rates for glioma dropped significantly (from 31 out of 200 MRIs (15.5%) using texture/shape features only to 21 out of 200 MRIs (10.5%) and from 38 out of 200 MRIs (19%) using texture/shape features only to 28 out of 200 MRIs (14%)) when all features were utilized. This highlights the importance of hybrid feature engineering for capturing tumor heterogeneity.
- **Model performance:** The Random Forest classifier consistently outperformed the MLP neural network across all configurations, particularly in terms of interpretability (e.g., feature importance rankings) and robustness (AUC metrics).

4.5 Classifier Optimization

At this point, hyperparameter optimization is explored in order to achieve better classification results utilizing the created models. In our workflow, first a Random Forest classifier was utilized with hyperparameter tuning to identify the 30 most important features from the dataset. The hyperparameters such as the number of trees (n_estimators), tree depth (max_depth), minimum samples per split and leaf (min_samples_split, min_samples_leaf), and others are optimized using randomized search with cross-validation.

This ensures that the Random Forest model is tuned for the purpose of capturing the most relevant patterns in the data but also for preventing overfitting. After determining the feature importances, the top 30 features are selected as inputs for the next step. These selected features are then fed into a neural network (MLPClassifier) with fixed hyperparameters (e.g., activation function, solver, learning rate, and number of iterations) for performing the final classification. Using the best subset of features reduces dimensionality and noise, and can improve the neural network's performance and training efficiency, while benefiting from complementary strengths of both models for robust classification.

The models who yielded the best performance, have also been modified from their previous configuration. For the Random Forest, the decision trees we increased by a factor of ten to 1000 decision trees, each using the Gini impurity criterion for spliting nodes and a random subset of 15 features per split to ensure diversity and reduce overfitting. In addition, the Neural Network remained a two layer MLP with 100 neurons in the first hidden layer and 50 in the second, activated by ReLU non-linearity. It uses the L-BFGS optimizer, a memory-efficient gradient based method suitable for smaller datasets.

Results from tuned-modified models in this way are presented below.

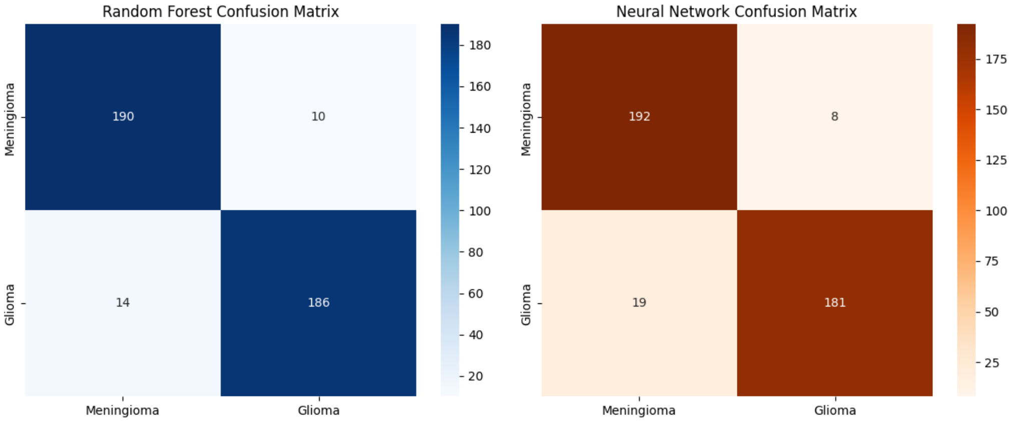


Figure 14: Metric results.

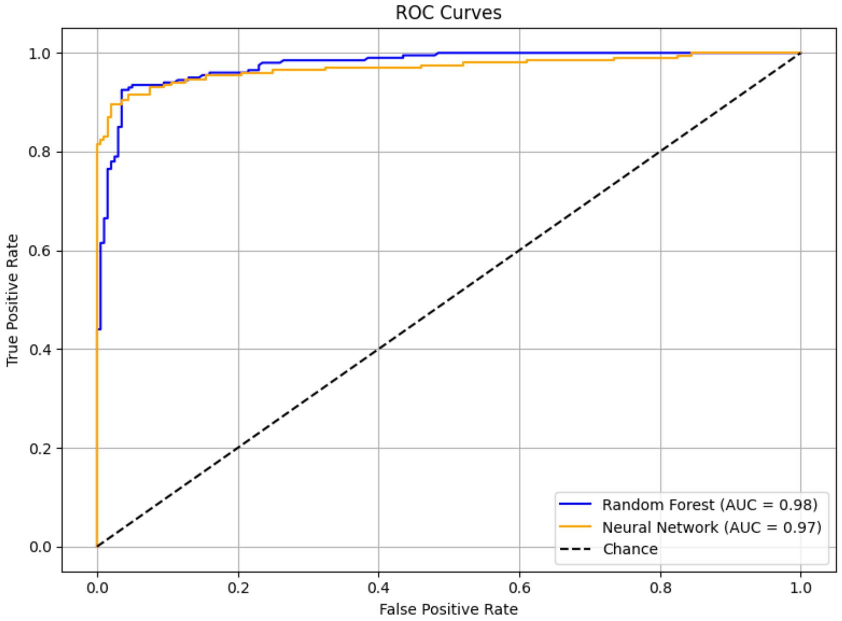


Figure 15: RoC curves for RF and MLP-nn.

After hyperparameter tuning, the Random Forest classifier's accuracy improved from 0.922 to 0.94, indicating a notable gain of approximately 1.8This improvement suggests that optimizing parameters like the number of trees, tree depth, and feature selection/reduction helped the model to better capture patterns in the data, resulting in more accurate predictions.

In a similar way, the neural network 's accuracy increased from 0.915 to 0.932, a gain of about 1.7%. This demonstrates that tuning hyperparameters such as the network architecture, learning rate, and using only the top 30 features selected by Random Forest's model allowed better generalization on the test data.

In overall, both models benefited from hyperparameter optimization, with the Random Forest showing a slightly higher accuracy gain. Improvements like these highlight the importance of fine tuning model settings in order to achieve maximization in classification performance, even in cases of starting from already strong baseline models.

As for the model agreement, a comparative analysis plot in Figure X, shows that both classifiers largely agree and perform well together, with most samples correctly identified by both. Each model also has some unique correct predictions, demonstrating by this their complementary strengths. More strategies could be employed (f.e voting etc) that could improve an overall system's performance. There is a relatively small number of samples misclassified by both models(11/400) samples, highlighting area for further improvements for this project.

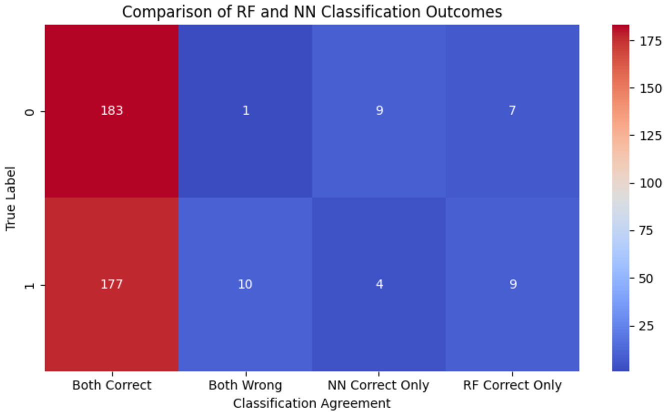


Figure 16: RoC curves for RF and MLP-nn.

5 Histogram Matching enchancement

Analysis of the histograms of the MRIs on the dataset is performed and is presented below. Results of analysing the intensities of the MRIs per category (Figure 17) demonstrate that unnormalized intensity distributions introduce scanner or acquisition protocol-dependent variability that could dominate the signal of interest. Via the application of histogram matching technique image intensities should be aligned on the intensity distributions while preserving the relative contrast relationships essential for the exact tumor classification.

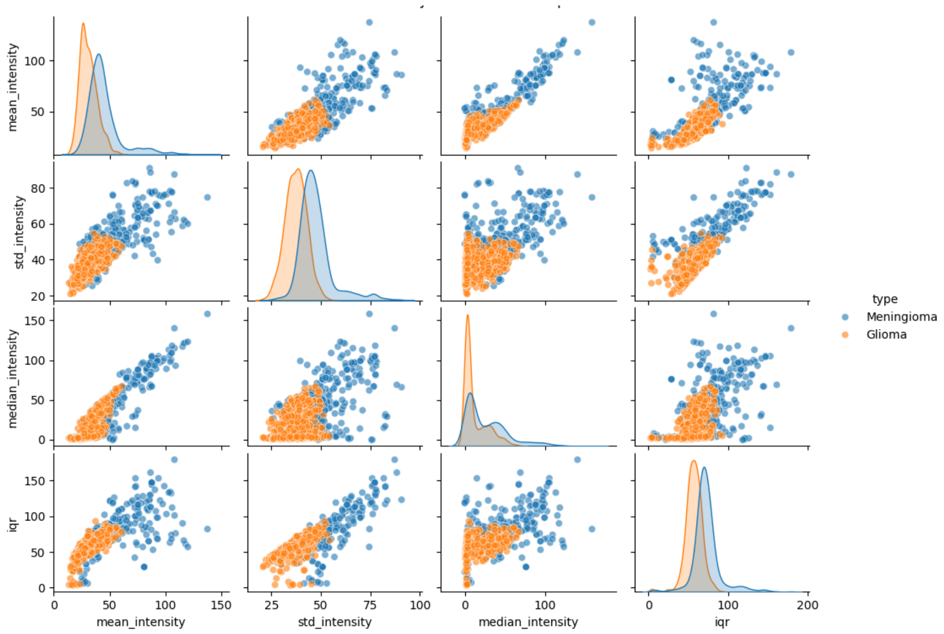


Figure 17: Analysis plot of all the MRI histograms.

Also in Figure 18 we can observe graphical the distribution intensities of the MRIs for both classes.

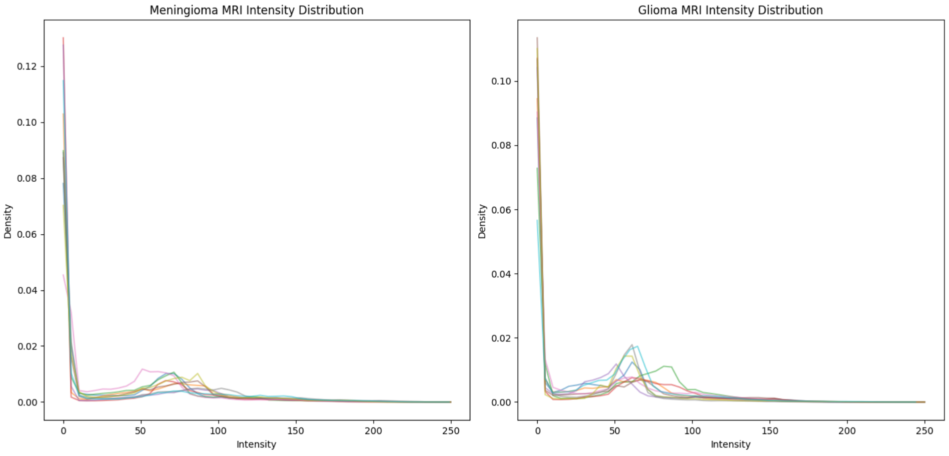
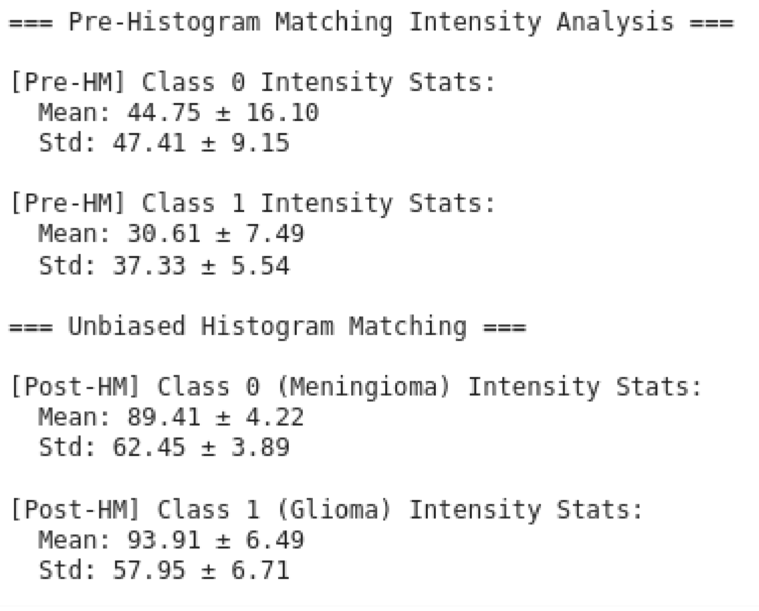


Figure 18: Intensity distributions of MRI histograms.

5.1 Reference image selection

For the proccess of the selection of an optimal reference image for histogram matching, MRIs have been systematically evaluated from the dataset (sampling 300 candidates by default) and choose the one that maximized class separability after histogram matching. Each candidate reference image, is tested on a subset of images (20 per class) by performing histogram matching and then calculating a separability score based on the difference in mean intensities between glioma and meningioma classes normalized by their combined standard deviation. The candidate that produces the highest score (indicating by that the best separation between classes after histogram matching) is selected as the final reference image. By following such an approach chooses a reference that maintains discriminative intensity characteristics between the two tumor types while using a single unbiased reference for all images.

The effect of histogram matching on 500 sampled MRIs belonging to meningioma category and 500 sampled MRIs belonging to glioma category is visualized in Figure 19.



```
=== Pre-Histogram Matching Intensity Analysis ===

[Pre-HM] Class 0 Intensity Stats:
  Mean: 44.75 ± 16.10
  Std: 47.41 ± 9.15

[Pre-HM] Class 1 Intensity Stats:
  Mean: 30.61 ± 7.49
  Std: 37.33 ± 5.54

=== Unbiased Histogram Matching ===

[Post-HM] Class 0 (Meningioma) Intensity Stats:
  Mean: 89.41 ± 4.22
  Std: 62.45 ± 3.89

[Post-HM] Class 1 (Glioma) Intensity Stats:
  Mean: 93.91 ± 6.49
  Std: 57.95 ± 6.71
```

Figure 19: Intensity metrics of 1000 MRIs prior histogram matching.

Prior to histogram matching, there was a greater difference in intensity statistics(mean and standard deviation) between the two classes. Meningioma MRIs exhibited a higher mean intensity (44.75 ± 16.10) and standard deviation (47.41 ± 9.15) compared to glioma MRIs (30.61 ± 7.49 mean; 37.33 ± 5.54 std), indicating class-specific variations in brightness and contrast. After performing histogram matching, the intensity distributions of both classes are aligned more closely to a common reference, resulting in similar mean intensities (89.41 ± 4.22 for healthy, 93.91 ± 6.49 for tumor) and reduced variation in standard deviation between the classes (62.45 ± 3.89 vs. 57.95 ± 6.71). This transformation effectively normalized intensity-related differences, allowing downstream radiomic feature extraction and classification models to focus more on structural and textural patterns.

5.2 Evaluation of classification with histogram matching

Utilizing the histogram matching with the previously described reference image selection method, and by extracting all the features, classification has been re-evaluated. Results of using all the available features extracted from histogram matched MRIs are presented in following figures.

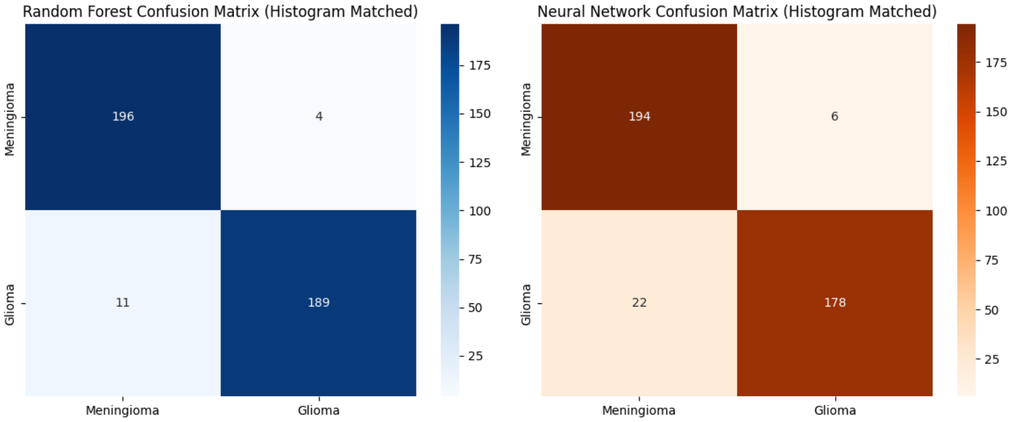


Figure 20: Metric results.

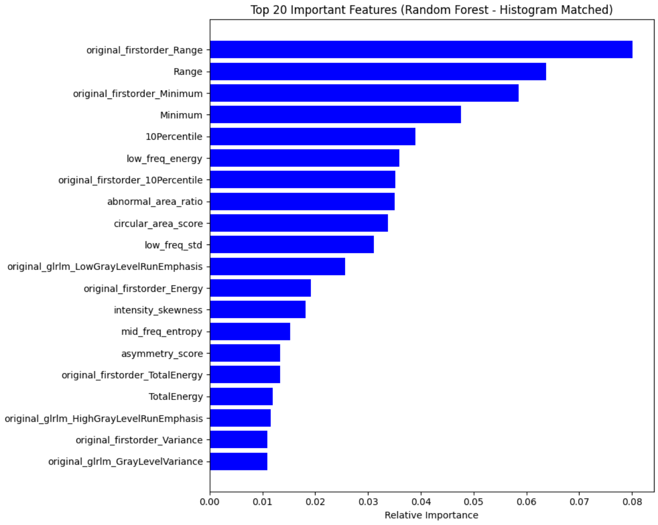


Figure 21: Top20 features of RF.

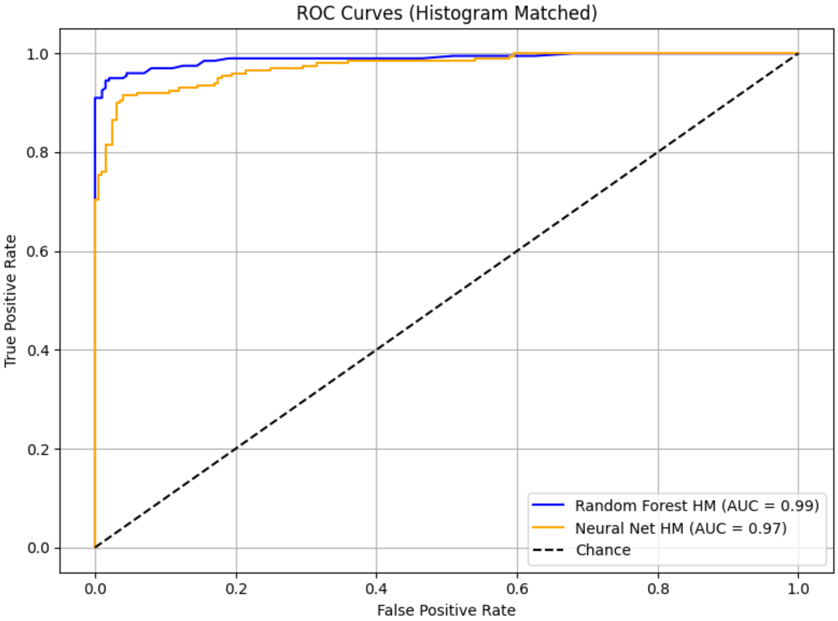
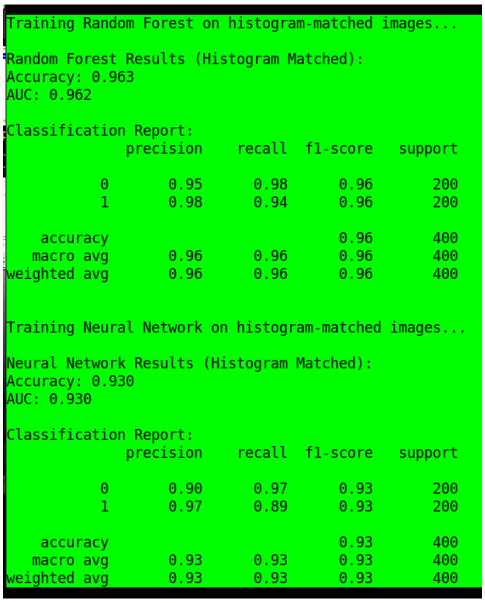


Figure 22: RoC curves for RF and MLP-nn.



```
Training Random Forest on histogram-matched images...

Random Forest Results (Histogram Matched):
Accuracy: 0.963
AUC: 0.962

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.98      0.96       200
           1       0.98      0.94      0.96       200

    accuracy                           0.96       400
   macro avg       0.96      0.96      0.96       400
weighted avg       0.96      0.96      0.96       400


Training Neural Network on histogram-matched images...

Neural Network Results (Histogram Matched):
Accuracy: 0.930
AUC: 0.930

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.97      0.93       200
           1       0.97      0.89      0.93       200

    accuracy                           0.93       400
   macro avg       0.93      0.93      0.93       400
weighted avg       0.93      0.93      0.93       400
```

Figure 23: Classification report

Histogram matching with an optimized reference image improved classification accuracy by 4.2% (RF) and 1.5% (MLP-nn), suggesting that the intensity normalization enhances model generalizability. The larger gain was observed in the Random Forest and by observing the figure refering to the top 20 features contribution for the RF-classification we can inspect that four (20%) of the custom extracted features are included in the list.

6 Conclusion

In this project has been attempted to create an application of an automated system for classifying gliomas and meningiomas using MRI scans, utilizing a combination of radiomic, custom-designed, and frequency-domain features. The Random Forest and Multi-Layer Perceptron classifiers demonstrated strong performance, with the Random Forest achieving slighter superior classification accuracy.

The integration of all feature types yielded the highest classification accuracy, suggesting a complementary nature of texture, shape, statistical, and frequency-domain features. Fine-tuning the models led to improvements, underscoring the importance of parameter selection from all the categories for maximizing performance. In addition, intensity normalization via histogram matching further enhanced model generalizability, reducing scanner dependent variability and improving the outcomes.

References

[1] `https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset` Accessed: 2025-05-09.

[2] Source code and data used in this project: `https://github.com/NikosMouzakitis/cv_master_25`

[3] Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. (2019) Gray-level discretization impacts reproducible MRI radiomics texture features. PLoS ONE 14(3): e0213459. `https://doi.org/10.1371/journal.pone.0213459`

[4] Fan, H., Luo, Y., Gu, F. et al. Artificial intelligence-based MRI radiomics and radiogenomics in glioma. Cancer Imaging 24, 36 (2024). `https://doi.org/10.1186/s40644-024-00682-y`

[5] Li, M., Liu, L., Qi, J. et al. MRI-based machine learning models predict the malignant biological behavior of meningioma. BMC Med Imaging 23, 141 (2023). `https://doi.org/10.1186/s12880-023-01101-7`