**Title:**

Prediction of Brain Tumour (Meningioma) Grade Using MRI Radiomic Features and Machine Learning

**Author:**

10643941

**Affiliations:**

University of Plymouth

School of Engineering, Computing and Mathematics

A logo for a university

Description automatically generated

Prediction of Brain Tumour (Meningioma) Grade Using MRI Radiomic Features and Machine Learning

*10643941*

# ABSTRACT

**Purpose** Meningiomas account for 36% of brain tumours and the preoperative understanding of brain tumour WHO grades is vital for patient treatment. The purpose of this study was to investigate the potential of MRI radiomics features to predict meningioma grades using different models.

**Methods** 89 patients from a 94-patient cohort with meningioma of low-grade and high-grade meningioma was analyzed in a retrospective study. Patients were randomly split into a training set (70%) and test set (30%). 16 out of 50 radiomics features were selected for model training using elastic net and RFECV. Prediction models were built using Support Vector Machine, Logistic Regression and Random Forest methods, hyperparameter tuning was conducted and a ROC curve was used for further evaluation of the best performing model.

**Results** The performance of the 3 classifiers was assessed using the test set. Logistic Regression classifier was the best performing based on performance metrics with an accuracy of 70.3%, recall of 64.3%, precision of 75%, F1-score of 69.3% and an AUC of 0.708.

**Conclusions** This study demonstrated that MRI-based Machine learning radiomics models can predict meningioma grade, further research on feature selection methods and model comparisons may prove to further eliminate knowledge gaps.

**Supplementary Information** Code can be found in a GitHub repository: LINK

***Index Terms*— Logistic Regression, Machine Learning, Magnetic Resonance Imaging, Meningioma, Radiomics, Support Vector Machine**

# INTRODUCTION

Meningiomas, or meningeal tumours, are the most commonly occurring intracranial (primary) tumors comprising over 36% of brain tumours. Meningiomas have an annual incidence rate of up to 76.1 per 1 million individuals worldwide and the incidence rate for women is over double that in men (Wiemels, Wrensch and Claus, 2010; Yan et al., 2017; Zhao *et al.*, 2024). The World Health Organization (WHO) categorizes meningeal tumours into three grades (Louis et al., 2021). Whilst grade I meningiomas are benign and account for most cases, high grades tumours patients (grade II and III) have shown to have poor prognosis with a 30-50% recurrence incidence rate greater than grade I patients over a 10-year follow up period (Zhao *et al.*, 2024).

Currently, the gold standard for diagnosis of meningioma is radiological diagnosis and investigation using magnetic resonance imaging (MRI), however, limitations to this still exist such as the tumour grading from preoperative scanning alone is not yet defined. The use of machine learning (ML) with radiomic features shows promise for meningeal tumour grading using existing preoperative scans and radiomic data (Ugga et al., 2021; Coroller et al., 2017). The use of radiomics for prediction of brain tumor grades would serve as a non-invasive technique and has already shown to identify the pathophysiological situation of tumours to levels which are non-identifiable with the human eye (Han et al., 2021; Han et al., 2023).

This study aims to investigate the use of supervised ML models and radiomic features to predict meningioma grades.

# LITERATURE REVIEW

## A. Machine Learning and Applications in Brain Tumour Assessment

Machine learning (ML) is a specialized branch of artificial intelligence (AI) which has recently gathered increasing interest for its potential use in medical fields such as radiology for cardiovascular applications and tumour identification in the oncology field (Cuocolo et al., 2019; Cuocolo et al., 2020). ML is able to essentially learn from existing data to make predictions without the need for manual, or explicit, programming beforehand. The three main ML algorithms include: unsupervised learning where ML models identify patterns from unlabeled data, supervised learning in which known labelled data is used to train ML models and reinforcement learning whereby a dynamic environment is established through software training utilizing a trial-and-error technique (Cuocolo et al., 2019). There are several ML techniques which can be applied in brain tumour (meningioma) assessment mostly falling into the unsupervised, e.g., convolutional neural networks (CNNs), and supervised learning models e.g., support vector machines (SVM) and random forests (RF) (REFERENCE).

This study will look at supervised ML techniques. Both SVM and RFs are utilized for prediction, regression tasks, and for classification making them ideal ML models for the analysis of MRI data for tumour classification and grading, however, do so differently. SVM is a type of non-linear classier which, through the data gathered from radiomic features in MRI, produces a hyperplane that is able to divide different meningioma classes. Whereas, RF, a type of ensemble method, provides predictions through use of multiple decision trees and learning patterns between known meningioma grades and radiomic features. As RF is non-parametric, it does not make assumptions on data distribution nor the relationship between target and data features, it is therefore robust to feature collinearity thus allowing for accurate predictions to be made (Gu et al., 2020).

Thus far, no ML methods for applications in brain tumour assessment have been chosen as the most fitting for analysis, instead the choice is up to individual researchers (REFERENCE). Table 1 presents the commonly used ML methods for prediction of meningioma grades and analysis with results from various studies, these include, SVM, LR, RF, naïve bayes (NB), CNN and linear discriminate analysis (LDA). Based on statistical ML performance of studies presented in table 1, it seems that LR is the best performing algorithm for the prediction of brain tumour grades, in this study researchers used features from MRI based on T1W1 images and research concluded an accuracy of 92.9%, sensitivity of 91.7%, specificity of 100% and a AUC of 0.948 using LR. However, this study had a relatively small study group of 98 patients with most, 82 patients, presenting with WHO grade I meningiomas which could suggest an increased risk of overfitting (Chu et al 2020). In addition, it seems that RF and SVM are the most wildly researched ML methods for applications in bran tumour assessment, with some researchers offering a comparison of the two within their study. Park et al 2018 study on meningioma grade prediction and histological subtypes found that SVM was a better performing method in comparison to RF, using validation set data with RFE feature selection and SMOTE as subsampling, presenting an accuracy of 89.7% and 82.4% respectively, furthermore the AUC for SVM was also higher than that of RF (0.86 and 0.82 respectively) (Park et al 2018). On the other hand, another study found there to be no difference between the AUC of Rf and SVM (0.884) (Duan et al., 2022). Having said this, attributes of RF such as it being a non-parametric ensemble method, as mentioned previously, and robustness towards outliers resulting in a reduced risk to overfitting suggest that it may be a better fit for the prediction of meningioma grades (Gu et al., 2020; REFERENCE).

**Table 1: Commonly used Machine Learning Methods for Meningioma Grade Prediction and their Performance**

|  |  |  |
| --- | --- | --- |
| ML method | ML performance | Reference |
| Support Vector Machine | (Validation set)  Accuracy= 89.7%  Sensitivity= 75.0%  Specificity= 93.5%  AUC = 0.86 | (Park et al 2018) |
| Logistic Regression | (Test set performance)  Accuracy = 92.9%  Sensitivity = 91.7%  Specificity = 100%  AUC = 0.948 | (Chu et al 2020) |
| Random Forest | (Validation set with RFE feature selection and SMOTE as subsampling)  Accuracy =82.4%  Sensitivity=58.3%  Specificity=89.1%  AUC= 0.82 | (Park et al 2018) |
| Naïve Bayes | Accuracy= 0.86-0.89  Sensitivity= 0.57-0.76  Specifcity= 0.92  AUC= 0.88-0.91 | (Yan et al 2017) |
| Convolutional Neural Network | Accuracy = 83.3% | (Zhu et al., 2019) |
| Linear Discriminate Analysis | Accuracy = 75.6% | (Chen et al., 2019) |

Furthermore, as mentioned above, LR was the best performing method for grade prediction and NB, despite having low sensitivity (0.57-0.76) has a fairly high accuracy up to 89%. However, it is worth noting that both ML methods have limitations (Chu et al., 2020; Yan et al., 2017). No multicollinearity amongst features is assumed in logistic regression and therefore if features have high correlations, incorrect estimates will be established, and incorrect interpretability may occur, whereas NB requires features to be discretized into categorical variables which may lead to the loss of some information and thus risk of decreased precision (REFERENCE). A study by Chen et al 2019 also identified a higher AUC with LDA over SVM when using LASSO for the prediction of WHO grade 1 meningioma (0.934 and 0.840 respectively) further reinforcing the idea that method selection is dependent on researcher’s preference (Chen et al., 2019; REFERENCE). However, due to LDA’s linear limitations, SVM therefore may be a better method choice in cases where the linearity of the data is unknown.

## B. Radiomics and Applications in Brain Tumour Assessment

Within the medical research field, radiomics is a quickly developing sector focusing on identifying quantitative features, also known as radiomic features, from medical diagnostic images such as MRIs. Using radiomics, tumour phenotypes can be distinguished using the quantitative data of various radiomic features, in addition it enables for the characterization of intralesional heterogeneity which could result in patient specific personalized treatment (Niu et al., 2019; Zhao *et al.*, 2024).

.

Studies utilizing radiomics for meningioma assessment have, on the most part, been established using MRI(Han et al., 2023; Li et al., 2023; Zhou et al., 2018). Some reasons for this are its ability to deliver superior anatomical information such as spatial location and their ability to characterize physiological processes in association with brain tumours such as meningiomas due to varying MRI sequences sensitivity to tumour physiology (Zhou et al., 2018). Radiomic application in brain tumour, in this case, meningioma, diagnosis and treatment usually aims to predict tumour grading, however, other applications also exist. A summary of some studies, their aim, results and relevant ML information is presented in table 1.

**Table 2: Radiomic Applications in Meningiomas and Summary of Current Studies**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Aim | Features | Feature Selection | Analysis | Statistical results & Conclusion | Reference |
| Tumour Grading | Wavelet transform, Co-occurrence matrix, Histogram |  | Support Vector Machine (SVM), Linear Regression (LR), (Naïve Bayes) (NB) | Training data presented with area under the curve (AUC) = 0.87 and p-value = <0.0001. SVM performed the best an accuracy of 0.87. | Yan et al., 2017 |
| Tumour Grading | 3,376 features were extracted | Selectprecentile, LASSO (Least absolute shrinkage and selection operator) | Random Forest (RF) | Training and validation sets showed AUC = 0.949 and 0.838 respectively. Results showed RF to be the best performing model and high accuracy for grade prediction using clinical radiomics model. | Han et al., 2023 |
| Tumour Grading | Wavelet (transform?), Histogram, Texture |  | RF | Validation data (?) presented with AUC = 0.78 and P-value = <0.0001. Results show radiomic application was able to distinguish different meningioma grades. | Coroller et al 2017 |
| Tumour Grading | CE-T1, T2 and BTI image features? | LASSO | SVM | AUC = 0.912 and 0.896 for training and testing data respectively for nomogram. Results showed the combined use of radiomic scores and clinical data had better diagnostic performance compared to only radiomic or clinical models alone. | Zhao et al 2024 |
| Histological subtype and tumour grading | Texture, histogram and morphology | Recursive feature elimination (RFE) or no subset selection | RF and SVM | AUC, accuracy, sensitivity and specificity calculated (0.86, 89.7%, 75.0% and 93.5% respectively). From the best performing feature selection using a combination of RF, SVm and SMOTE (DEFINE SMOTE). Results showed possibility to distinguish between tumour grades based radiomic features using ML classifiers. | (Park et al., 2018) |

From table 1, it is clear that many studies focus on predicting tumour grading utilizing radiomics this may be because current grading, which is vital to be established for selecting suitable treatment options, uses invasive procedures such as biopsy of tumour or surgery to gather decisive histopathology information (Nowosielski et al., 2017). It may be that radiomics could prove to be a non-invasive preoperative alternative for the prediction of meningioma grades.

In the recent years, a range of research groups have began exploring the accuracy and reliability of radiomic features for prediction of meningioma WHO grades, with many presenting that features such as wavelet-transform, texture, histogram and more are able to do so (Park et al., 2018; Yan et al 2017; Chu et al., 2021). A 2018 study with a validation set of 12 high-grade and 46-low grade patients, found 90 radiomic texture features including 3 categories (run-length matrix-based parameters, co-occurrence matrix and histogram) and had an accuracy of 89.7% and specificity of 93.5% suggesting radiomic texture features might be a promising application for tumour assessment, furthermore this study identified a significant difference in texture parameters for fibroblasts and non-fibroblasts (Park et al., 2018). A previous study also identified significant texture features from the run-length matrix-based parameter category for both low and high-grade WHO prediction further implying the potential importance of texture features for radiomic use in brain tumour assessment (Yan et al., 2017). Furthermore, it seems that multiple radiomic features are commonly used throughout studies, for example morphology features (Huang et al., 2019). From these, an increase in heterogeneity and abnormal malignancy formation has been identified to be in association with high-grade meningiomas and could suggest an increased risk of tumour recurrence/progression. However, it is vital to acknowledge that there is no specific morphologic feature able to distinguish solely malignant meningiomas, this means similar applications could lead to similar findings in low-grade meningiomas (Huang et al., 2019; (Mahmood et al., 1993). It seems that the lack of specificity from radiomic morphologic features alone implies a difficulty to differentiate between brain tumour grades using only this.

Whilst radiomic features have the ability to increase probability of specific grades, a combination of radiomic findings with the histopathological and patient clinical data is vital for accurate diagnosis as well as patient treatment and prognosis for those with meningiomas. Studies already show promise of enhanced classification model performance when the combined use of radiomic features from varying feature groups, patient clinical data and more is employed. For example, a study, with details in table1, used over 3000 features and presented with a training set AUC of 0.949, whereas the study by Yan et al 2017 on wavelet-transform and co-occurrence matrix feature has an accuracy of 0.87 for its best performing model (Han et al., 2023; Yan et al., 2017). Not only this but, corroler et al 2017 also found promising results for radiomic use in differentiating between meningioma grades (AUC = 0.78) (Coroller et al., 2017). However, it can be noted that these papers all had different best performing ML models, implying that further optimization is necessary in future research to establish a standard protocol.

Other applications of radiomics in brain tumour assessments have also been researched, including histological subtypes and the prediction of meningioma relapse. One study reported ability to identify between low-grade meningioma subtypes, specifically, angiomatous mengiomas using MRI features only (Zhang et al., 2018). In addition, Park et al 2018 identified the non-fibroblastic and fibroblastic subtypes from variations in texture features (Park et al., 2018). Some studies have instead focused on the capabilities of radiomic features to predict meningiomas recurrence. For example, a study on 155 patients with WHO grade 2 meningiomas identified the combined clinical, pathological and radiomics model to outperform the study’s model 1 (clinicopathological model) with a significant AUC of 0.78 for the training set (Park et al., 2022). This suggests that radiomics may significantly contribute to the accuracy of brain tumour recurrence further presenting as another potential use for radiomics and its applications in brain tumour assessment.

## C. Knowledge Gaps

The use of radiomics for the diagnosis, progression monitoring and treatment of meningiomas is a relatively new tool and therefore comes with its limitations and current knowledge gaps. Firstly, current research consists of retrospective studies, utilizing existing patient data, which suggests and implication with possible selection bias. In addition, unicentric studies also pose with the same limitation, a multicentric study could therefore reduce selection bias and improve reliability of results (Coroller et al., 2017; Li et al., 2023; Zhao *et al.*, 2024). In addition, studies present with small sample sizes and imbalance in group sizes for meningioma grades increasing the change of overfitting and reducing the statistical power of these studies (Yan et al., 2017; Li et al., 2023). This suggests potential increased risk of type II errors and incorrect generalizations of the population, in this case, patients with meningioma. Furthermore, heterogeneity of patient study cohorts poses as a significant limitation to meningioma ML research as it results in lower-quality raw data and reduces study impact, heterogeneity therefore can further increase inappropriate generalization and study conclusions (Court et al., 2016; Zhao *et al.*, 2024). Improvements to patient cohorts/data collection methods are therefore vital for the enhancement of meningioma research quality in turn enabling for better patient care/treatment.

## D. Abbreviations and Acronyms

Machine Learning- ML

Artificial Intelligence- AI

Support Vector Machines- SVM

Convolutional Neural Networks – CNN

Random Forest – RF

Area Under Curve – AUC

Least absolute shrinkage and selection operator- LASSO

Linear Discriminate Analysis – LDA

Logistic Regression – LR

Naïve Bayes - NB

# METHODOLOGY

## A. Dataset

Dataset (received in csv. format) for this study used MRIs from The Cancer Imaging Archive (TCIA), a public dataset (Vassantachart et al., 2023). This dataset includes the radiomic features from pre-operative MRI scans extracted from TCIA for 94 patients with Grade I and Grade II meningiomas diagnosed in 2010 to 2019. No clinical data, aside from meningioma grade, was received. The original dataset included 96 patients, however, patients with Grade III meningiomas were not included due to the low number of Grade III patients. The ‘Target’ represents the meningioma grade where the value ‘0’ signifies Grade I and the value ‘1’ signifies Grade II meningioma, all other columns represent the various radiomic features which were extracted from the pre-operative MRI scans. Meningioma grade confirmation was established based on the current guides for classification.   Data was processed (?) using python and code can be found on the Juptyr Notebook uploaded on Github (LINK)

## B. Image Acquisition

MRI scans were conducted using a range of 3T/1.5T GE scanner, full details can be found in previous research by Vassantachart et al ., 2022 (Vassantachart et al., 2022). All patients in this study underwent MRI before surgery with a slice thickness of 1-2 mm. This included two standard sequences for meningioma assessment: T1-CE and T2-FLAIR imaging. Following MRI scans, patients underwent surgical subtotal or gross total resection. Tumour tissue was then analyzed for diagnosis confirmation and meningioma grading (Grade I or II) (Vassantachart et al., 2022).

## C. Radiomic Feature Collection and Extraction

Radiomics features were extracted by researchers at the University of Plymouth using Python. 48 features were extracted from the two MRI sequences (T1C and T2-FLAIR) including shape, first-order statistics and texture features. Radiomics shape features relayed the size and morphology of the tumours whilst first- order statistics describe the voxel intensity distributions for lesion characterization and texture features allow for measurement of voxel intensity variation between the two MRI sequences. Texture features were obtained by using Grey-level Co-occurrence Matrices (GLCMs, Grey level size zones matrices (GLSZMs, gray level run length matrices (GLRLMs), and gray level difference matrices (GLDMs). After the wavelet radiomics features were analyzed through wavelet decomposition on original images at low and high frequencies. (REFERENCES?)

## D. Feature Preprocessing

Data was processed using Python code in the computing platform, Jupyter notebook. A repository on GitHub to store all python codes produced for this study can be found using the link : (ADD LINK). Firstly, the dataset received was inspected to identify the number of entries and features, after this, data cleaning was conducted. A check for missing data was done on each feature by working out the percentage missingness. Following this, data was checked for any duplicates and non-numeric features were removed. Descriptive statistics were complete and the Log2 fold-change between grades was established.

Before feature selection and model development, data scaling was complete on the extracted radiomic features using the StandardScaler from the scikit-learn library (REFERENCE). ‘StandardScaler’ is a widely used technique for feature standardization ensuring all features have a 0 mean and unit variance allowing for a consistent model performance through the removal of feature scale discrepancies (MAYBE REFERENCE?)

Outlier detection was then performed using Isolation Forest on the scaled data. Outliers were filtered and removed from the dataset, by incorporating this technique, the quality and reliability of the subsequent models should be improved.

## E. Data Training and Splitting

Analysis was conducted using Python and a random seed was set for consistent model evaluation and to increase reproducibility of research. The dataset was randomly split into two subsets: a training and a testing set. A ratio of 70:30 was used as this ratio is a common standard practice in ML (REFERENCE?). The training set, consisting of 70% of the data, was utilized to train the ML models whilst the testing set, the remainder of the data (30%), was used for evaluating the models’ performances. The training set was used to train 3 ML classifiers: RF, LR and SVM. Accuracy on the test data was checked for all 3 prior to feature selection to allow for a later comparison. In addition, RF feature importance was completed and plotted using the permutation importance function of the scikit-learn library. This allows to analyse the importance of each radiomics feature (REFERENCE LIBRARY AND FUNCTION).

## F. Radiomic Feature Selection

As discussed previously, this study was conducted to accomplish one aim using machine learning models: predict the WHO grade of meningioma. Radiomics feature selection is vital to reduce risk of overfitting. Feature ranking by normalized importance was conducted to allow for an evaluation of the contribution of each radiomics feature, scores were normalized to increase interpretability based on the estimator values from RF classifier. These results were not used for feature selection, however, still bring insight on the radiomics features for predicting meningioma grades. Feature ranking using a wrapper-based method, recursive feature elimination (RFE), was also conducted but used as the final feature selection method.

Feature selection was based on a wrapper-based feature selection process including Elastic Net and recursive Feature Elimination with Cross-Validation (RFECV) to select the optimal number of features for model training.

## G. Development Radiomics Models

A total of 16 radiomics features were selected and three classification methods were considered for the model training including RF, LR and SVM due to their common use in prediction of meningioma grade research. Each model was trained with the optimal features selected using the training dataset and mean accuracies were calculated. In addition, a comparison of the algorithms was conducted using the training dataset and boxplots were produced to visually present the results.

To assess the performance of each algorithm to predict meningioma grades, accuracy, precision, recall, F1-score and AUC was calculated by evaluating the trained models on the test dataset.

## H. Hyperparameter Tuning and Further Model Evaluation

Hyperparameter tuning was performed using a grid search method with cross-validation, ‘GridSearchCV’ from the scikit-learn library was used. For the 3 models, LR, RF and SVM, appropriate hyperparameters were tuned to enhance performance. For LR this included the regularization strength, C, and the penalty type L2. The number of estimators and maximum depth of trees were tuned for RF and for SVM, C and a linear/radial basis function for the kernel type was selected. A CV equal to 5 was used for all models as this is common practice for ML models. Models were then re-evaluated on the test dataset and performance metrics were calculated again.

In addition, the Receiver Operating Characteristic (ROC) curve was plotted for the best performing model, LR. The ROC curve graph plots the sensitivity i.e. the true positive rate against the specificity i.e., the false positive rate using the information gathered from the tuned test set LR model. AUC was also calculated to show the overall model performance.

## I. Statistical Analysis

Feature preprocessing, data training and splitting, radiomics feature selection and development of radiomics models were all conducted using version 3 of Python. The library “scikit-learn” was used throughout this project and therefore was a major contributor to the selection of features as well as the development of models.

# RESULTS

## A. Data Cleaning and Outlier Detection

This dataset consisted of 94 entries (patients) and 50 radiomics features, percentage missingness was conducted and found 0% missingness for data in all features. In addition, no duplicates were found and 1 non-numeric feature, ‘Subjects’, was removed. Finally, outlier detection using Isolation Forest was able to detect 5 outliers, these were removed from the data. In this study, radiomics data from 89 patients was used and 49 radiomics features including meningioma grade.

from the data. 89 entries and 49 features, including grade, were used for this study.

## B. Exploratory Statistics

Log2 fold changes for all radiomics features were calculated and a visual representation of the fold changes between the grades can be seen in figure 1. A positive fold change implies an upregulation in a group whilst a negative fold change suggests the downregulation of the feature between the two grades. Results from figure 1 show a log2 fold change of approximately 2 for the ‘wavelet-HLH\_firstorder\_Mean\_t1c’ and ‘wavelet-HHH\_firstorder\_Skewness\_t1c’ features meaning a 4-fold increase in these features’ expression in grade 2 meningioma in comparison to grade 1. In addition, there is a log2 fold change of approximately -9 and -8 for features ‘original\_glszm\_LowGrayLevelZoneEmphasis\_t2f’ and ‘original\_glszm\_HighGrayLevelZoneEmphasis\_t2f’ respectively. This analysis revealed a difference for these features between grade 1 and grade meningioma indicating a 512- and 256-fold change reduction for grade 2 in comparison to grade 1 meningioma. In figure 1, it can also be identified that most other radiomics features have been downregulated.

Box plots for the 4 features above were produced (see Figure A1 in the Appendix). Box plots all presented with overlapping interquartile ranges, therefore, no significant difference between the tumour grades was found for all 4 features.

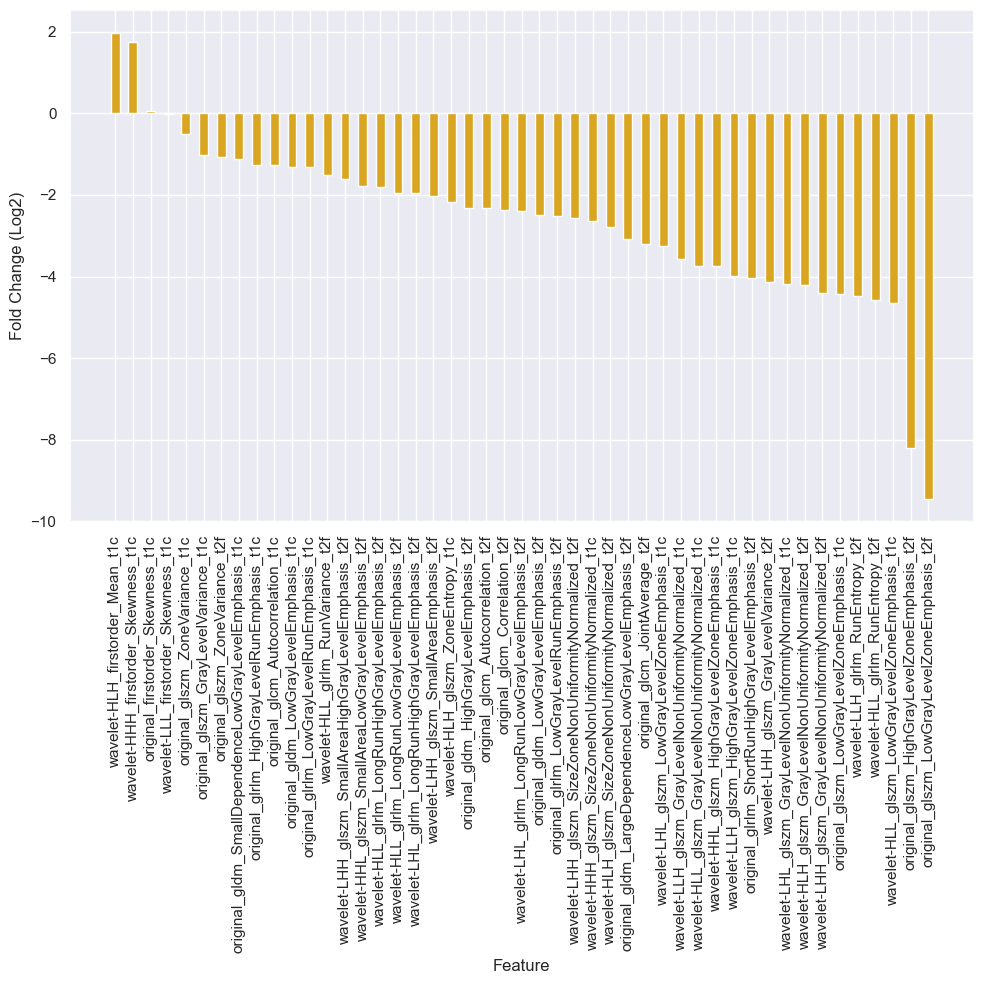


Figure 1: Log2 Fold Change between meningioma grades 1 and 2 for all 48 radiomics features.

## C. Data Training and Feature Selection

Dataset was initially trained prior to feature selection with 3 ML classifiers: RF, SVM and LR. Accuracy on test data was checked for all, results presented with RF having the highest accuracy (0.85) and the LR to have the lowest performance (0.63), in addition SVM performed relatively like RF with an accuracy of 0.81.

For feature selection initially, two methods were employed for feature ranking by importance: normalized importance and RFE. The normalized importance method found, as seen in figure 2, found the feature ‘wavelet-HLH\_glszm\_GrayLevelNonUniformityNormalized\_t2f’ to be the highest-ranking feature. In comparison, in the model using RFE (please see Figure A2 in the Appendix) this feature was ranked as the 5th most important. ‘Wavelet-HLL\_glrlm\_RunVariance\_t2f’ was the 2nd highest ranking feature for both methods. Furthermore, ‘original\_glszm\_LowGrayLevelZoneEmphasis\_t1c’ was one of the lowest ranking features by importance for both methods (ranked 48th by normalised importance and 45th by RFE).

Whilst analysis was conducted using these methods and have provided further insight into the relevance of features, they were not used as the final tools for feature selection. Instead, a more robust feature selection method was conducted through the combined use of Elastic Net and RFECV. Elastic Net was able to reduce the number of optimal features from 48 to 19, the use of RFECV on features selected allowed for further feature elimination. RFECV found 16 features to be the optimal number for analysis. Selected features have been displayed in table 3, ranking in both techniques previously mentioned have also been added as a comparison. The final method had 9 features in common with feature importance by normalized importance and 7 in common features with the RFE method for feature importance by ranking.

A graph with blue and black text

Description automatically generated

Figure 2 Radiomics Feature Ranking Based on Normalized Importance.

Table 3: A comparison of the final 16 Radiomics Features Selected with feature ranking by improtance.

|  |  |  |  |
| --- | --- | --- | --- |
| Selected Features | | Importance Rank | |
| **Normalized Importance Method** | **RFE Method** |
| 1 | Original\_firstorder\_Skewness\_t1c | 19 | 7 |
| 2 | Wavelet-LHL\_glszm\_GrayLevelNonUniformityNormalized\_t1c | 16 | 29 |
| 3 | Wavelet-HLL\_glszm\_GrayLevelNonUniformityNormalized\_t1c | 12 | 12 |
| 4 | Wavelet-HLH\_firstorder\_Mean\_t1c | 21 | 24 |
| 5 | Wavelet-HHL\_glszm\_HighGrayLevelZoneEmphasis\_t1c | 7 | 4 |
| 6 | Wavelet-HHH\_firstorder\_Skewness\_t1c | 11 | 3 |
| 7 | Wavelet-HHH\_glszm\_SizeZoneNonUniformityNormalized\_t1c | 29 | 26 |
| 8 | Original\_glszm\_HighGrayLevelZoneEmphasis\_t2f | 35 | 37 |
| 9 | Wavelet-LHH\_glszm\_GrayLevelNonUniformityNormalized\_t2f | 13 | 16 |
| 10 | Wavelet-LHH\_glszm\_SizeZoneNonUniformityNormalized\_t2f | 18 | 31 |
| 11 | Wavelet-LHH\_glszm\_SmallAreaHighGrayLevelEmphasis\_t2f | 33 | 25 |
| 12 | Wavelet-HLH\_glszm\_SizeZoneNonUniformityNormalized\_t2f | 14 | 11 |
| 13 | Wavelet-HHL\_glszm\_SmallAreaLowGrayLevelEmphasis\_t2f | 30 | 9 |
| 14 | Wavelet-LLH\_glszm\_HighGrayLevelZoneEmphasis\_t1c | 10 | 18 |
| 15 | Original\_glszm\_ZoneVariance\_t2f | 4 | 27 |
| 16 | Wavelet-LHH\_glszm\_GrayLevelVariance\_t2f | 6 | 22 |

## D. Radiomics Models Performance, Evaluation and Selection

The performance of each radiomics model (classification algorithms: LR, RF and SVM) was examined. Each model was trained using the trained dataset and the CV was set to 5. The mean accuracy scores were obtained, this can provide further insight to the performance of each model for the prediction of meningioma grades. The accuracy scores for the 3 classifiers were:

* Logistic Regression: 0.82
* Random Forest: 0.78
* Support Vector Machine: 0.78

In addition, box plots were also produced, Figure 3, to allow for the visual comparison of the accuracy scores distribution between the 3 models. The LR and RF models both present relatively similar small interquartile ranges whereas SVM has a significantly larger box size signifying a higher distribution of scores. F

These results suggest that LR has the highest mean accuracy score and therefore in currently the best performing model followed by the RF model and finally SVM due to its higher distribution although has an equal mean accuracy score to RF.

A graph of a machine learning algorithm

Description automatically generated

Figure 3 Boxplots for the Comparison of Accuracy Scores of 3 Machine Learning Algorithms.

For further evaluation of the 3 model performances, performance metrics, without CV input, was conducted and presented on table 4 to facilitate interpretation. Results showed that the LR model continued to be the best performing model with an accuracy of 0.703 and an AUC of 0.708. RF (accuracy = 0.630, AUC= 0.633) and SVM (accuracy =

0.667, AUC= 0.675) had similar performance, however, SVM was the 2nd best performing model for the prediction of meningioma grades. Other performance metrics (precision, recall and F1-score) were also calculated and can be found in table 4.

Hyperparameter tuning was conducted using a grid-search method, models were re-evaluated using the test dataset for performance metrics, results can be seen in table 4. No change to performance of the LR model was noted. Whilst the SVM model performance was enhanced with model tuning, performance now matching that of LR, a decrease in performance can be seen for the RF model. The LR model was the best preforming model for the prediction of meningioma grade and therefore was chosen as the final model for this study.

Table 4: Machine Learning Models Performance for Meningioma Grade Prediction on Test Data Before and After Hyperparameter Tuning.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Machine Learning Model | Accuracy | | Precision | | Recall | | F1-Score | | AUC | |
| **Before** | **After** | **Before** | **After** | **Before** | **After** | **Before** | **After** | **Before** | **After** |
| Logistic Regression | 0.703 | 0.703 | 0.643 | 0.643 | 0.750 | 0.750 | 0.693 | 0.693 | 0.708 | 0.708 |
| Random Forest | 0.630 | 0.593 | 0.571 | 0.538 | 0.667 | 0.583 | 0.615 | 0.560 | 0.633 | 0.592 |
| Support Vector Machine | 0.667 | 0.703 | 0.600 | 0.643 | 0.750 | 0.750 | 0.667 | 0.693 | 0.675 | 0.708 |

Finally, an ROC curve graph (figure 4) was built for the evaluation of the LR model, i.e., the best performing model. The ROC curve visualizes the true positive rate against the false positive rate for grade prediction across a threshold in the test set. This model showed an AUC score of 0.78.

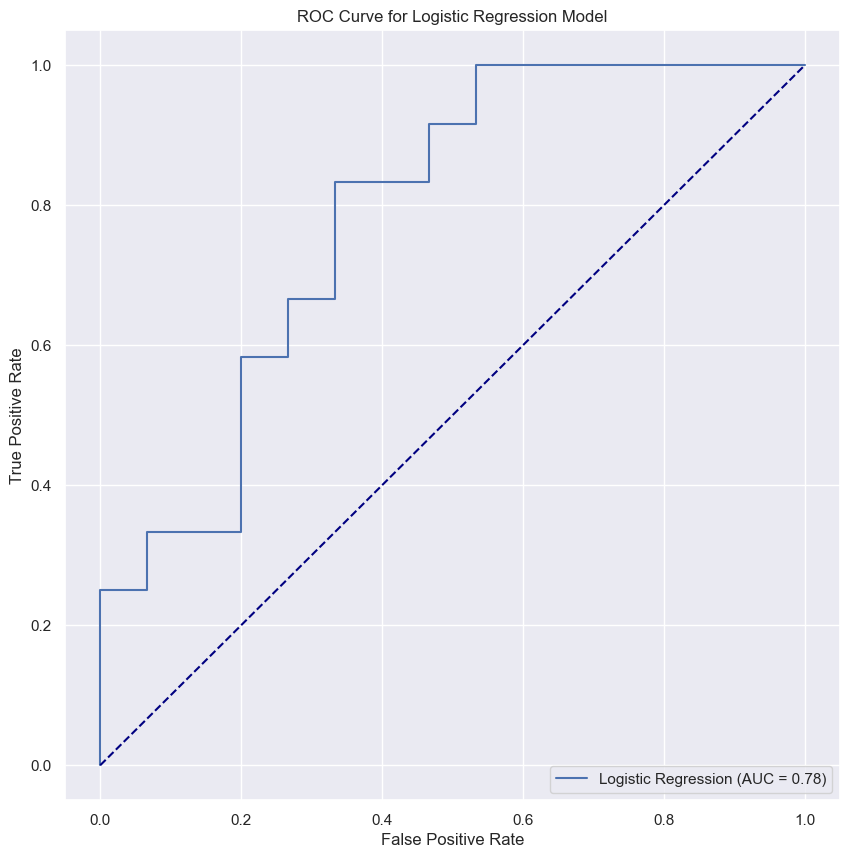


Figure 4: ROC Curve for best performing model, Logistic Regression on test set.

# DISCUSSION

In this study, as previously mentioned, three classifiers were utilized (LR, RF and SVM) and their performance was evaluated for the prediction of meningioma grades. This study assessed the performance of three models to predict meningioma grades from MRI radiomics features and show the potential for machine-learning models to assist in accurate grading of brain tumours. Performance was calculated using the accuracy, precision, recall, F1-score and AUC metrics on the test set. Overall, Logistic Regression was the best performing classifier with an accuracy of approximately 70%. Whilst after hyperparameter tuning, the SVM model had equal performance to that of LR across all metric analysis, results showed that the

distribution in this model was less consisted with a wide distribution for the accuracy metric (figure 3) therefore suggesting lower reliability. Finally, RF had the poorest performance for grade prediction with a test set AUC of 0.592 and so was not considered. However, RF has some advantages. Firstly, it has the ability to manage non-linear relationships which can be beneficial in cases where the linearity of data is unknown. RF was also considered for analysis due to its robustness to noise which therefore decreases the risk of an overfitted model. In addition, this classifier is able to handle missing data without direct input making it a prime option when assessing the predictive capabilities of ML for brain tumour grading (Auret and Aldrich, 2012). However, research has shown that the generalization error of an RF ensemble stabilizes with the increase of decision trees, tree numbers. That is, however, if all separate trees are considered diverse and accurate. This implies that an optimal number of trees can be established, and the increase of tree numbers may not significantly enhance the model performance further suggesting that a tree number increase beyond the optimal number may result in overfitting, therefore, reducing the interpretability of results (Breiman, 2001; Auret and Aldrich, 2012).

Like RF, SVM has also been identified as an attractive classfier for prediction, with multiple paper sighting its use (REFERENCE X2). Some notable advantages are that SVM is able to handle nonlinear and linear relationships and works effectively in higher-dimensional spaces, furthermore, a range of kernel types can be utilized (Saidani et al., 2023). However, the use of large datasets can result in a longer training period and SVM comes with increased interpretation difficulty of the final model. Due to the relatively small dataset used in this study, the time and computational cost of using SVM did not pose as an issue, however, SVM performance is also limited by unbalanced class sizes within datasets (Cervantes et al., 2020). In this study, as classes were close in size, class imbalance, i.e., number of grade 1 and grade 2 groups, was not resolved. It could therefore be that the solving on unbalanced data may result in the better performance of the SVM model.

Finally, LR was chosen as the best performing classifier due to performance metrics exceeding or equaling that of all models throughout this study. Due to the simplicity behind the relationship between the grade outcome and feature variables, LR has a reduced risk of overfitting. Furthermore, it has easy interpretability therefore improving transparency of results. Its simplicity also results in reduced training time and therefore reduced computational cost which may be a consideration point when selecting a final model. One study found LR to take only 8 minutes for total building time and with new entries, after building, taking seconds in the risk estimation of breast cancer. This suggests that the use of LR may be beneficial in the future if it were to be implemented in a large scale. Due to the limited number of entries in this study, computational time was not assessed as training time was minimal (Ayer et al., 2010).

This study identified LR model as a promising prediction for meningioma grade (70.3% accuracy = 70.3%, AUC = 0.708, this is supported by a previous study which observed a test set accuracy of nearly 93% and an AUC of 0.948 using LR (Chu et al., 2020). However, AUC results between table 4 and figure 4 for LR differ (0.708 and 0.78 respectively). This is likely due factors such as sample variability and differing threshold values, however, it is also possible that hyperparameter tuning may have overfit the model, further future research using different hyperparameter techniques for tuning are necessary to improve current knowledge on ML model’ predictive performance for meningioma grades. Moreover, this study used Elastic Net and RFECV for feature selection, other feature selection methods were also conducted but not utilized in the assessment of performance for ML models, further research could be conducted to identify an optimal feature selection protocol.

The current study has its limitations. Firstly, this is a retrospective study and some patients were removed thus selection bias may have occurred. Secondly, clinical data regarding the study cohort, such as sex, age and other factors which may increase heterogeneity were unknown for both groups which may affect the predictive performance of this study. Additionally, due to the lack of standard protocol across research groups, model performance may vary if repeated by different centers, moreover, model evaluation for LR presented with differing AUC results which could also suggest a reduction in reproducibility probability. Furthermore, only pre-operative MRIs/radiomics features were utilized for analysis, the combined use with clinical data and current diagnostics may prove to increase predictive performance and therefore should be further studied.

## VI. CONCLUSION

In conclusion, machine learning models using MRI radiomics features presents as an effective way to predict high and low WHO grade of meningiomas, results from this study showed promise for a high accuracy and the potential of ML models to assist in diagnosis. Further studies using a comparison of feature selection techniques on a range of classifiers could provide valuable insights and eliminate current knowledge gaps in the field.

# VII. APPENDIX

A screenshot of a graph

Description automatically generated

Figure A1: A boxplot for 4 radiomic features. 'Grade 0' signifies Grade 1 Meningioma whilst 'Grade 1' represents Grade Meningioma. Figures were chosen based on fold change analysis.

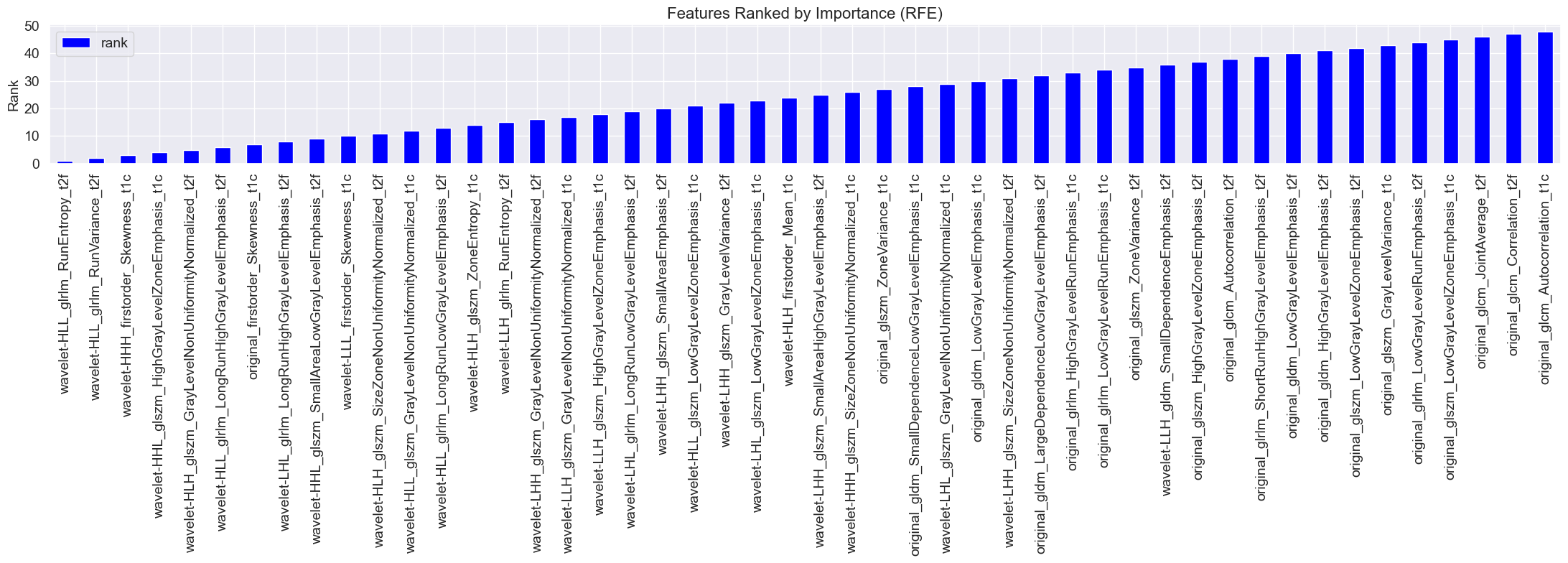


Figure A2: Ranking of all 48 features by importance using RFE.

# REFERENCES

(Wiemels, Wrensch and Claus, 2010)  
Wiemels, J., Wrensch, M. and Claus, E.B., 2010. Epidemiology and etiology of meningioma. *Journal of neuro-oncology*, [online] 99(3), pp.307–314. <https://doi.org/10.1007/s11060-010-0386-3>.

Zhao, Z., Nie, C., Zhao, L., Xiao, D., Zheng, J., Zhang, H., Yan, P., Jiang, X. and Zhao, H., 2024. Multi-parametric MRI-based machine learning model for prediction of WHO grading in patients with meningiomas. *European radiology*, [online] 34(4), pp.2468–2479. <https://doi.org/10.1007/s00330-023-10252-8>.

Yan, P.-F., Yan, L., Hu, T.-T., Xiao, D.-D., Zhang, Z., Zhao, H.-Y. and Feng, J., 2017. The potential value of preoperative MRI texture and shape analysis in grading meningiomas: A preliminary investigation. *Translational oncology*, [online] 10(4), pp.570–577. <https://doi.org/10.1016/j.tranon.2017.04.006>.

Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., Soffietti, R., von Deimling, A. and Ellison, D.W., 2021. The 2021 WHO classification of tumors of the Central Nervous System: A summary. *Neuro-oncology*, [online] 23(8), pp.1231–1251. <https://doi.org/10.1093/neuonc/noab106>.

Ugga, L., Perillo, T., Cuocolo, R., Stanzione, A., Romeo, V., Green, R., Cantoni, V. and Brunetti, A., 2021. Meningioma MRI radiomics and machine learning: systematic review, quality score assessment, and meta-analysis. *Neuroradiology*, [online] 63(8), pp.1293–1304. <https://doi.org/10.1007/s00234-021-02668-0>.

Coroller, T.P., Bi, W.L., Huynh, E., Abedalthagafi, M., Aizer, A.A., Greenwald, N.F., Parmar, C., Narayan, V., Wu, W.W., Miranda de Moura, S., Gupta, S., Beroukhim, R., Wen, P.Y., Al-Mefty, O., Dunn, I.F., Santagata, S., Alexander, B.M., Huang, R.Y. and Aerts, H.J.W.L., 2017. Radiographic prediction of meningioma grade by semantic and radiomic features. *PloS one*, [online] 12(11), p.e0187908. <https://doi.org/10.1371/journal.pone.0187908>.

Han, Y., Wang, T., Wu, P., Zhang, H., Chen, H. and Yang, C., 2021. Meningiomas: Preoperative predictive histopathological grading based on radiomics of MRI. *Magnetic resonance imaging*, [online] 77, pp.36–43. <https://doi.org/10.1016/j.mri.2020.11.009>.

Han, T., Liu, X., Long, C., Xu, Z., Geng, Y., Zhang, B., Deng, L., Jing, M. and Zhou, J., 2023. Prediction of meningioma grade by constructing a clinical radiomics model nomogram based on magnetic resonance imaging. *Magnetic resonance imaging*, [online] 104, pp.16–22. https://doi.org/10.1016/j.mri.2023.09.002.

Cuocolo, R., Caruso, M., Perillo, T., Ugga, L. and Petretta, M., 2020. Machine Learning in oncology: A clinical appraisal. *Cancer letters*, [online] 481, pp.55–62. <https://doi.org/10.1016/j.canlet.2020.03.032>.

Cuocolo, R., Perillo, T., De Rosa, E., Ugga, L. and Petretta, M., 2019. Current applications of big data and machine learning in cardiology. *Journal of geriatric cardiology: JGC*, [online] 16(8), pp.601–607. <https://doi.org/10.11909/j.issn.1671-5411.2019.08.002>.

Niu, L., Zhou, X., Duan, C., Zhao, J., Sui, Q., Liu, X. and Zhang, X., 2019. Differentiation researches on the meningioma subtypes by radiomics from contrast-enhanced magnetic resonance imaging: A preliminary study. *World neurosurgery*, [online] 126, pp.e646–e652. <https://doi.org/10.1016/j.wneu.2019.02.109>.

Gu, H., Zhang, X., di Russo, P., Zhao, X. and Xu, T., 2020. The current state of radiomics for meningiomas: Promises and challenges. *Frontiers in oncology*, [online] 10. <https://doi.org/10.3389/fonc.2020.567736>.

Li, M., Liu, L., Qi, J., Qiao, Y., Zeng, H., Jiang, W., Zhu, R., Chen, F., Huang, H. and Wu, S., 2023. MRI-based machine learning models predict the malignant biological behavior of meningioma. *BMC medical imaging*, [online] 23(1). <https://doi.org/10.1186/s12880-023-01101-7>.

Court, L.E., Fave, X., Mackin, D., Lee, J., Yang, J. and Zhang, L., 2016. Computational resources for radiomics. *Translational cancer research*, [online] 5(4), pp.340–348. <https://doi.org/10.21037/tcr.2016.06.17>.

Zhou, M., Scott, J., Chaudhury, B., Hall, L., Goldgof, D., Yeom, K.W., Iv, M., Ou, Y., Kalpathy-Cramer, J., Napel, S., Gillies, R., Gevaert, O. and Gatenby, R., 2018. Radiomics in brain tumor: Image assessment, quantitative feature descriptors, and machine-learning approaches. *AJNR. American journal of neuroradiology*, [online] 39(2), pp.208–216. <https://doi.org/10.3174/ajnr.a5391>.

Park, Y.W., Oh, J., You, S.C., Han, K., Ahn, S.S., Choi, Y.S., Chang, J.H., Kim, S.H. and Lee, S.-K., 2018. Radiomics and machine learning may accurately predict the grade and histological subtype in meningiomas using conventional and diffusion tensor imaging. *European radiology*, [online] 29(8), pp.4068–4076. <https://doi.org/10.1007/s00330-018-5830-3>.

Nowosielski, M., Galldiks, N., Iglseder, S., Kickingereder, P., von Deimling, A., Bendszus, M., Wick, W. and Sahm, F., 2017. Diagnostic challenges in meningioma. *Neuro-oncology*, [online] 19(12), pp.1588–1598. <https://doi.org/10.1093/neuonc/nox101>.

Chu, H., Lin, X., He, J., Pang, P., Fan, B., Lei, P., Guo, D. and Ye, C., 2021. Value of MRI radiomics based on enhanced T1WI images in prediction of meningiomas grade. *Academic radiology*, [online] 28(5), pp.687–693. <https://doi.org/10.1016/j.acra.2020.03.034>.

Huang, R.Y., Bi, W.L., Griffith, B., Kaufmann, T.J., la Fougère, C., Schmidt, N.O., Tonn, J.C., Vogelbaum, M.A., Wen, P.Y., Aldape, K., Nassiri, F., Zadeh, G., Dunn, I.F., Aldape, K., Au, K., Barnhartz-Sloan, J., Bi, W.L., Brastianos, P.K., Butowski, N., Carlotti, C., Cusimano, M.D., DiMeco, F., Drummond, K., Dunn, I.F., Galanis, E., Giannini, C., Goldbrunner, R., Griffith, B., Hashizume, R., Hanemann, C.O., Herold-Mende, C., Horbinski, C., Huang, R.Y., James, D., Jenkinson, M.D., Jungk, C., Kaufman, T.J., Krischek, B., Lachance, D., Lafougère, C., Lee, I., Liu, J.C., Mamatjan, Y., Mansouri, A., Mawrin, C., McDermott, M., Munoz, D., Nassiri, F., Noushmehr, H., Ng, H.-K., Perry, A., Pirouzmand, F., Poisson, L.M., Pollo, B., Raleigh, D., Sahm, F., Saladino, A., Santarius, T., Schichor, C., Schultz, D., Schmidt, N.O., Selman, W., Sloan, A., Spears, J., Snyder, J., Suppiah, S., Tabatabai, G., Tatagiba, M., Tirapelli, D., Tonn, J.C., Tsang, D., Vogelbaum, M.A., Deimling, A. von, Wen, P.Y., Walbert, T., Westphal, M., Workewych, A.M., Zadeh, G. and International Consortium on Meningiomas, 2019. Imaging and diagnostic advances for intracranial meningiomas. *Neuro-oncology*, [online] 21(Supplement\_1), pp.i44–i61. <https://doi.org/10.1093/neuonc/noy143>.

Mahmood, A., Caccamo, D.V., Tomecek, F.J. and Malik, G.M., 1993. Atypical and malignant meningiomas: A clinicopathological review. *Neurosurgery*, [online] 33(6), pp.955–963. <https://doi.org/10.1227/00006123-199312000-00001>.

Zhang, T., Yu, J.-M., Wang, Y.-Q., Yin, D.-D. and Fang, L.-J., 2018. WHO grade I meningioma subtypes: MRI features and pathological analysis. *Life sciences*, [online] 213, pp.50–56. <https://doi.org/10.1016/j.lfs.2018.08.061>.

Park, C.J., Choi, S.H., Eom, J., Byun, H.K., Ahn, S.S., Chang, J.H., Kim, S.H., Lee, S.-K., Park, Y.W. and Yoon, H.I., 2022. An interpretable radiomics model to select patients for radiotherapy after surgery for WHO grade 2 meningiomas. *Radiation oncology (London, England)*, [online] 17(1). <https://doi.org/10.1186/s13014-022-02090-7>.

Zhu, H., Fang, Q., He, H., Hu, J., Jiang, D. and Xu, K., 2019. Automatic prediction of meningioma grade image based on data amplification and improved convolutional neural network. *Computational and mathematical methods in medicine*, [online] 2019, pp.1–9. <https://doi.org/10.1155/2019/7289273>.

Chen, C., Guo, X., Wang, J., Guo, W., Ma, X. and Xu, J., 2019. The diagnostic value of radiomics-based machine learning in predicting the grade of meningiomas using conventional magnetic resonance imaging: A preliminary study. *Frontiers in oncology*, [online] 9. <https://doi.org/10.3389/fonc.2019.01338>.

Duan, C.F., Li, N., Li, Y., Liu, F., Wang, J.C., Liu, X.J. and Xu, W.J., 2022. Comparison of different radiomic models based on enhanced T1-weighted images to predict the meningioma grade. *Clinical radiology*, [online] 77(4), pp.e302–e307. <https://doi.org/10.1016/j.crad.2022.01.039>.

Vassantachart, A., Cao, Y., Shen, Z., Cheng, K., Gribble, M., Ye, J. C., Zada, G., Hurth, K., Mathew, A., Guzman, S., & Yang, W. (2023). Segmentation and Classification of Grade I and II Meningiomas from Magnetic Resonance Imaging: An Open Annotated Dataset (Meningioma-SEG-CLASS) (Version 1) [Data set]. The Cancer Imaging Archive.

Vassantachart, A., Cao, Y., Gribble, M., Guzman, S., Ye, J.C., Hurth, K., Mathew, A., Zada, G., Fan, Z., Chang, E.L. and Yang, W., 2022. Automatic differentiation of Grade I and II meningiomas on magnetic resonance image using an asymmetric convolutional neural network. *Scientific reports*, [online] 12(1). <https://doi.org/10.1038/s41598-022-07859-0>.

Auret, L. and Aldrich, C., 2012. Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals engineering*, [online] 35, pp.27–42. <https://doi.org/10.1016/j.mineng.2012.05.008>.

Breiman, L., 2001. Random Forests. *Machine learning*, [online] 45(1), pp.5–32. https://doi.org/10.1023/a:1010933404324.

Saidani, O., Aljrees, T., Umer, M., Alturki, N., Alshardan, A., Khan, S.W., Alsubai, S. and Ashraf, I., 2023. Enhancing prediction of brain tumor classification using images and numerical data features. *Diagnostics (Basel, Switzerland)*, [online] 13(15), p.2544. https://doi.org/10.3390/diagnostics13152544.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L. and Lopez, A., 2020. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, [online] 408, pp.189–215. <https://doi.org/10.1016/j.neucom.2019.10.118>.

Ayer, T., Chhatwal, J., Alagoz, O., Kahn, C.E., Jr, Woods, R.W. and Burnside, E.S., 2010. Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics: a review publication of the Radiological Society of North America, Inc*, [online] 30(1), pp.13–22. https://doi.org/10.1148/rg.301095057.