Feature Extraction from MRI ADC Images for Brain Tumor Classification Using Machine Learning Techniques

Sahan M. Vijithananda
University of Peradeniya

Mohan L. Jayatilake (jayatiml@gmail.com)
University of Peradeniya

Badra Hewavithana
University of Peradeniya

Teresa Gonçalves
University of Évora

Luis M. Rato
University of Évora

Bimali S. Weerakoon
University of Peradeniya

Tharindu D. Kalupahana
University of Sri Jayewardenepura

Anil D. Silva
National Hospital of Sri Lanka

Karuna D. Dissanayake
National Hospital of Sri Lanka

Research Article

Keywords: Magnetic Resonance Imaging, Diffusion Weighted Imaging, Apparent Diffusion Coefficient, Brain Tumor Classification, ANOVA f-test feature selection, Machine Learning, Random Forest

Posted Date: January 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1186157/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Feature Extraction from MRI ADC Images for Brain Tumor Classification Using Machine Learning Techniques

Sahan M. Vijithananda¹, Mohan L. Jayatilake²*, Badra Hewavithana¹, Teresa Gonçalves³, Luis M. Rato³, Bimali S. Weerakoon², Tharindu D. Kalupahana⁴, Anil D. Silva⁵ and Karuna D. Dissanayake⁶

Abstract

Background: Diffusion-weighted (DW) imaging is a well-recognized magnetic resonance imaging (MRI) technique that is being routinely used in brain examinations in modern clinical radiology practices. This study focuses on extracting demographic and texture features from MRI Apparent Diffusion Coefficient (ADC) images of human brain tumors, identifying the distribution patterns of each feature and applying Machine Learning (ML) techniques to differentiate malignant from benign brain tumors.

Methods: This prospective study was carried out using 1599 labeled MRI brain ADC image slices, 995 malignant, 604 benign from 195 patients who were radiologically diagnosed and histopathologically confirmed as brain tumor patients.

The demographics, mean pixel values, skewness, kurtosis, features of Grey Level Co-occurrence Matrix (GLCM), mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence and shade, were extracted from MRI ADC images of each patient.

At the feature selection phase, the validity of the extracted features were measured using ANOVA f-test. Then, these features were used as input to several Machine Learning classification algorithms and the respective models were assessed.

Results: According to the results of ANOVA f-test feature selection process, two attributes: skewness (3.34) and GLCM homogeneity (3.45) scored the lowest ANOVA f-test scores. Therefore both features were excluded in continuation of the experiment. From the different tested ML algorithms, the Random Forest classifier was chosen to build the final ML model since it presented the highest accuracy. The final model was able to predict malignant and benign neoplasms with an 90.41% accuracy after the hyper parameter tuning process.

Conclusion: This study concludes that the above mentioned features (except skewness and GLCM homogeneity) are informative to identify and differentiate malignant from benign brain tumors. Moreover, they enable the development of a high-performance ML model that has the ability to assist in the decision-making steps of brain tumor diagnosis process, prior to attempting invasive diagnostic procedures such as brain biopsies.

Keywords: Magnetic Resonance Imaging; Diffusion Weighted Imaging; Apparent Diffusion Coefficient; Brain Tumor Classification; ANOVA f-test feature selection; Machine Learning; Random Forest
Background

Brain tumors are neoplastic tissue masses in which cells multiply and grow uncontrollably without being checked by the mechanisms that control normal cell division. It can occur at any age [1] and is one of the major diseases that affects the human central nervous system. According to a study done in United States, 29.9 per 100000 adults (20 years or older) are vulnerable to have a brain tumor at any stage of their lifetime [2]. Approximately one-third of these brain tumors are malignant and the others remain as benign tumors [2, 3].

Although the computed tomography (CT), positron emission tomography (PET) medical imaging techniques are frequently involved in brain tumor diagnosis process, MRI is considered the most effective tumor imaging method due to its superior contrast properties in current radiological practices [4]. However, the noise within medical images, including MRI ones, and non-systematic search of patterns by humans (radiologists) affects the accuracy of the diagnosis. Therefore patients often need to go through invasive biopsy procedures to confirm, through histopathological analysis, the type (including its malignant or benign status) and the WHO grade of the tumour [5].

Magnetic Resonance Imaging

Diffusion Weighted (DW) imaging is a form of magnetic resonance imaging (MRI) technique that is widely used in tumor identification and classification in modern clinical radiology practices [6, 7]. This technology is based on measurements of random brownian motion of water molecules within a voxel of a biological tissue [8, 9, 10]. The technique allows to visualize the net direction of diffusion of water molecules or collective flow of water molecules in a live tissue. Hence, it has the ability to provide information on the microscopic behaviour of living biological tissues (such as the presence and permeability of membranes and the presence of macro-molecules and intracellular–extracellular water equilibrium) by measuring and imaging the transitional mobility of water molecules [11, 12, 13, 14]. Due to the characteristic features of DW images, they are appreciated as an indispensable tool for investigating central nerves system (CNS) diseases such as brain neoplasms, brain and spinal cord injuries, degenerative brain diseases, etc.

The resistance for the diffusion of water molecules inside a tissue is quantitatively assessed by calculating the apparent diffusion coefficient (ADC) values [10]. To generate an ADC map, there should be at least two types of DW images differing from each other in terms of the diffusion sensitization level (b-value). In most cases, it is common to utilize $b = 0s/mm^2$ for the lower limit and images with b-value in the range of 600 to 1000s/mm$^2$ for the upper limit [15, 16]. However, there are evidences of using a b-value greater than 1000s/mm$^2$ as the upper limit of ADC image generation [17].

Texture features

The generated ADC images reflect the magnitude of diffusion of water molecules within tissues and these images are rich in texture allowing the analysis of image in terms of these features. The texture of an image can be defined as a constant repetition of an element or pattern on the surface of an image which represents
its structure [18, 19]. Texture analysis focuses on finding a specific way of representing the hidden characteristics of textures and express them in a simplified and unique form. Grey level co-occurrence matrices (GLCM) of MRI ADC images can be identified as a rich source of statistical texture features which can be utilized in training robust machine learning (ML) models, which is a powerful method that is commonly utilize to identify the unique patterns of the distribution of texture features within an image [20] [20, 21, 22].

GLCM texture feature extraction can be defined as a statistical method that reveals specific properties about the spatial distribution of gray levels in image texture considering the spatial relationship of pixels [23]. Here measures the relation of grey intensities between two adjacent pixels (reference pixel (i), neighbor pixel (j)) of an image at a time to have information about variation in intensity at a pixel of interest. The GLCM matrices are computed using 2 parameters such as the relative distance between the pixel pair and the relative orientation (angular relationship) of the pixel pair. Most frequently, the orientation quantified as 0°, 45°, 90° and 135° angles and the average of the resultant values for all four directions used to extract the features [24, 25]

**Higher order moments**

Higher order moments can be identified as functions that use high power of a sample (higher than 2nd order statistics), that is opposed to the conventional first or second order statistics (lower order statistics). The higher order statistic provides powerful tools in identifying problems in non linear systems [26]. However, skewness and kurtosis are the examples of third order and fourth order statistics respectively [8, 27]. Here, skewness measures the asymmetry around the mean of probability distribution of a real valued random variable and the values for skewness can be zero (0), positive (+), negative (−) or undefined. The kurtosis use to describe the shape of a probability distribution of a real valued random variable and measures the tailedness of it. The kurtosis values for any uni-variate normal distributions remain as 3. However, the distributions with kurtosis values more than 3 are considered as platykurtic distributions while considering the distributions with kurtosis values less than 3 as leptokurtic distributions [28]

**Supervised Machine Learning**

Machine Learning is a scientific area that allows computers to "learn" from data. The algorithms are used to find out natural patterns in data aiming to aid and/or support decisions and predictions. Considering the goal and the nature of data, ML methods can be further classified as supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, transduction and learning to learn [29]. Among the above mentioned machine learning techniques, the supervised learning uses known input and output data to train a model [30, 31, 32] to solve classification problems [29, 33] and is one of the most common ML paradigms.

The above mentioned texture features of MRI brain can be extracted from ADC images of brain neoplasms and used by machine learning techniques to train classifiers. The process can be described in a few basic steps as illustrated in Figure 1).
Figure 1 Supervised learning flow chart. The flow chart illustrates the steps of building a classification model to differentiate brain neoplasms using supervised learning technique. According to the study plan, the problem was identified as a classification problem at the initial stage and then the necessary data was collected as the second step. Data pre-processing was executed as the third step and at the fourth step, the data set was split into training and testing sets. Then a suitable ML algorithm for the collected data was selected as the fifth step of the study flow and then, the selected algorithm was trained with the training data as the sixth step. Finally, the developed algorithm was evaluated with the test data and the hyperparameter of the developed model was tuned to reach the optimum accuracy level of the model.

Objectives
Using the hidden characteristics of image textures of higher order moments of ADC (skewness and kurtosis) [8, 21], GLCM statistical features [1, 34, 35] and patients’ demographic data, this work proposes a Machine Learning classification model that can be used to differentiate benign and malignant brain tumors.

Methods
This prospective study was designed to address the above mentioned objectives of the study and hypothesized that there is a correlation between the extracted features and the benign and malignant status of the tumors. According to the nature of the features extracted, the study plan was designed and the Figure 1 summarises the supervised learning process which was used to develop a robust automated technique to discriminate malignant from benign brain tumors.

Data Acquisition and Preparation
This study includes 1599 MRI brain image slices from 195 patients of both sexes (53.41% male and 46.59% female) and all the subjects were within the 12 to 80 year age range with an average of 45.51 years. The MRI Digital Imaging and Communications in Medicine (DICOM) data of each subject was acquired after confirming the pathological condition by referring both radiological and histopathological reports of each patient. All patients data was obtained from the Departments of Radiology and Histopathology, National Hospital of Sri Lanka (NHSL) and Anuradhapura Teaching Hospital, Sri Lanka, followed by the informed consent of the patients and
ethical clearance certificate from the institutional ethical review committee of NHSL and the Faculty of Medicine, University of Peradeniya.

From the initial set of 1896 ADC image slides, 297 were removed according to the exclusion criteria such as lack of information, corrupted MRI images, and the selected tumor not within the considered area (Brain). Therefore, the study was conducted with the remaining 1599 image slices which consisted with 62.22% malignant tumors and 37.77% benign brain tumor slices (See Table 1).

All scans in this study were performed with a 3T Siemens Skyra MR system using head coil and utilized the EPI (Echo Planner Imaging) sequence to acquire axial DW MRI data in both \( b = 0 \) and \( b = 1000 \) diffusion sensitization levels with a \( \text{flip}_\text{angle} = 90^\circ, \text{TE} = 68ms \) and \( TR = 4300ms \) (being \( TE \) the time of echo and \( TR \) the time of repetition), \( FOV = 219mm \times 219mm, \text{matrix}_\text{size} = 124 \times 124 \) and \( \text{slice}_\text{thickness} = 1mm \).

**ADC Image Generation and ROI Selection**

The ADC images were generated by merging two different DW images with different diffusion sensitization levels \( (b = 0 \) and \( b = 1000 \)) according to the Equation 1, where \( i \) is the image number, \( S_i \) the \( i^{th} \) image (image acquired with a diffusion pulse of \( i \)), \( S_0 \) the first image (image acquired without any diffusion pulses), \( n \) the number of images and \( b_i \) the diffusion gradient value.

\[
ADC = \sum_{i=1}^{n} \ln \frac{S_i}{S_0} \frac{1}{b_i} \tag{1}
\]

The tumor area of each 2D ADC image slice was selected by drawing a 3D region of interest (ROI) encompassing the tumor (see Figure 2) and extracted the pixel values within the selected area.

![Figure 2](image.png) MRI ADC brain image of a 14 years old female patient diagnosed with pilocytic astrocytoma which was radiologically and histo-pathologically identified as a benign tumor. The tumor area is surrounded by the ROI. The texture features were extracted from the selected area.
Feature Extraction

We have evaluated the mean, higher order moments skewness \((n = 3)\) and kurtosis \((n = 4)\) and GLCM based statistical texture features of MRI-ADC brain tumors and the patients demographics. The mean pixel value and the higher order moment values were calculated within the ROI using Equations 2 and 3, respectively. Here, \(P_i\) represents the signal intensity in \(i^{th}\) pixel and \(N\) is the total number of pixels within the ROI, \(P\) is the mean of the pixel values and \(f(P_i)\) the probability of the signal intensity of pixel.

\[
Mean = \frac{\sum_{i=1}^{N} P_i}{N} \quad (2)
\]

\[
n^{th} \text{ moment} = \sum_i (P_i - P)^nf(P_i) \quad (3)
\]

MATLAB 2019 Simulink software was used in all the image processing steps and Python 3.7 in all the feature extraction and analysis processes [8]. The GLCM matrices of each 2D parametric map of ADC brain tumor were derived according to the Equation 4 (see Appendix). The statistical texture features of GLCM (mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence and shade values) were extracted from the generated GLCM matrices. Moreover, the GLCM features were extracted according to the Equations 5 to 13 (see Appendix) [24, 25, 36].

\[
M_{f,\delta}(k, l) = \sum_{x,y=1}^{n} \begin{cases} 
1 & \text{if } f(x, y) = k \text{ and } f(x + \delta_x, y + \delta_y) = y \\
0 & \text{otherwise}
\end{cases} \quad (4)
\]

Here the \(f\) consider as 2D parametric ADC map, \(M_{f,\delta}(k, l)\) is the co-occurrence matrix which represents the joint probability occurrence of pixel pairs with grey level value \(k\) and \(l\) for, \(\delta = (\delta_x, \delta_y)\) specific spatial offset between the pixel pair. \(n\) is the bar of grey levels in 2D parametric ADC map of brain tumor.

However, lower and higher order moments (see Equation 3); mean pixel value \((n = 1)\), skewness \((n = 3)\), kurtosis \((n = 4)\) and texture features of GLCM such as mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence and shade values were studied in this pattern recognition process; The GLCM features were extracted according to Equations 5, 6, 7, 8, 9, 10, 11, 12 and Equation 13 respectively [24] [25] [36]. Here \(P_{i,j}\) be the co-occurrence matrix, \(N\) be the number of grey levels in the image, \(\mu\) be the mean of \(P_{i,j}\), \(\mu_i\) be the mean of row \(i\), \(\mu_j\) be the mean value of column \(j\), \(\sigma_i\) be the standard deviation of row \(i\) and \(\sigma_j\) be the standard deviation of column \(j\). The extracted feature values were stored in a CSV file for data preparation and further analysis.
**GLCM mean**
Left sided equation calculates the mean based on the reference pixel \((\mu_i)\) while right sided equation calculates the mean using neighbouring pixels \((\mu_j)\)

\[
\mu_i = \sum_{i,j=0}^{N-1} i \times (P_{i,j}) \\
\mu_j = \sum_{i,j=0}^{N-1} j \times (P_{i,j})
\] (5)

**GLCM Variance**
Left sided equation calculates the dispersion of the reference pixel values \((\sigma_i^2)\) around \((\mu_i)\) and the right sided equation calculates the dispersion of the neighbour pixel values \((\sigma_j^2)\) around \((\mu_j)\)

\[
\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j} \times (i - \mu_i)^2 \\
\sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j} \times (j - \mu_j)^2
\] (6)

**GLCM Energy (En)**
Energy expresses the uniformity of the texture (within a scale between 0 to 1) by measuring the sum of squared elements in the GLCM. GLCM energy value is 1 when the texture is uniform.

\[
En = \sum_{i,j=0}^{N-1} P_{i,j}^2
\] (7)

**Entropy (Etr)**
Entropy describes the degree of disorder among pixels within the matrix, which is approximately inversely correlated with uniformity. The Larger the number of grey levels within the image express larger entropy values.

\[
Etr = \sum_{i,j=0}^{N-1} P_{i,j} \times (-\ln P_{i,j})
\] (8)

**GLCM Contrast (Con)**
GLCM Contrast expresses the amount of local gray level variation in an image. Presence of edges, noise, or wrinkled textures within an image returns high contrast values.

\[
Con = \sum_{i,j=0}^{N-1} P_{i,j} \times (i - j)^2
\] (9)

**Homogeneity (Hom)**
Homogeneity expresses the smoothness of the distribution of gray levels within an image, which is approximately, inversely correlated with contrast.

\[
Hom = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2}
\] (10)
Correlation (Cor)
Correlation expresses the amount of linear dependency of gray levels among two neighboring pixels within the matrix. Texture with high GLCM correlation has high predictability of pixel relationships.

\[
\text{Cor} = \sum_{i,j=0}^{N-1} P_{i,j} \left[ \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \tag{11}
\]

Cluster Shade (Shd)
Cluster Shade evaluates the tendency of clustering of the pixels by measuring the skewness of pixel values within the matrix.

\[
\text{Shd} = \sum_{i,j=0}^{N-1} \{i + j - \mu_i - \mu_j\}^3 P_{i,j} \tag{12}
\]

Cluster Prominence (Prom)
Cluster Prominence measures local intensity variation of pixels and the asymmetry of an image. The high prominence value indicates less symmetry of an image while image with less cluster prominence value shows peak in GLCM matrix around the mean.

\[
\text{Prom} = \sum_{i,j=0}^{N-1} \{i + j - \mu_i - \mu_j\}^4 P_{i,j} \tag{13}
\]

Feature Selection and Model Training
The derived features from each image slice (see Section Feature extraction) were split into train and test sets with 70%:30% proportions respectively as 1119 image slices for training and 480 image slices for testing data. The best set of input variables in terms of increasing the prediction accuracy of the classification model and reducing the computational cost to train a machine learning model was selected through an ANOVA f-test feature selection method.

The extracted features were normalized to have zero mean and unit variance. A subset of the input normalized features most relevant to the target variable (malignant or benign) was selected using the ANOVA (Analysis of Variance) f-test feature selection technique. The first 14 attributes with the highest scores at the ANOVA f-test (see Table 2) were selected to proceed the K-fold cross-validation experiment (see Figure 3). To reduce the effect of the stochastic nature of the algorithm and differences of numerical precision, the process was repeated 5 times and the mean score for each attribute was calculated (see Table 2).

A 10-Folds cross-validation method was utilized to figure out the most promising algorithm that discriminates the malignant and benign brain tumors with highest prediction accuracy. Here, the cross validation scores for several common classification algorithms such as Logistic Regression, Linear Discriminant Analysis, K Nearest Neighbor (KNN), Decision Tree Classifier, Gaussian Naive Bayes (NB), Support Vector Classifier (SVC) and Random Forest Classifier were calculated. However the
Random Forest Classifier algorithm was selected to build the ML model as it perform the highest cross validation score at the experiment (see Table 3). Python 3.7 along with scikit-learn library were used to normalize the data, apply the ANOVA f-test and build and assess the classification models [37].

Parameter Tuning and Prediction

Then the selected normalized features were fed into the Random Forest Classifier to develop a tumor classification model and the performance was evaluated for different parameters of the algorithm. The accuracy, Precision, Recall and F1 measures obtained with the set of parameters that produced the best model is presented on Table 4.

To search for the best set of parameters, a random grid search was performed aiming to increase the tumor prediction accuracy and a decision threshold adjustment was done to optimize the sensitivity and specificity of the classification model [38]. The considered tunable hyperparameters of the algorithm were: min_samples_split,
n_estimators, max_depth and max_features and each hyper parameter was tested within a pre-defined ranges of values (min_samples_split : [2, 5, 10], n_estimators : from 200 to 1000 (with step of 10), max_depth: 10 to 100 (with step of 10), and max_features : [3, 5, 10, 20]). Here the optimum values for each hyperparameters that maximize the precision and recall of the developed classification model were measured separately (See Table 7).

The decision threshold (the operating point) of the developed ML model was adjusted to improve either sensitivity or specificity. It was adjusted with the guidance of the precision-recall curve shown in Figure 4 and the receiver operating characteristic (ROC) curve presented on Figure 5. Finally, the performance of the tuned classification model was assessed by observing accuracy score, the precision, recall and F1 scores over the test set.

The precision and recall curve (see Figure 4) utilized to visualize sensitivity and specificity trade-off in the classifier. With the assistance of the information from precision and recall curve, the decision threshold was adjusted and set to 0.45 which maximize the sensitivity and specificity of the developed ML model. The ROC curve (see Figure 5) was implemented visualize the performance of the ML model at all classification thresholds. As a result of adjusting the decision threshold and the hyperparameter tuning process the developed ML model was able gained the ability of predicting malignant and benign brain tumors with 90.41% average accuracy level.

**Results**

According to the Table: 2 skewness (3.3444) and GLCM Homogeneity (3.4572) reported the minimum scores at the ANOVA f-test feature selection experiment while the feature "patient gender" reporting the highest (73.7926) score (See Figure 3).

As a result of 10-fold cross-validation experiment for the training and testing data sets, the Random Forest Classifier expressed the highest accuracy level (84.36%)
while Logistic Regression, Linear Discriminant Analysis, k-nearest neighbors Classifier, GaussianNB and SVC obtained 75.33%, 74.89%, 82.84%, 80.07%, 74.89%, and 81.50% accuracy levels respectively (See Table 3). Hence, the study was continued through the Random Forest Classifier to build most accurate ML model to differentiate malignant and benign brain tumor MRI-ADC images.

As a result of training the Random Forest Classifier, the generated model acquired the ability to predict the tumor type (malignancy and benign status) with 85% accuracy level. The model performance over the test data is presented on Table 4 and revealing that the ML model was able to identify the malignant tumors with 85% Precision, 92% Recall which corresponds to a 89% F1-score. Moreover, a performance of 85% for Precision and 73% for Recall (which corresponds to a F1-score of 79%) was obtained for the benign tumors.

The best hyper-parameters for maximum Precision were $max\_\text{depth} = 70$, $max\_\text{features} = 10$, $min\_\text{samples\_split} = 2$ and $n\_\text{estimators} = 500$, and the ML model predicted 229 malignant and 117 benign tumors correctly with 34 false negatives and 20 false positives from the test data (see Table 5). The best parameters which returned the maximum recall score was $max\_\text{depth} : 30, max\_\text{features} : 10, min\_\text{samples\_split} : 2, n\_\text{estimators} : 300$ (see Table 7) and 231 malignant tumors, 119 benign tumors were accurately predicted while expressing 32 fails negatives and 18 fails positives (see Table 6).

Finally the decision threshold was adjusted to 0.45 with the assistance of the information provided by the precision and recall curve (see Figure 4). As a result, the accuracy score of the optimized ML classification model increased up to 90.41% while the precision, recall and f1 score for predicting malignant tumors maintaining at 92.02%, 92.64% and 92.33% respectively. Also for the benign tumor prediction, the precision, recall and f1 scores reported as 87.71%, 86.74% and 87.22% (see Figure 6).

**Discussion**

The excellent soft tissue differentiation ability of MRI allows to visualize the exact location of the tumor, and aids to therapeutic, diagnosis and evaluation process of human brain tumors [30, 39, 40]. Moreover the modern MRI techniques used in clinical setup such as DWI, DTI and DSCI utilize to assist in tumor characterization and treatment process [41]. This study focused on developing an automated method using currently available MRI technique (DWI) to differentiate malignant and benign brain tumors. However, developing an accurate, automated and noninvasive method to differentiate benign and malignant brain tumors leads to increase the accuracy of the diagnosis process in terms of sensitivity and the specificity. Since such a tool has the ability to assist the clinician in decision-making at the brain tumor diagnosis process, it can be applied as an intermediate step in between tumor imaging and the brain biopsies which allows the clinician to decide the necessity of requesting a biopsy for further investigation.

To achieve these goals, a ML model was developed using extracted features from DWI images and the demographics of patients. The mean pixel value, skewness, kurtosis, GLCM features (mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence and shade values) and demographic features (age and gender) were extracted using BLeDIA, an home made software which was specifically
designed for this study. We hypothesized that the above mentioned features correlate with malignant and benign brain tumors in complex and non-linear ways. Therefore, the combination of all features except skewness and GLCM homogeneity were used to develop a machine learning model able to distinguish the malignant and benign status of brain tumors. As a byproduct of ANOVA f-test feature selection process, we have observed that the feature "patient gender" showed the highest ANOVA f-test score (73.7926) (See Table 2). Therefore, it is possible to assume that the patient’s gender has high impact on predicting benign and malignant brain tumors. Such information can be studied as an extension of this study.

The Selected normalized features were used on several classification algorithms to find out the best fit algorithm for the data set. According to performance obtained (presented on Table 3) the Random Forest classifier showed the highest score being then selected to develop a ML model.

The Random Forest Classifier could predict the malignant and benign brain tumors with 85% accuracy level (see Table 4). However, the accurate interpretation is not straightforward due to the numeric nature of the extracted features. Therefore, hyper parameter tuning and the decision threshold adjustment was utilized to increase the overall accuracy level of the ML model in terms of sensitivity and specificity. As visualized in Table 7 the optimized Hyper parameter values for precision and the recall was measured. However, according to precision recall curve (see Figure 4) and the ROC curve (see Figure 5) the decision threshold value was adjusted as the ML model returns the optimum precision and recall values. As result of tuning the ML model in two steps, finally the ML model able to predict the malignant and benign brain tumors with 90.41% accuracy level with high recall score (92.64%) for malignant tumor identification which indicates that there is less probability to not detect malignant tumors (see Figure 6).
Conclusion
The study concludes that mean ADC, kurtosis of ADC and the GLCM features of ADC (mean, variance, energy, entropy, contrast, correlation, prominence and shade) and demographics features can be used as potential bio-markers to identify and differentiate benign and malignant brain tumors.

Given the findings just presented, one can say that this study reveals that there is a great potential on using the developed ML mode in clinical practices to differentiate benign and malignant brain tumors. The results of this study encourage to develop an advanced ML model to predict WHO grading of brain tumors and specifically identifying brain tumors.

Appendix
Acknowledgements
I wish to acknowledge the support provided by the University of Peradeniya (Sri Lanka), University of Évora (Portugal), the ERASMUS Plus program, the National Hospital of Sri Lanka, Anuradhapura Teaching Hospital, and all the staff members including consultants of the departments of radiology and histopathology in the above-mentioned hospitals in providing the opportunity, knowledge, resources and immense support needed to accomplish the study.

Funding
Not Applicable...

Abbreviations
MRI: Magnetic Resonance Imaging; DW: Diffusion Weighted; DWI: Diffusion Weighted Imaging; CNS: Central Nervous System; ADC: Apparent Diffusion Coefficient; GLCM: Grey Level Co-occurrence Matrix; ML: Machine Learning; ROC: Receiver Operating Characteristic curve

Availability of data and materials
Text for this section...

Ethics approval and consent to participate
The study was approved by two institutional ethics review committees.
1. The ethics review committee of the Faculty of Medicine, University of Peradeniya under 2019/EC/50 reference number.
2. The ethics review committee of the National Hospital of Sri Lanka, Colombo 10, under ETH/COM/2019/AUGUST/05 reference number.
...

Competing interests
The authors declare that they have no competing interests...

Consent for publication
I, as the corresponding author, give my consent for the publication of identifiable details, which can include images, tables and details within the manuscript to be published in the BMC Medical Imaging journal...

Authors' contributions
Conceptualization: Sahan M. Vijithananda, Mohan L. Jayatilake, Badra Hewavithana, Teresa Gonçalves, Luis M. Rato, Bimali S. Weerakoon
Design of the work: Sahan M. Vijithananda, Mohan L. Jayatilake, Teresa Gonçalves
Data curation: Sahan M. Vijithananda, Badra Hewavithana, Karuna D. Dissanayake, Anil D. Silva
Formal analysis: Sahan M. Vijithananda, Teresa Gonçalves, Mohan L. Jayatilake, Luis M. Rato, Tharindu D. Kalupahana
Interpretation of data: Sahan M. Vijithananda, Teresa Gonçalves, Luis M. Rato
creation of new software used in the work: Sahan M. Vijithananda, Tharindu D. Kalupahana
Writing – original draft: Sahan M. Vijithananda, Mohan L. Jayatilake
Writing – review editing: Sahan M. Vijithananda, Mohan L. Jayatilake, Teresa Gonçalves, Badra Hewavithana, Luis M. Rato...

Authors’ information
Sahan M. Vijithananda: Lecturer (Temporary), Department of anatomy, Faculty of Medicine and Allied Sciences, Rajarata University, Sri Lanka
Mohan L. Jayatilake: Senior Lecturer, Department of Radiography and Radiotherapy, Faculty of Allied Health Sciences, University of Peradeniya.
Badra Hewavithana: Senior Professor in Radiology, Department of Radiology, Faculty of Medicine, University of Peradeniya.
Teresa Gonçalves: Associate professor, Department of Informatics, University of Evora, Evora, Portugal.
References

1. Jafarpour, S., Sedghi, Z., Amirani, M.C.: A robust brain mri classification with glcm features. International Journal of Computer Applications 37(12), 1–5 (2012)

2. Ostrom, Q.T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., Barnholtz-Sloan, J.S.: Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011–2015. Neuro-oncology 20(suppl 4), 1–86 (2018)

3. Kohler, B.A., Ward, E., McCarthy, B.J., Ries, L.A., Schymura, M.J., Rees, L.A., Eheman, C., Jemal, A., Anderson, R.N., Amin, M.B., Edwards, B.K.: Annual report to the nation on the status of cancer, 1975–2017, featuring tumors of the brain and other nervous system. Journal of the national cancer institute 103(9), 714–736 (2011)

4. Arakeri, M.P., Reddy, G.R.M.: Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. Signal, Image and Video Processing 9(2), 409–425 (2015)

5. Desroches, J., Jermyn, M., Pinto, M., Picot, F., Tremblay, M.-A., Obaid, S., Marple, E., Urmey, K., Trudel, D., Soulez, G., et al.: A new method using raman spectroscopy for in vivo targeted brain cancer tissue biopsy. Scientific reports 8(1), 1–10 (2018)

6. Maier, S.E., Sun, Y., Muliken, R.V.: Diffusion imaging of brain tumors. NMR in biomedicine 23(7), 849–864 (2010)

7. Kono, K., Inoue, Y., Nakayama, K., Shakudo, M., Morino, M., Ohata, K., Wakasa, K., Yamada, R.: The role of diffusion-weighted imaging in patients with brain tumors. American journal of neuroradiology 22(6), 1081–1088 (2001)

8. Vijithananda, S.M., Jayatilake, M.L., Weerakoon, B.S., Wathsala, P., Thevapriya, S., Thasanky, S., Kalupahana, T.D., Wijerathne, T.K.: Skewness and kurtosis of apparent diffusion coefficient in human brain lesions to distinguish benign and malignant using mri. In: International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 189–199 (2018). Springer

9. Rajeshkannan, R., Moorthy, S., Sreekumar, K., Rupa, R., Prabhu, N.: Clinical applications of diffusion weighted mri imaging: A review. Indian Journal of Radiology and Imaging 16(4), 705 (2006)

10. Balyan, V., Das, C.J., Sharma, R., Gupta, A.K.: Diffusion weighted imaging: technique and applications. World journal of radiology 8(9), 785 (2016)

11. Montagne, A., Nation, D.A., Pa, J., Sweeney, M.D., Toga, A.W., Zlokovic, B.V.: Brain imaging of the liver: usefulness of adc values in the differential diagnosis of focal lesions and effect of roi methods on adc measurements. Magnetic Resonance Materials in Physics, Biology and Medicine 26(3), 303–312 (2013)

12. Di Trani, M.G., Manganaro, L., Antonelli, A., Guerri, M., De Feo, R., Catalano, C., Capuani, S.: Apparent diffusion coefficient assessment of brain development in normal fetuses and ventriculomegaly. Front. Phys. 7: 160. doi: 10.3389/fphy (2019)

13. Thörmer, G., Otto, J., Reis-Zimmermann, M., Seiwerts, M., Moehe, M., Garnov, N., Franz, T., Do, M., Stolzenburg, J.-U., Horn, L.-C., et al.: Diagnostic value of adc in patients with prostate cancer: influence of the choice of b values. European radiology 22(8), 1820–1828 (2012)

14. Sener, R.: Diffusion mri: apparent diffusion coefficient (adc) values in the normal brain and a classification of brain disorders based on adc values. Computerized medical imaging and graphics 25(4), 299–326 (2001)

15. Kim, C.K., Park, B.K., Lee, H.M., Kwon, G.Y.: Value of diffusion-weighted imaging for the prediction of prostate cancer location at 3t using a phased-array coil: preliminary results. Investigative radiology 42(12), 842–847 (2007)

16. Srinivasan, G., Shobha, G.: Statistical texture analysis. In: Proceedings of World Academy of Science, Engineering and Technology, vol. 36, pp. 1264–1269 (2008)

17. Lerski, R.A., Straughan, K., Schad, L., Boyce, D., Blüml, S., Zuna, I.: Vii. mr image texture analysis—an approach to tissue characterization. Magnetic resonance imaging 11(6), 873–887 (1993)

18. Ali, K.H., Talal, E.B., Alsaad, N.J.: Texture features analysis using gray level co-occurrence matrix for a spine mri images. International Journal of Computer Science and Information Security 14(9), 666 (2016)

19. Mohanaiah, P., Sathyarayana, P., Gurusubramanian, L.: Image texture feature extraction using glcm approach. International journal of scientific and research publications 3(5), 1 (2013)

20. Bhatti, P., Choudhary, S., Singh, K.M.: A comparative study for brain tumor detection in mri images using texture features. In: Sensors for Health Monitoring, pp. 259–287. Elsevier, ??? (2019)
23. Zulpe, N., Pawar, V.: Glcm textural features for brain tumor classification. International Journal of Computer Science Issues (IJCSI) 9(3), 354 (2012)
24. PS, S.K., VS, D.: Extraction of texture features using glcm and shape features using connected regions. International Journal of Engineering and Technology 8(6), 2926–2930 (2016)
25. Yang, X., Tridandapani, S., Beitzler, J.J., Yu, D.S., Yoshida, E.J., Curran, W.J., Liu, T.: Ultrasound glcm texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. Medical physics 39(9), 5732–5739 (2012)
26. Emara-Shabaik, H.E.: Nonlinear systems modeling identification using higher order statistics/polyspectra. In: Leonides, C.T. (ed.) Stochastic Digital Control System Techniques. Control and Dynamic Systems, vol. 76, pp. 289–322. Academic Press, ??? (1996). doi:10.1016/S0090-5267(96)80020-X. https://www.sciencedirect.com/science/article/pii/S009052679680020X
27. Dean, S., Illowsky, B.: Descriptive statistics: skewness and the mean, median, and mode. Connexions website (2018)
28. Joanes, D.N., Gill, C.A.: Comparing measures of sample skewness and kurtosis. Journal of the Royal Statistical Society: Series D (The Statistician) 47(1), 183–189 (1998)
29. Ayodele, T.O.: Types of machine learning algorithms. New advances in machine learning 3, 19–48 (2010)
30. Sharma, K., Kaur, A., Gujral, S.: Brain tumor detection based on machine learning algorithms. International Journal of Computer Applications 103(1), 7–11 (2014)
31. Juntu, J., Sijbers, J., De Backer, S., Rajan, J., Van Dyck, D.: Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in t1-mri images. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine 31(3), 680–689 (2010)
32. Zacharaki, E.I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E.R., Davatzikos, C.: Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine 62(6), 1609–1618 (2009)
33. Mohammed, M., Khan, M.B., Bashier, E.B.M.: Machine Learning: Algorithms and Applications. Crc Press, ??? (2016)
34. Aggarwal, N., Agrawal, R.: First and second order statistics features for classification of magnetic resonance brain images (2012)
35. Jayati, M., Gonçalves, T., Rato, L.: Classification of breast lesions using quantitative dynamic contrast enhanced-mri. In: International Symposium Computational Modeling of Objects Represented in Images, pp. 108–119 (2018). Springer
36. Hall-Beyer, M.: Glcm texture: a tutorial v. 3.0 march 2017 (2017)
37. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. Journal of machine learning research 12(Oct), 2825–2830 (2011)
38. Chen, J., Tsai, C.-A., Moon, H., Ahn, H., Young, J., Chen, C.-H.: Decision threshold adjustment in class prediction. SAR and QSAR in Environmental Research 17(3), 337–352 (2006)
39. Thamaraielv, B., Yamuna, G.: Gray level co-occurrence matrix features based classification of tumor in medical images. J. Eng. Appl. Sci 11, 11403–11414 (2016)
40. Provenzale, J.M., Mukundan, S., Barborkia, D.P.: Diffusion-weighted and perfusion mr imaging for brain tumor characterization and assessment of treatment response. Radiology 239(3), 632–649 (2006)
41. Lee, S.-K.: Diffusion perfusion tensor imaging and of brain tumors in high-field mr imaging. Neuroradiology Applications of High-Field MR Imaging, An Issue of Neuroimaging Clinics-E-Book 22(2), 123 (2012)

Tables
Additional Files
Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.
Table 1 Tumor types and percentages belonging to each benign and malignant categories

According to the radiological and histopathological reports, there were 995 malignant brain image slices, including WHO (World Health Organization) Grade IV tumors; 442 glioblastomas, 109 medulloblastoma, 170 metasasis/residual malignancies, and WHO Grade III tumors; 147 high grade gliomas, 22 anaplastic astrocytomas, 11 anaplastic meningioma, 29 anaplastic oligodendro glioma, 65 central astrocytomas within the population. Also there were 604 benign brain tumors slices with WHO Grade I; 13 pilocytic astrocytoma, 262 meningioma, 135 schwannoma, 16 hemangioblastoma, 11 cranoephrangygioma, 13 Dermoid cysts, and WHO Grade II; 10 astrocytoma, 21 meningiomas, 112 low grade gliomas, 7 epedymomas, 4 frontal cavernoma

| Category   | WHO Grading | Tumor Type                  | image slices | Percentage |
|------------|-------------|-----------------------------|--------------|------------|
| Benign     | WHO I       | Meningioma                 | 262          | 43.38%     |
|            |             | Schwannoma                 | 135          | 22.35%     |
|            |             | Pilocytic astrocytoma      | 13           | 2.15%      |
|            |             | Hemangioblastoma           | 16           | 2.65%      |
|            |             | Cranepharyngioma           | 11           | 1.82%      |
|            |             | Dermoid cyst               | 13           | 2.15%      |
|            | WHO II      | Low grade gliomas          | 112          | 18.54%     |
|            |             | Meningioma                 | 21           | 3.48%      |
|            |             | Astrocytoma                | 10           | 1.67%      |
|            |             | Ependymoma                 | 7            | 1.16%      |
|            |             | Frontal cavernoma          | 4            | 0.66%      |
| Malignant  | WHO III     | High grade gliomas         | 147          | 14.77%     |
|            |             | anaplastic astrocytomas    | 22           | 2.21%      |
|            |             | anaplastic meningioma      | 11           | 1.10%      |
|            |             | anaplastic oligodendro glioma | 29     | 2.91%      |
|            |             | central astrocytomas       | 65           | 6.53%      |
|            | WHO IV      | Glioblastomas              | 442          | 44.42%     |
|            |             | Medulloblastoma            | 109          | 10.95%     |
|            |             | Metasasis                  | 170          | 17.08%     |

Table 2 ANOVA f-test Feature Selection. The table visualize the performance of each feature at the ANOVA f-test Feature Selection process. The data set went through the ANOVA f-test Feature Selection process for 5 times and the mean values were calculated. There were slight differences of values received at each time due to stochastic nature of the algorithm, or differences in numerical precision or evaluation procedure.

| Feature                      | ANOVA f-test Score (mean value) |
|------------------------------|---------------------------------|
| Mean Pixel Value of ADC      | 32.3343                         |
| Skewness                     | 3.3444                          |
| Kurtosis                     | 9.6250                          |
| GLCM Mean 1                  | 32.6372                         |
| GLCM Mean 2                  | 29.1327                         |
| GLCM Var 1                   | 14.0761                         |
| GLCM Var 2                   | 27.5219                         |
| GLCM Energy                  | 33.9675                         |
| GLCM Entropy                 | 4.989                           |
| GLCM Contrast                | 47.9462                         |
| GLCM Homogeneity             | 3.4572                          |
| GLCM Correlation             | 48.6392                         |
| GLCM Prominence              | 15.4134                         |
| GLCM Shade                   | 17.1677                         |
| Patient Age                  | 9.4337                          |
| Patient Gender               | 73.7926                         |

Table 3 Results of the cross validation experiment.

| Algorithm                    | Mean accuracy | Standard deviation[SD] |
|------------------------------|---------------|------------------------|
| Logistic Regression          | 0.753378      | 0.034451               |
| Linear Discriminant Analysis| 0.748898      | 0.036810               |
| k-nearest neighbors Classifier| 0.828459      | 0.030710               |
| Decision Tree Classifier     | 0.800764      | 0.045553               |
| GaussianNB                   | 0.748922      | 0.052582               |
| SVC                          | 0.815082      | 0.043396               |
| Random Forest Classifier     | 0.843629      | 0.042054               |
Table 4 Classification Report (without optimizing the model) shows a binary classification of the data set with Random Forest Classifier.

| Tumor Type | Precision | Recall | F1-score | Support |
|------------|-----------|--------|----------|---------|
| Malignant  | 0.85      | 0.92   | 0.89     | 299     |
| Benign     | 0.85      | 0.73   | 0.79     | 181     |
| Accuracy   |           |        | 0.85     | 480     |
| Macro average | 0.85  | 0.83   | 0.84     | 480     |
| Weighted average | 0.85 | 0.85   | 0.85     | 480     |

Table 5 Classification Report: Performance of Random Forest after hyperparameter optimization to have best precision score.

| Tumor Type | Precision | Recall | F1-score | Support |
|------------|-----------|--------|----------|---------|
| Malignant  | 0.89      | 0.94   | 0.92     | 299     |
| Benign     | 0.90      | 0.81   | 0.85     | 181     |
| Accuracy   |           |        | 0.89     | 480     |
| Macro average | 0.89  | 0.87   | 0.88     | 480     |
| Weighted average | 0.89 | 0.89   | 0.89     | 480     |

Table 6 Classification Report: Performance of Random Forest after hyperparameter optimization to have best recall score.

| Tumor Type | Precision | Recall | F1-score | Support |
|------------|-----------|--------|----------|---------|
| Malignant  | 0.91      | 0.93   | 0.92     | 299     |
| Benign     | 0.88      | 0.85   | 0.86     | 181     |
| Accuracy   |           |        | 0.90     | 480     |
| Macro average | 0.87  | 0.85   | 0.86     | 480     |
| Weighted average | 0.87 | 0.87   | 0.87     | 480     |

Table 7 Optimum level of hyper parameters for maximum precision score and the maximum Recall score for the selected features where; n estimators is the number of trees in random forest, min sample split is the minimum number of samples required to split a node and max depth is the maximum number of levels in tree.

| Hyper parameter | Best Condition for Precision | Best Condition for Recall |
|-----------------|------------------------------|---------------------------|
| n estimator     | 500                          | 300                       |
| min sample split| 2                            | 2                         |
| max features    | 10                           | 10                        |
| max depth       | 70                           | 30                        |
Figures

Figure 1

Supervised learning flow chart. The flow chart illustrates the steps of building a classification model to differentiate brain neoplasms using supervised learning technique. According to the study plan, the problem was identified as a classification problem at the initial stage and then the necessary data was collected as the second step. Data pre-processing was executed as the third step and at the fourth step, the data set was split into training and testing sets. Then a suitable ML algorithm for the collected data was selected as the fifth step of the study flow and then, the selected algorithm was trained with the training data as the sixth step. Finally, the developed algorithm was evaluated with the test data and the hyperparameter of the developed model was tuned to reach the optimum accuracy level of the model.

Figure 2

MRI ADC brain image of a 14 years old female patient diagnosed with pilocytic astrocytoma which was radiologically and histo-pathologically identified as a benign tumor. The tumor area is surrounded by the ROI. The texture features were extracted from the selected area.

Figure 3

ANOVA f-test Results chart. ANOVA f-test score for attributes 0 to 15 are illustrated in the graph; mean pixel value of ADC 32.3343, Skewness 3.3444 Kurtosis 9.6250, GLCM Mean1 32.6372, GLCM mean2 29.1327, GLCM variance1 14.0761, GLCM variance2 27.5219 GLCM energy, GLCM Homogeneity 3.4572, 33.9675, GLCM Entropy 4.989, GLCM contrast 47.9462, GLCM Correlation 48.6392, GLCM prominence 15.4134, GLCM Shade 17.1677, Patient Age 9.4337 and Patient Gender 73.7926.
Figure 4

Precision Recall Curve; visualize the sensitivity-specificity trade-off in the classifier the information provided by the curve used to set the decision threshold of the model to maximize the sensitivity and specificity.
Figure 5

Receiver Operating Characteristic curve (ROC Curve). The curve illustrates the behaviour of the false positive rate (x-axis) and true positive rate (y-axis) for a series of different decision threshold values in between 1.00 to 0.00. The smaller values of the X-axis represent the lower false positive rate, and the higher true negative rate. Also, the larger values of Y-axis represent the lower false negative rates and higher true positive rates.
Confusion matrix for the model performance after hyperparameter tuning.

Figure 6

The final confusion matrix. The confusion matrix express the performance of the optimized benign malignant brain tumor brain tumor classification model over the test set.