



## OPEN Texture feature analysis of MRI-ADC images to differentiate glioma grades using machine learning techniques

Sahan M. Vijithananda<sup>1</sup>, Mohan L. Jayatilake<sup>2✉</sup>, Teresa C. Gonçalves<sup>3</sup>, Luis M. Rato<sup>3</sup>, Bimali S. Weerakoon<sup>2</sup>, Tharindu D. Kalupahana<sup>4</sup>, Anil D. Silva<sup>5</sup>, Karuna Dissanayake<sup>6</sup> & P. B. Hewavithana<sup>1</sup>

Apparent diffusion coefficient (ADC) of magnetic resonance imaging (MRI) is an indispensable imaging technique in clinical neuroimaging that quantitatively assesses the diffusivity of water molecules within tissues using diffusion-weighted imaging (DWI). This study focuses on developing a robust machine learning (ML) model to predict the aggressiveness of gliomas according to World Health Organization (WHO) grading by analyzing patients' demographics, higher-order moments, and grey level co-occurrence matrix (GLCM) texture features of ADC. A population of 722 labeled MRI-ADC brain image slices from 88 human subjects was selected, where gliomas are labeled as glioblastoma multiforme (WHO-IV), high-grade glioma (WHO-III), and low-grade glioma (WHO I-II). Images were acquired using 3T-MR systems and a region of interest (ROI) was delineated manually over tumor areas. Skewness, kurtosis, and statistical texture features of GLCM (mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence, and shade) were calculated using ADC values within ROI. The ANOVA f-test was utilized to select the best features to train an ML model. The data set was split into training (70%) and testing (30%) sets. The train set was fed into several ML algorithms and selected most promising ML algorithm using K-fold cross-validation. The hyper-parameters of the selected algorithm were optimized using random grid search technique. Finally, the performance of the developed model was assessed by calculating accuracy, precision, recall, and F1 values reported for the test set. According to the ANOVA f-test, three attributes; patient gender (1.48), GLCM energy (9.48), and correlation (13.86) that performed minimum scores were excluded from the dataset. Among the tested algorithms, the random forest classifier ( $0.8772 \pm 0.0237$ ) performed the highest mean-cross-validation score and selected to build the ML model which was able to predict tumor categories with an accuracy of 88.14% over the test set. The study concludes that the developed ML model using the above features except for patient gender, GLCM energy, and correlation, has high prediction accuracy in glioma grading. Therefore, the outcomes of this study enable to development of advanced tumor classification applications that assist in the decision-making process in a real-time clinical environment.

### Abbreviations

MRI	Magnetic resonance imaging
DW	Diffusion weighted
DWI	Diffusion weighted imaging
CNS	Central nervous system
GBM	Glioblastoma multiforme

<sup>1</sup>Department of Radiology, Faculty of Medicine, University of Peradeniya, Peradeniya 20400, Sri Lanka. <sup>2</sup>Department of Radiography/Radiotherapy, Faculty of Allied Health Sciences, University of Peradeniya, Peradeniya 20400, Sri Lanka. <sup>3</sup>Department of Informatics, University of Évora, 7000 Évora, Portugal. <sup>4</sup>Department of Computer Engineering, Faculty of Engineering, University of Sri Jayawardhanapura, Dehiwala-Mount Lavinia, Sri Lanka. <sup>5</sup>Department of Radiology, National Hospital of Sri Lanka, Colombo 10 01000, Sri Lanka. <sup>6</sup>Department of Histopathology, National Hospital of Sri Lanka, Colombo 10 01000, Sri Lanka. ✉email: Jayatml@pdn.ac.lk

LGG	Low grade glioma
HGG	High grade glioma
ADC	Apparent diffusion coefficient
GLCM	Grey level co-occurrence matrix
ML	Machine learning
ROC	Receiver operating characteristic curve

Glioma is the most common primary neoplasm type in the central nervous system (CNS). According to the epidemiology of intracranial neoplasms, 30% of all the primary CNS neoplasms, as well as 80% of intracranial malignancies that occur in adults, are gliomas<sup>1–3</sup>. However, gliomas can be defined as an abnormal and uncontrollable proliferation of glial cells/neuroglia by bypassing the mechanisms that control the normal cell division to form a heterogeneous group of neoplastic masses belonging to multiple histologic types and malignancy grades<sup>4,5</sup>. Glial cells refer to all the non-neuronal cells that are associated with both CNS and the peripheral nervous system (PNS). The ependymal, oligodendrocytes, astrocytes, and choroid plexus cells are identified as the glial cells that are responsible to maintain the structural integrity of the CNS and PNS while providing metabolic support; nutrient and waste transportation, communication, and insulation to the adjacent neurons. Therefore, the term glioma is considered as a non-specific term that indicates the origin of the tumor is from one of the types of glial cells; i.e., ependymoma, oligodendroglioma, astrocytoma, and choroid plexus papilloma arises from the ependymal, oligodendrocytes, astrocytes, and choroid plexus cells respectively<sup>6</sup>. According to the World Health Organization (WHO), gliomas are classified into four grades (I, II, III, and IV) by considering the aggressiveness (histological and molecular features) of the tumors<sup>7–9</sup>. However, medical imaging including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography and Computed Tomography (PET-CT), PET-Magnetic Resonance and other nuclear imaging modalities such as scintigraphy plays a major role in brain tumor diagnosis, identification, and therapeutic procedures<sup>10–12</sup>.

### Magnetic resonance imaging

Among the above-mentioned medical imaging modalities, MRI is one of the most promising neuroimaging modalities that are being used in the current clinical setup to produce diagnostic medical images of brain tumors<sup>13</sup>. There are a number of MRI sequences; T1 weighted, T2 weighted, Fluid Attenuation and Inversion Recovery (FLAIR), Diffusion-Weighted Imaging (DWI), T1 post-contrast fast-spin echo (T1 FSE), and susceptibility-weighted imaging (SWI) are currently being used in routine neuroimaging practices. Among the above sequences, the DWI images have the ability to probe the random Brownian motions of water molecules within tissues on a voxel basis<sup>6,14</sup>. As the DW images provide information about the net direction of the water molecules within tissues, it is widely appreciated in observing the microscopic behavior of biological tissues; the existence of membranes, cellularity, the intracellular-extracellular water equilibrium, and the presence of macromolecules. Changes in the microscopic diffusion of water molecules within tissue indicate the alteration of homeostasis at the cellular level<sup>15</sup>. Therefore, DWI became an indispensable tool in clinical neuroimaging. It became widely popular among clinicians as a powerful imaging tool for the diagnosis of some life-threatening conditions such as ischemic, tumors, trauma, and non-life-threatening conditions like schizophrenia, multiple sclerosis, dyslexia, etc<sup>16–19</sup>.

### Diffusion weighted imaging and apparent diffusion coefficient

DWI can be acquired at different diffusion sensitization levels by changing the critical parameters; the amplitude and duration of the applied diffusion sensitization gradient. These parameters have the ability to encode different properties of tissues into DWI signals while controlling the magnitude of diffusion weighting in the resultant image. The sensitivity of the acquired DW image is indicated by the *b-value* that is measured in seconds per square millimeter ( $s/mm^2$ ). The *b-value* is also proportional to the duration, and square of the amplitude of the applied diffusion sensitization gradient. The diffusion of water molecules within tissues is qualitatively assessed by trace DW images and it is being quantitatively assessed by calculating the apparent diffusion coefficient (ADC) parameter (see Eq. 1). According to Eq. (1), it is mandatory to have the involvement of at least two DW images with different *b-values* to calculate ADC values. The images with  $b = 0 s/mm^2$  are utilized as the lower limit in common radiology practices while *b-values* from 600 to 1000  $s/mm^2$  are used for the upper limit<sup>20,21</sup>. However, *b-values* greater than 1000  $s/mm^2$  are also applied to generate ADC in non-routine studies<sup>22</sup>. The degree of the diffusion of water molecules through adjacent structures is visualized by plotting the calculated ADC values as a parametric map. High ADC values represent less impedance for the diffusion of water molecules within tissues, and such tissues are hyperintense in ADC while hypointense in trace DW images. As a result, these hyperintense and hypo-intensities express different textures for different tissue types according to their microscopic behavior.

MRI-ADC imaging provides information about tissue microstructure by assessing water diffusion. It detects changes in cellular density and organization, which are indicative of diseases. Lower ADC values suggest higher cellular density and disrupted tissue architecture, highlighting pathological alterations. MRI-ADC imaging is sensitive to early microstructural changes, enabling early disease diagnosis. It offers diagnostic value in oncology, neurology, and other medical fields. Being non-invasive, it provides valuable insights without the need for invasive procedures.

### Image texture

Texture describes the structure and surface of an image by considering the regular repetition of an element or pattern on the surface. Image texture provides important information about the spatial arrangement of intensities or colors in an image<sup>23,24</sup>. The ADC images have the ability to visualize the structures with different diffusivity in

different grey levels/different intensities which make the image enriched with texture. Texture analysis is based on finding the specific patterns of hidden characteristics of the texture and presenting them in a more simplified and unique way. Grey Level Co-occurrence Matrix (GLCM) can be identified as a promising statistical method to examine the texture of an image by considering the spatial relationship of pixels<sup>25</sup>. GLCM is a square matrix with dimensions equal to the number of grey levels ( $n \times n$ ) contained in the 2D parametric ADC image ( $I$ ) and it counts the co-occurrence of neighboring grey levels of pixels within the image along  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  orientations and summed<sup>26,27</sup>.

### Higher order moments

Apart from the first and second-order statistics such as mean, and variance, the higher-order statistics (HOS) have also played a tremendous role in signal processing and system analysis in recent history. The higher-order statistics are the statistical functions that use high power of sample; higher than 2nd order (lower order) statistics, provide useful tools for addressing issues in nonlinear systems<sup>28</sup>. Higher-order statistics such as third-order (Skewness), and fourth-order (Kurtosis) carry more useful information due to their phase sensitiveness. Such information is critical in developing robust statistical modes to identify non-minimum phase systems<sup>29,30</sup>. In third-order statistics; skewness measures the asymmetry around the mean of a probability distribution of a data set. The skewness of a normal distribution remains at zero. However, the distributions skewed to left are indicated by negative ( $-$ ) values while the distributions skewed to right are indicating positive ( $+$ ) values. The distributions with skewness value less than  $-0.5$  or higher than  $0.5$  are considered highly skewed distributions. Kurtosis measure and compare the shape (tail) of the probability distribution of a real-valued random variable with a normal distribution. The kurtosis value of any univariate normal distributions remains at 3 and the distributions with kurtosis of more than 3 are considered as platykurtic distributions. In contrast, the distributions with kurtosis of less than 3 are identified as leptokurtic distributions<sup>31</sup>.

### Machine learning

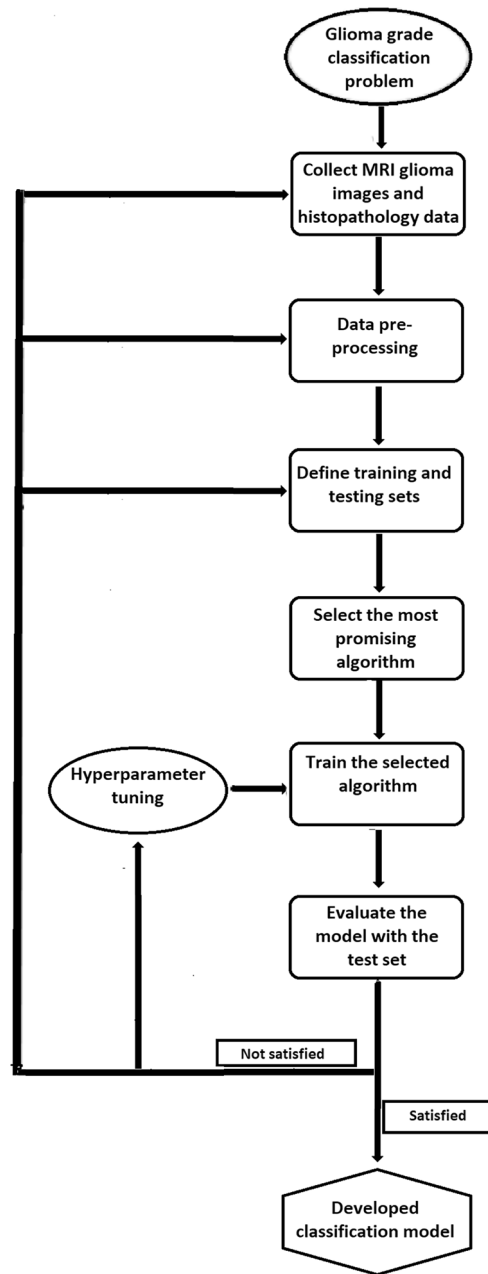
Machine learning is a branch of artificial intelligence (AI) that allows computers to “learn” from data and develop analytical models to aid and/or support in making decisions and predictions with minimal human involvement. Here, the ML algorithms are used to identify the hidden characteristics/patterns of data to develop analytical models. ML approaches can be classified into three main categories: Supervised learning, Unsupervised learning, and Reinforcement learning<sup>32</sup>. Supervised learning uses labeled datasets to train ML algorithms while unsupervised learning uses unlabeled datasets to train. Reinforcement learning is a type of machine learning that learns as it goes by using trial and error. Supervised learning is a powerful learning method that is being used to address a variety of real-world classification, and regression problems<sup>33,34</sup>.

Therefore, the supervised learning method can be identified as one of the most common ML paradigms that use labeled input data to train ML algorithms<sup>25,35–37</sup>. When the data is fed into the algorithm, it identifies the hidden characteristics, patterns, and correlations for each class and makes ML models using such information. The process iterates until the algorithm achieves the highest prediction accuracy and the developed model is able to address the intended problem with high accuracy level (see Fig. 1). The accuracy level of the developed ML model is optimized by tuning the hyper-parameters of the model<sup>38</sup>. Among the various types of supervised learning algorithms, Neural Networks, Naïve Bayes, Linear Regression, Logistic Regression, Support Vector Machines, K-Nearest Neighbor, Decision Tree, and Random Forest algorithms are the most commonly used ones. From the above algorithms, the Random Forest algorithm can be identified as an ensemble method that uses a collection of decision trees to generate a decision in classification and regression problems.

### Literature survey

There are numerous studies available in the literature that focus on the development of glioma grade classification models. In recent years, several notable studies have contributed to this field. When we consider few of most recent studies, in year 2019, A. Vamvakas et al., developed support vector machine (SVM) binary classification model to predict glioma types (High grade glioma, Low grade glioma) using the radiomics features extracted from several MRI image sequences including T1 pre/post-contrast, T2-FSE, T2-FLAIR (Fluid Attenuation and Inversion Recovery) Diffusion Tensor, Perfusion Imaging and 1H-MR Spectroscopy. As a result, they could predict these two classes of gliomas with 95.5% Accuracy<sup>39</sup>. Another study conducted in 2019 by Nidhi Gupta et al. involved the development of a model to identify and classify gliomas using MRI images in T1, T1-post contrast, T2, and FLAIR sequences. The researchers incorporated image texture features, as well as morphological and inherent characteristics of the tumor such as solidity, perimeter, area, and orientation. Their classification model achieved an accuracy of 97.76%<sup>40</sup>.

In another study conducted in 2017 by Xin Zhang et al., various machine learning methods were compared for glioma grading, specifically distinguishing between low grade and high grade gliomas, using multi-parametric MRI data. The study extracted quantitative parameters, including parametric histogram and image texture attributes, from perfusion, diffusion, and permeability maps of gliomas. The SVM method achieved a classification accuracy of 94.5% in differentiating between the two glioma classes<sup>41</sup>. In the year 2012, Nitish Zulpe1 and Vrushen developed a brain tumor classification model using the GLCM texture features extracted from T2 weighted and proton density (PD) MRI image sequences obtained from four subjects with four different types of brain tumors. However, the model developed in a two-layered Feedforward Neural Network predicted the tumor types with 97.5% average accuracy<sup>42</sup>. Jiang et al., in the year 2017 developed a statistical model to discriminate low grade and high-grade gliomas by using the texture features extracted from multiple types of MRI sequences such as T2-FLAIR and T1WI-Contrast enhanced DWI sequences and found GLCM cluster shade, entropy and homogeneity as the best features to use in differentiating low grade and high grade gliomas<sup>43</sup>. Rajagopal et al.,



**Figure 1.** Application of supervised learning method for multiclass tumor classification problem. The flow chart illustrates the steps followed in developing a multiclass classification model. After identification of the nature of the problem, the necessary MRI and histopathology data are collected. At the data pre-processing step, the texture features are extracted from MRI images, and the extracted data is prepared to be compatible with training the machine learning model (data labeling, removing defected data, binarization). The next step splits the dataset into train and test sets. The most promising machine learning algorithm for the dataset is selected and fed to the algorithm with the training set to build the classification model. Finally, the performance of the developed model is assessed. When the performance did not meet the required level of performance, the hyperparameters of the developed model are tuned and find the most suitable combination of hyperparameters. Also, sometimes it is necessary to revise the data collection, data pre-processing, and repeat training and testing steps until meeting the required performance of the model.

(2019) developed a glioma detection and segmentation model using GLCM features extracted from the MRI brain images and they utilized the random forest classifier to build the classification model with an accuracy of 97.7%<sup>44</sup>.

In the year 2019, Reza et al. proposed a high grade and low-grade glioma classification model developed in random forest classifier was able to classify gliomas with significantly high accuracy the model developed in SVM<sup>45</sup>. However, in the study, the texture features of MRI brain images have been acquired from multiple MRI sequences such as T1 weighted, T2 weighted, T1- post-contrast, and FLAIR.

In the year 2019, Deniz Alis et al., developed machine learning model to predict IDH1 status in high-grade gliomas. This study used texture features extracted from axial T2WI FLAIR, post-contrast T1WI, and ADC maps to feed random forest classifier. The developed model was able to predict IDH1 status of high-grade gliomas with 86.94% accuracy<sup>46</sup>. Similarly, Han et al. (2018) used ADC-based texture features along with other clinical and radiological features to classify gliomas into three different grades and achieved an accuracy of 89.6%<sup>47</sup>.

The study conducted by Radwa et al., in the year 2021 was able to find a significant difference of mean ADC values between high-grade glioma (HGG) and low-grade glioma (LGG) by analyzing the features extracted from ADC images of gliomas<sup>48</sup>. Similarly, in the year 2018, Fusun et al. developed a machine learning model based on support vector machine to differentiate between high-grade glioma (WHO III and IV) and low-grade gliomas (WHO I and WHO II) using the features extracted from T1 and T2-weighted, diffusion-weighted, diffusion tensor, MR perfusion and MR spectroscopic imaging. However, their binary classification model was able to classify two glioma classes with an accuracy of 93.0%<sup>49</sup>. In summary, classifying glioma using MR images is a prevalent research problem among the scientific community focused on the advancement of medical imaging.

Almost all the studies discussed in the literature use at least T1-post-contrast images that involve invasive procedures. However, the literature currently lacks strong evidence for a method developed to differentiate glioma grades solely based on texture features extracted from MRI-ADC images and avoid any kind of invasive procedures. Here in this study, our aim was to address this gap in the literature and generate novel insights and contribute to the existing knowledge base in this field. The proposed non-invasive approaches aim to provide accurate classification results while minimizing patient discomfort and potential risks associated with contrast agents.

### Hypothesis

The study is based on the hypothesis; that there is an existence of a correlation between the extracted features (patients' demographics, higher-order moments of ADC, and GLCM texture features of ADC) and the severity level of the glioma (WHO glioma grading levels).

### Objectives

Objectives of this study include the development of a robust and non-invasive method for distinguishing between low-grade glioma (WHO I/II), high-grade glioma (WHO III), and glioblastoma (WHO IV) based on features extracted from MRI-ADC images. This will be achieved through the analysis of patients' demographics, higher-order moments of ADC, and GLCM texture features of ADC using machine learning techniques. The primary aim of the study is to improve the accuracy of glioma diagnosis, which will ultimately lead to better patient outcomes with zero invasive procedures and minimum patient discomfort. This research will contribute to the advancement of medical knowledge in the field of neuro-oncology and may have significant implications for clinical practice.

### Main contributions

The main contributions of this work are:

- Development of a robust and non-invasive method: The study aims to develop a robust and non-invasive method for distinguishing between low-grade glioma (WHO I/II), high-grade glioma (WHO III), and glioblastoma (WHO IV). This method will be achieved through the analysis of patients' demographic information, higher-order moments of ADC, and GLCM texture features of ADC using machine learning techniques.
- Improvement of glioma diagnosis accuracy in a noninvasive manner: The primary aim of the study is to improve the accuracy of noninvasive glioma classification, which will ultimately lead to better patient outcomes with minimum patient discomfort.
- Advancement of medical knowledge in the field of neuro-oncology: This research will contribute to the advancement of medical knowledge in the field of neuro-oncology by providing new insights into the diagnosis and classification of gliomas. The study may also lead to the discovery of new biomarkers or imaging features that can be used to improve the diagnosis and treatment of gliomas.
- Potential implications for clinical practice: The findings of this study may have significant implications for clinical practice by providing clinicians with a more accurate and reliable method for diagnosing gliomas. This could lead to improved patient outcomes and a reduction in the number of unnecessary biopsies or surgeries.

### Results

According to the results of the analysis of variance (ANOVA) F-test feature selection, the patient gender (1.4850), GLCM Energy (9.4805), and the GLCM Correlation (13.8695) were excluded from the dataset as such features reported the minimum scores (see Table 1) (see Fig. 2). Among the seven ML algorithms tested in the tenfold cross-validation process, the Random Forest Classifier reported the maximum mean-cross-validation score (mean-accuracy) for both balanced ( $0.8772 \pm 0.0237$ ) and imbalanced ( $0.7901 \pm 0.0495$ ) datasets. Therefore, the Random Forest Classifier was selected as the basic tool for building the glioma classification model (see Table 2).

However, the classification model built by training the Random Forest Classifier algorithm with the train set predicted the glioma categories at 86.08% overall accuracy with a 13.26% average error (see Table 3). According to the area under the curve (AUC) of receiver operating characteristic curve (ROC), the base model performance was glioblastoma vs rest: 0.9434, high-grade glioma vs rest: 0.9521, and low-grade glioma vs rest: 0.9885 (see Fig. 3). After identifying the *min\_samples\_split*, *n\_estimators*, *max\_depth*, *bootstrap*, *min\_samples\_leaf*, and

Feature	ANOVA F-test score
Mean ADC	230.8198
Skewness	14.3727
kurtosis	135.5591
GLCM mean 1	262.2695
GLCM mean 2	212.9521
GLCM variance 1	50.6215
GLCM variance 2	38.5333
Energy	9.4805
Entropy	30.3616
Contrast	44.8493
Homogeneity	27.6501
Correlation	13.8695
Prominence	258.8106
Shade	249.7605
Patients' age	128.6092
Patients' gender	1.4850

**Table 1.** ANOVA F-test scores for each feature. The table illustrates the performance of each feature at the ANOVA F-test feature selection process.

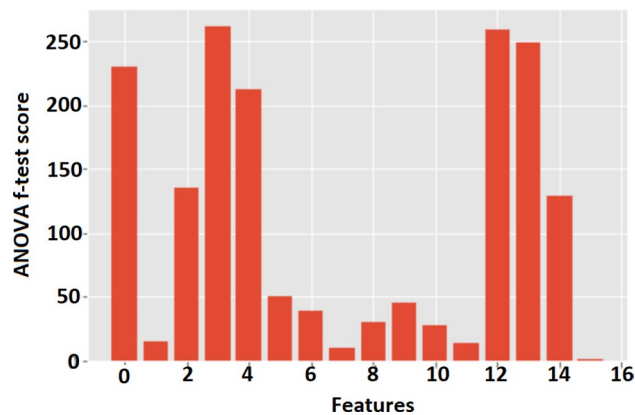
	Algorithm	Mean accuracy	(SD)	Accuracy as percentage
With SMOTE	k-Nearest neighbors classifier	0.8398	0.0227	83.98%
	Linear discriminant analysis	0.7181	0.0385	71.81%
	Gaussian Naïve Bayes	0.6640	0.0420	66.40%
	Decision tree classifier	0.8176	0.0507	81.76%
	Support vector machine	0.8132	0.0407	81.32%
	Random forest classifier	0.8772	0.0237	87.72%
	Logistic regression	0.7457	0.0371	74.57%
Without SMOTE	k-Nearest neighbors classifier	0.6969	0.0541	69.69%
	Linear discriminant analysis	0.7167	0.0501	71.67%
	Gaussian Naïve Bayes	0.6927	0.0525	69.27%
	Decision tree classifier	0.7443	0.0877	74.43%
	Support vector machine	0.6772	0.0484	67.72%
	Random forest classifier	0.7901	0.0495	77.83%
	Logistic regression	0.6910	0.0457	69.10%

**Table 2.** The mean cross-validation scores, standard deviation (SD) and the accuracy from different algorithms for the balanced and imbalanced datasets. The table illustrates the mean K-fold cross-validation scores and the corresponding standard deviations acquired by each classification algorithm with and without the application of the synthetic minority over-sampling technique (SMOTE) over the dataset.

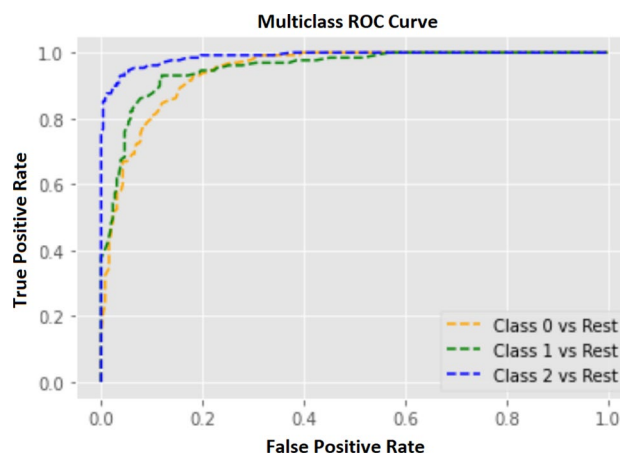
	Glioma category	Precision	Recall	f1-score	Support	Accuracy
Base model	0	0.84	0.79	0.81	130	
	1	0.83	0.85	0.84	129	86.08%
	2	0.91	0.94	0.92	129	
Tuned model	0	0.84	0.86	0.85	130	
	1	0.86	0.84	0.85	129	88.14%
	2	0.95	0.94	0.94	129	

**Table 3.** Performance of the developed machine learning model with and without hyperparameter tuning. The table illustrates the precision, recall, and f1-score acquired by each glioma category in both base model and the tuned classification modes. The glioma categories 0, 1, and 2 represent glioblastoma, high-grade glioma, and low-grade glioma, respectively.





**Figure 2.** ANOVA F-test results in a bar chart. The figure illustrates the bar chart of the feature importance scores for each input feature. The standardized features; mean ADC, skewness, kurtosis, GLCM mean 1, GLCM mean 2, GLCM variance 1, GLCM variance 2, energy, entropy, contrast, homogeneity, correlation, prominence, shade patient' age, and gender are indicated by 0 to 15 numbers in the bar chart respectively.



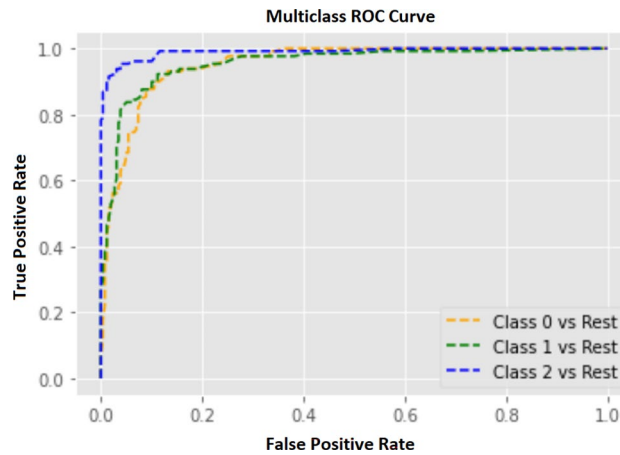
**Figure 3.** Multiclass receiver operating characteristic (ROC) curve for the base model. The ROC curve illustrates the trade-off between true positives and false positives that reflects the performance of the classification model at various threshold settings. The performance of multiclass classification models is displayed in ROC curves using the one vs rest technique. Class 0, 1, and 2 represent glioblastoma, high-grade glioma, and low-grade glioma, respectively. The area under the curve (AUC) for each curve; yellow: 0.9434, green: 0.9521, and blue: 0.9885.

*max\_features* as tunable hyperparameters of the base model, the grid search cross-validation technique found the optimum conditions/ combinations of above parameters; *n\_estimators*: 108, *bootstrap*: False, *max\_depth*: 50, *max\_features*: auto, *min\_samples\_leaf*: 1, and *min\_samples\_split*: 2. As a result of assessing the performance of the tuned model using the test set, the tuned model was able to predict the glioma categories with 88.14% accuracy and 11.86% error which is a 2.40% of increment from the accuracy of the base model (see Table 3). Moreover, the tuned classification model correctly predicted 121 out of 129 low-grade glioma image slices, 109 of 129 high-grade gliomas slices, and 112 out of 130 image slices of glioblastomas (see Fig. 5). According to the ROC-AUC values, the tuned model was performed at glioblastoma vs rest: 0.9525, high-grade glioma vs rest: 0.9545, and low-grade glioma vs rest: 0.9901 (see Fig. 4).

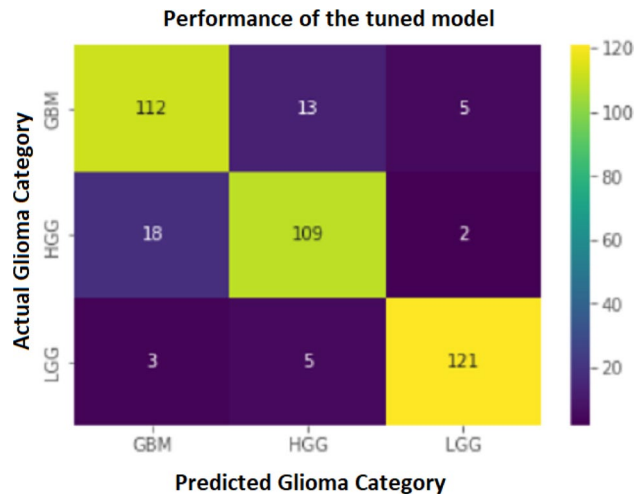
## Discussion

Finding a robust way to identify the severity level or tumor grades of glioma using MRI images has been a leading scientific research area in the past few decades<sup>36</sup>. However, in this study, we discussed about developing an automated and non-invasive method to differentiate gliomas according to the severity level/WHO grades using the information acquired from patient demographics, statistical texture features of GLCM, the mean, skewness, and kurtosis of ADC. However, the intended texture features of each image slice were extracted using homemade software called Brain Lesion Differentiation and Identification Assistant (BLE DIA) which was specifically designed to extract the texture features of MRI brain tumors<sup>29</sup>.

The whole glioma ADC image population acquired from both institutes National Hospital of Sri Lanka (NHSL) and Anuradhapura Teaching Hospital (ATH) was divided into three categories according to the severity



**Figure 4.** Multiclass receiver operating characteristic (ROC) curve for the base model after hyperparameter tuning. The performance of tuned multiclass classification models is displayed in ROC curves using the one vs rest technique. Class 0, 1, and 2 represent glioblastoma, high-grade glioma, and low-grade glioma respectively. The area under the curve (AUC) for each curve; yellow: 0.9525, green: 0.9545, and blue: 0.9901.



**Figure 5.** Confusion matrix illustrating the performance of the tuned classification model. According to the confusion matrix, the tuned model predicted 112 out of 130 cases of glioblastoma multiforme (GBM), 109 out of 129 cases of high-grade glioma (HGG), and 121 out of 129 cases of low-grade glioma (LGG).

of glioma; LGG (WHO I/II), HGG (WHO III), and GBM (WHO IV). According to the statistics, each category contained an unequal number of image slices; 109, 182, and 431 for LGG, HGG, and GBM categories respectively. To avoid the effects of the imbalanced sample sizes of data between each category, the synthetic minority over-sampling technique (SMOTE) over-sampling technique was implemented and as a result, the sample sizes of each category were equalized to the sample size of GBM as it has the highest sample size within the population<sup>50,51</sup>.

The results of the cross-validation for seven machine learning classification algorithms indicate that Random Forest Classifier has the highest accuracy score of 0.7901 and Decision Tree Classifier has the second-highest accuracy score of 0.7443 before applying SMOTE. However, after applying SMOTE, the accuracy scores of all the algorithms improved significantly, especially for Gaussian Naïve Bayes, which had the lowest accuracy score before SMOTE application<sup>52,53</sup>. The Random Forest Classifier also had a substantial improvement in accuracy score, with a score of 0.8772, which is the highest accuracy score among all the algorithms after SMOTE application (see Table 2).

Overall, the application of SMOTE technique has positively impacted the performance of all the algorithms, except for Gaussian Naïve Bayes, which had slight decreases in accuracy scores after SMOTE application. The cross-validation results suggest that the application of SMOTE technique can significantly improve the performance of machine learning classification algorithms, especially for imbalanced datasets. However, the impact of SMOTE can vary across different algorithms, and it is essential to evaluate the performance of different algorithms before and after applying SMOTE to determine its effectiveness.



The dataset with equalized sample sizes for each glioma category was split into train and test sets. The most promising algorithm for the data was selected using 10-fold cross-validation. Within this process, the seven most popular supervised learning algorithms were tested and the algorithm that performed the highest cross-validation score with a lesser standard deviation (Random Forest algorithm) was selected to build the classification model (see Table 2).

However, the developed model (base model) could predict the glioma categories with an accuracy of 86.08%, and the high ROC-AUC values calculated in the one versus rest (OVR) method witnessed the high classification power of the developed classifier (see Fig. 4). Also, the performance of the base model over the test set was measured by calculating precision, recall, and f1-score for each glioma category (see Table 3). The accuracy of the base model was optimized by changing the parameters that are critical for the learning process, also known as hyperparameter tuning<sup>54</sup>. At last, the performance of the tuned model was estimated using the test data set and measured by calculating the accuracy and the values of precision, recall, and f1-score for each glioma category. Comparing the precision, recall, and f1-score values of the base model and the tuned model for each category, all the values in the tuned model except the recall score of the HGG category are higher or equal to the precision, recall, and f1-score values of the base model (base model:0.85 > tuned model:0.84) (see Table 3). In addition, the overall classification power of the tuned model for each glioma category was drafted as ROC curves (OVR technique), and the behaviors of the AUC values received for each category in both the base model and the tuned model were compared. As a result, we could identify the improvements in the degree of separability of the tuned model than the base model.

By comparing the results of this study with the study conducted by Alksas et al., in the year 2022, they could reach 95.8% overall prediction accuracy. However, the methodology they used was vastly different from this study. according to their methodology, they have extracted data from several MRI sequences including intravenous (IV) contrast-enhanced sequences such as T1 weighted post-contrast sequence<sup>55</sup>. Young Jin et al., in the year 2014 conducted a study to differentiate gliomas into WHO-II, WHO-III and WHO-IV categories using the features extracted from ADC maps of tumors. They calculated P-value for each feature and observed that high-grade gliomas reported significantly higher entropy values and lower fifth percentiles of the ADC cumulative histogram than low-grade tumors. Entropy was the only parameter that was significantly different between grades III and IV, and its diagnostic accuracy was superior to that of the fifth percentile of the ADC histogram in distinguishing high- from low-grade gliomas<sup>56</sup>.

Although the results of this study are promising, there were two main limitations to address when practically executing the study process. The major limitation is drawing the ROIs of 3D tumors in a 2D plane. According to the shape and volume of the tumor, it may appear on several image slices as well as several spots in the same slice. To overcome this problem, we decided to take several ROIs in the same image slice but in different locations and draw ROIs on each image slice that contains the details of the tumor. The next limitation was the lack of patient details. Most of the data collected in this study were accomplished in a retrospective manner. Therefore, tracing the medical records (MRI images, radiological reports, and histopathology reports) of each subject was a challenging event.

## Conclusion

The study concludes that the features extracted and applied in this study such as mean ADC, skewness, kurtosis, GLCM mean 1, GLCM mean 2, GLCM variance 1, GLCM variance 2, entropy, contrast, homogeneity, shade, patients' age can be collectively used as potential biomarkers to differentiate gliomas according to its severity. Moreover, due to the high accuracy level and the high AUC values of the developed classification model, it can be implemented in clinical setup with further advancements as assistance for clinicians who are involved in the tumor diagnosis process.

## Methods

This prospective study was designed to address the above objective which is building a robust ML model to predict the severity of glioma using the texture features and higher-order moments of MRI-ADC and the patients' demographics. According to the nature of the collected data and the concerned problem of the study, it was designed as a multi-class classification study and Fig. 1 illustrates the workflow of the supervised learning method utilized in the development of the glioma classification ML model.

## Data acquisition and preparation

The study was carried out using 722 labeled (431 for glioblastoma (GBM)-WHO IV, 182 for high-grade glioma (HGG)-WHO III, and 109 for low-grade glioma (LGG)-WHO I and II) MRI-ADC image slices of 88 human subjects being 57 males, and 31 females who were within the 8 to 90 age range. The pathological condition of each subject was confirmed using the radiological and histopathological reports provided by the experts. All the MRI-DW Digital Imaging and Communications in Medicine (DICOM) data, radiological reports, and corresponding histopathological reports were collaboratively obtained from the departments of Radiology and Histopathology at National Hospital Sri Lanka (NHSL) and the Teaching Hospital Anuradhapura (THA) after obtaining informed consent of the patients, and the ethical clearance approvals from the ethical review board of the Faculty of Medicine, University of Peradeniya, Sri Lanka and the ethical review board of the NHSL. All the data collection activities were carried out within a one-year period under the supervision of the consultants/experts of each institute and department. However, patients with insufficiently detailed or potentially inaccurate information, and damaged/artifact-affected MR images were excluded during the data preprocessing phase.

All the brain tumors that occur without the involvement of glial cells; Meningioma, metastasis, dermoid or epidermoid cysts, choristomas, chondrosarcoma, hamartoma, chordoma, etc., the tumors outside the interested

region (extracranial tumors), patients with weak radiological or histopathological histories and corrupted MRI images of brain tumors were excluded at the data preprocessing stage. According to the objectives of this study, the patients' demographics (age and gender) the mean, skewness (3rd order statistics), kurtosis (4th order statistics) of ADC, and the statistical texture features of GLCM (mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence, and shade) were extracted from the selected subjects.

### Generate ADC images

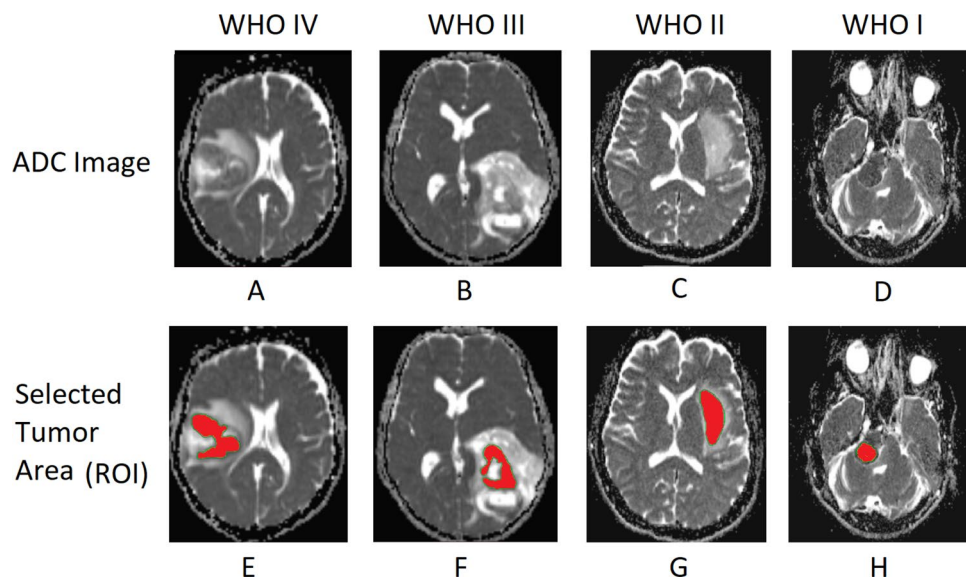
All the MRI-DW images of the selected subjects were acquired using 3T MR systems and head coils. The Echo Planner Imaging (EPI) sequence with the parameters; TR = 4300 ms, TE = 68 ms (being TR the time of repetition and TE the time of echo), flip angle = 90°, field of view (FOV) = 219mm × 219mm, matrix size = 124 × 124 and slice thickness = 1 mm were utilized to generate the required  $b = 0\text{s/mm}^2$ , and  $b = 1000\text{s/mm}^2$  DWI images. The DW images generated in two different diffusion sensitization levels ( $b$ -values);  $b = 0\text{s/mm}^2$  image, and its corresponding  $b = 1000\text{s/mm}^2$  image of each patient were collected and utilized to generate ADC images by merging them according to Eq. (1).

$$ADC = \sum_{i=1}^n \frac{\ln \frac{S_i}{S_0}}{b_i}. \quad (1)$$

Where  $i$  represents the image number while the  $S_i$  represents the  $i$ th image (the image acquired with a diffusion pulse of  $i$ ).  $S_0$  is the first image (image acquired without any diffusion pulses) and  $n$  is the number of images and  $b_i$  is the diffusion gradient value.

### Region of interest (ROI) selection and feature extraction

the tumor areas within the generated apparent diffusion coefficient (ADC) images of each patient were identified with the assistance of two board-certified consultant radiologists who possess extensive experience in the field of diagnostic radiology, with over 20 years of professional practice. The selection of ADC values within the tumor regions was carried out through the manual drawing of regions of interest (ROI) that encompassed predetermined tumor locations, as illustrated in Fig. 6. All ROIs were delineated manually by a radiology postgraduate student, who was under the strict supervision of the same two consultant radiologists. However, the mean ADC, the higher-order moments of ADC; skewness (3rd order statistics), and kurtosis (4th order statistics) values within the selected ROI were extracted according to Eqs. (2) and (3) respectively. All image processing, ROI selection, and feature extraction processes involved in this study were conducted using custom-made software named Brain Lesion Differentiation and Identification Assistant (BLeDIA) which was developed in Python 3.7.



**Figure 6.** Apparent diffusion coefficient (ADC) images of gliomas. (A) An ADC brain image of a 62-year-old male patient presented with glioblastoma multiforme (GBM) (WHO grade IV). (B) ADC brain image of a 16-year-old male patient with Anaplastic oligodendroglioma (WHO III). (C) ADC brain image of a 39-years-old female patient presented with low grade (WHO II) glioma. (D) ADC brain image of a 49-years-old male patient with presented with a schwannoma (WHO I) (E–H) illustrate the region of interest drawn over the tumor areas of (A–D) images respectively.

$$\text{Mean\_ADC} = \frac{\sum_{i=1}^N P_i}{N} \quad (2)$$

Where  $p_i$  is the signal intensity in  $i$ th pixel and  $N$  is the total number of pixels within the ROI

$$n^{\text{th}} \text{ moment} = \sum_i (P_i - P)^n f(P_i) \quad (3)$$

Where  $p_i$  represents the signal intensity in  $i$ th pixel,  $i$  represents the number of pixels within the ROI,  $p$  represents the mean of the pixel values and  $f(p_i)$  is the probability of the signal intensity of the pixel.

Using the same ROIs, the GLCM matrices corresponding to each tumor image were generated according to Eq. (4). The generated GLCM matrices were utilized to extract the statistical texture features of GLCM; mean, variance, energy, entropy, contrast, homogeneity, correlation, prominence, and shade, corresponding to each tumor image. However, most of the texture features are calculated as weighted averages of the normalized GLCM cell contents. Equations (5) to (13) describe the methods utilized to extract the statistical texture features of GLCM and within the equations,  $P_{i,j}$  represents the probabilities calculated for values in the GLCM matrix,  $N$  is the grey levels count in the image,  $\mu$  is the mean of  $P_{i,j}$  matrix,  $\mu_i$  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 represents the joint probability occurrence of the pixel pairs containing  $x$  and  $y$  grey level values for  $\delta_v$  and  $\delta_s$  specific spatial offset between the pixel pairs, and  $I$  represents the 2D parametric ADC map with dimensions of  $n \times n$  (number of grey levels).  $s$  and  $v$  are the spatial positions in image  $I$  (see Eq. 4).

$$L_{\delta_v \delta_s}(x, y) = \sum_{x,y=1}^n \begin{cases} 1, & \text{if } I(v, s) = x \text{ and } I(s + \delta_{s,v} + \delta_v) = y \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

GLCM Mean: The equation calculates the mean values based on the pixel values of adjacent pixels. In the equation, the left-sided equation calculates the mean based on the pixel with  $i$  value ( $\mu_i$ ), meanwhile the right-side equation calculates the GLCM mean based on the pixel with  $j$  value ( $\mu_j$ ). According to the equation, the similar values for  $\mu_i$  and  $\mu_j$  indicates that the GLCM matrix is identically symmetrical.

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}) \quad \mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j}) \quad (5)$$

GLCM Variance: GLCM variance measures the dispersion of cell values around the mean. The magnitude of the variance depends on the mean cell values and the dispersion around the mean cell value within the GLCM. Since the GLCM variance is calculated using the GLCM, there is always an involvement of two pixels (the reference ( $i$ ) and the adjacent ( $j$ ) pixel). GLCM variance gives the same values for both variances calculated based on pixels value  $i$  or  $j$  when the matrices are symmetrical

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

GLCM Energy (ENR): The GLCM energy measures the uniformity of the grey level distribution of an image. An identically uniform distribution of grey levels in an image (window is very orderly) expresses 1 for GLCM energy and it becomes 0 for images that have an identically nonuniform distribution of grey levels. Here, GLCM energy uses each  $P_{(i,j)}$  value as a weight for itself in the calculation of GLCM energy.

$$\text{ENR} = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (7)$$

GLCM entropy (ENT): 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 expresses larger entropy values.

$$\text{ENT} = \sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j}) \quad (8)$$

GLCM contrast (CNT): GLCM contrast, also known as the sum of squares variance, measures the intensity difference between two neighboring pixels ( $i$ , and  $j$ ) over the whole image. GLCM contrast becomes 0 for constant images ( $i=j$ ), while the weights continue to increase exponentially as the difference of pixel intensities ( $i-j$ ) increases. However, the edges, noise, or wrinkled textures within an image increase the contrast value.

$$\text{CTN} = \sum_{i,j=0}^{N-1} P_{i,j}(i - j)^2 \quad (9)$$

GLCM homogeneity (HOM): GLCM Homogeneity is the way of measuring the smoothness of distribution of grey levels within an image, which is inversely correlated with contrast.

$$HOM = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (10)$$

GLCM correlation (COR): The linear dependency of grey levels on neighboring pixels of the image is measured by the GLCM correlation. When there is a linear and predictable relationship between the two pixels, the corresponding correlation increases. Therefore, the images with high correlation values express that there is high predictability of pixel relationship.

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

GLCM cluster shade (CS): Evaluate the tendency of clustering of the pixels by measuring the skewness of pixel values within the matrix. GLCM Cluster shade measures the uniformity of a grey image and values fluctuate between 0 to 2. Therefore, the higher values for cluster shade indicate the nonuniform distribution of grey values in the image.

$$CS = \sum_{i,j=0}^{N-1} \{i + j - \mu_i - \mu_j\}^3 P_{i,j} \quad (12)$$

GLCM cluster prominence (CP): Measures local intensity variation of pixels and the asymmetry of an image. A high prominence value indicates less symmetry of an image while an image with a less cluster prominence value shows the peak in the GLCM matrix around the mean.

$$CP = \sum_{i,j=0}^{N-1} \{i + j - \mu_i - \mu_j\}^4 P_{i,j} \quad (13)$$

### Feature selection and model training

Following the extraction of GLCM texture features, mean ADC, and the higher-order moments of ADC, the demographic data corresponding to each subject was taken to a single spreadsheet. Then, all the feature values corresponding to each image slice were labeled manually with the final diagnosis. According to the labels, the dataset was divided into three classes; Glioblastoma (WHO IV), High-grade glioma (WHO III), and Low-grade glioma (WHO I and II). However, the sample size of each class was not equal to each other.

Therefore, the Synthetic Minority Over-sampling Technique (SMOTE) was utilized to balance the imbalanced sample sizes of each class. SMOTE generates synthetic examples of the minority class by following a set of steps. Initially, a random minority class example is selected, and its k-nearest neighbors from the same minority class are identified. Then, one of the k-nearest neighbors is randomly chosen. A new synthetic example is produced by interpolating between the selected example and the randomly selected nearest neighbor. The interpolation involves computing a weighted sum of the feature values of the two examples. The process continues until the required number of synthetic examples has been generated. The outcome of this algorithm is an increase in the number of minority class examples, which can enhance the performance of classifiers that are biased toward the majority class.

Data within each class of imbalanced (before applying SMOTE) and balanced (after applying SMOTE) datasets were split into train and test sets with a proportion of 70%:30%, respectively by keeping the random state at 42. The purpose of considering the states of the dataset before and after applying SMOTE was to evaluate the effect of SMOTE in developing ML models. Then the features in each train set were standardized as all the features centered around zero mean and unit variance. This standardization process avoids the domination of features with high variance in the learning process. Therefore, it leads the estimator to learn from other features correctly and unbiasedly (see Eq. 14).

$$A_n = \frac{A - A_{\min}}{A_{\max} - A_{\min}} \quad (14)$$

Where  $A_n$  is the normalized value of a feature value,  $A$  is the feature value,  $A_{\max}$  and  $A_{\min}$  represents the maximum and minimum values reported for the considering feature

Among the standardized features in both the balanced and balanced datasets, the subset of input features that are most relevant to the target variables (classes) was selected using the ANOVA (Analysis of Variance) f-test feature selection method. Specifically, the entire training dataset was subjected to the ANOVA f-test feature selection algorithm, and the three features that performed minimum scores on the test (i.e., features that are primarily independent of the target variable) were excluded from each dataset (see Fig. 2). The remaining features were then used in the subsequent K-fold cross-validation experiment to identify the most promising machine learning (ML) algorithm for each dataset.

Parameter	Status
Bootstrap	True
ccp_alpha	0.0
class_weight	None
Criterion	Gini
max_depth	None
max_features	Auto
max_leaf_nodes	None
max_samples	None
min_samples_leaf	1
min_samples_split	2
min_weight_fraction_leaf	0.0
n_estimators	100
n_jobs	None
oob_score	False
random_state	None
Verbose	0
warm_start	False

**Table 4.** Default parameters used to build the base model. The base model was developed using the illustrated default parameters and conditions.

To this end, a tenfold ( $K = 10$ ) cross-validation experiment was conducted over both training datasets using several common classification algorithms, including Logistic Regression, Linear Discriminant Analysis, Decision Tree Classifier, Gaussian Naïve Bayes, Support Vector Machine (SVM), K-nearest neighbor (KNN), and Random Forest Classifier. The ML algorithm that yielded the highest cross-validation score was considered as the most promising algorithm to develop the glioma classification model. However, the impact of the application of SMOTE oversampling technique was also examined by comparing the results of the 10-fold cross-validation of both datasets. Here, Python 3.7 along with the scikit-learn library was utilized to standardize the data, ANOVA F-test, SMOTE oversampling, and building and assessing the classification models.

According to the K-fold cross-validation experiment, the Random Forest Classifier was selected as the most promising algorithm to build the glioma grading classification model. The Random Forest Classifier algorithm was trained with the train data by keeping the random state at 42, and the developed classification model (base model) was evaluated using the test set. This developed base model consisted of default parameters (see Table 4), and the performance of the base model was assessed by using the overall accuracy, precision, recall, and F1 scores corresponding to each class. However, the combination of tunable hyperparameters; `min_samples_split`, `n_estimators`, `max_depth`, `bootstrap`, `min_samples_leaf`, and `max_features`, of the developed model was optimized (tuned) using the grid search cross-validation technique on the training set, in order to improve the performance of the model. Here, each hyperparameter was tested within a pre-defined range of values; `n_estimators`: from 100 to 1000 (with the step of 400), `min_samples_split`: (2, 5, 10), `max_depth`: 10 to 110 (with 11 steps), `max_features`: auto, and `bootstrap`: True, False. The tuned model was also evaluated using the test set and the performance of the tuned model was assessed by observing the overall accuracy, precision, recall, and the F1 scores. Also, the performance of the developed model was graphically illustrated using receiver operating characteristic (ROC) curves for each class using the one-vs-rest (OVR) technique and quantitatively measured the performance by calculating the area under the curve (AUC) (see Figs. 3 and 4).

### Ethics approval and consent to participate

The study was approved by two institutional ethics review committees and All the methods were carried out in accordance with relevant guidelines and regulations. (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.

### Data availability

The data that support the findings of this study are available on request from the corresponding author [M. L. Jayatilake]. The data are not publicly available due to them containing information that could compromise research participant privacy, and consent.

Received: 22 November 2022; Accepted: 24 August 2023

Published online: 22 September 2023

### References

1. Goodenberger, M. L. & Jenkins, R. B. Genetics of adult glioma. *Cancer Genet.* **205**, 613–621 (2012).



2. Wang, X. *et al.* Machine learning models for multiparametric glioma grading with quantitative result interpretations. *Front. Neurosci.* **12**, 1046 (2019).
3. Tessamma, T. & Ananda Resmi, S. *Texture Description of Low Grade and High Grade Glioma Using Statistical Features in Brain MRIS* (ACEEE, 2010).
4. Zuckerkandl, E. & Pauling, L. Evolutionary divergence and convergence in proteins. In: *Evolving Genes and Proteins*. 97–166 (Elsevier, 1965).
5. Ostrom, Q. T. *et al.* Cbtrus statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro-oncology* **16**, iv1–iv63 (2014).
6. Maier, S. E., Sun, Y. & Mulkern, R. V. Diffusion imaging of brain tumors. *NMR Biomed.* **23**, 849–864 (2010).
7. Marquet, G., Dameron, O., Saikali, S., Mosser, J. & Burgun, A. Grading glioma tumors using owl-dl and NCI thesaurus. In: *AMIA Annual Symposium Proceedings*. Vol. 2007. 508 (American Medical Informatics Association, 2007).
8. Hilton, D. *et al.* Accumulation of  $\alpha$ -synuclein in the bowel of patients in the pre-clinical phase of Parkinson's disease. *Acta Neuropathol.* **127**, 235–241 (2014).
9. Louis, D. N. *et al.* The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **131**, 803–820 (2016).
10. Joseph, R. P., Singh, C. S. & Manikandan, M. Brain tumor MRI image segmentation and detection in image processing. *Int. J. Res. Eng. Technol.* **3**, 1–5 (2014).
11. Miles, K. Cancer imaging-making the most of your gamma camera. *Cancer Imaging* **4**, S16 (2004).
12. Sarkar, S. D. Benign thyroid disease: What is the role of nuclear medicine? In: *Seminars in Nuclear Medicine*. Vol. 36. 185–193 (Elsevier, 2006).
13. Acampora, A. *et al.* High b-value diffusion MRI to differentiate recurrent tumors from posttreatment changes in head and neck squamous cell carcinoma: A single center prospective study. *BioMed. Res. Int.* **2016** (2016).
14. Kono, K. *et al.* The role of diffusion-weighted imaging in patients with brain tumors. *Am. J. Neuroradiol.* **22**, 1081–1088 (2001).
15. Bammer, R. Basic principles of diffusion-weighted imaging. *Eur. J. Radiol.* **45**, 169–184 (2003).
16. Lansberg, M. G. *et al.* Advantages of adding diffusion-weighted magnetic resonance imaging to conventional magnetic resonance imaging for evaluating acute stroke. *Arch. Neurol.* **57**, 1311–1316 (2000).
17. Nakahara, M., Ericson, K. & Bellander, B. Diffusion-weighted MR and apparent diffusion coefficient in the evaluation of severe brain injury. *Acta Radiol.* **42**, 365–369 (2001).
18. Filippi, M., Cercignani, M., Inglese, M., Horsfield, M. & Comi, G. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology* **56**, 304–311 (2001).
19. Moseley, M. *et al.* Diffusion-weighted MR imaging of acute stroke: correlation with t2-weighted and magnetic susceptibility-enhanced MR imaging in cats. *Am. J. Neuroradiol.* **11**, 423–429 (1990).
20. Thörmer, G. *et al.* Diagnostic value of ADC in patients with prostate cancer: Influence of the choice of b values. *Eur. Radiol.* **22**, 1820–1828 (2012).
21. Sener, R. Diffusion MRI: Apparent diffusion coefficient (ADC) values in the normal brain and a classification of brain disorders based on ADC values. *Comput. Med. Imaging Graph.* **25**, 299–326 (2001).
22. 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. *Invest. Radiol.* **42**, 842–847 (2007).
23. He, X., An, S. & Shi, P. Statistical texture analysis-based approach for fake iris detection using support vector machines. In *International Conference on Biometrics*. 540–546 (Springer, 2007).
24. Lerski, R. A. *et al.* VIII. MR image texture analysis—An approach to tissue characterization. *Magnet. Resonan. Imaging* **11**, 873–887 (1993).
25. Sharma, K., Kaur, A. & Gujral, S. Brain tumor detection based on machine learning algorithms. *Int. J. Comput. Appl.* **103** (2014).
26. Yang, X. *et al.* Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. *Med. Phys.* **39**, 5732–5739 (2012).
27. Shijin Kumar, P. S. & Dharun, V. S. Extraction of texture features using GLCM and shape features using connected regions. *Int. J. Eng. Technol.* **8**, 2926–2930 (2016).
28. Emará-Shabaik, H. E. Nonlinear systems modeling & identification using higher order statistics/polyspectra. In: *Control and Dynamic Systems*. Vol. 76. 289–322 (Elsevier, 1996).
29. Vijithananda, S. M. *et al.* 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*. 189–199 (Springer, 2018).
30. Dean, S. & Illowsky, B. Descriptive statistics: Skewness and the mean, median, and mode. *Connexions website* (2018).
31. Joanes, D. N. & Gill, C. A. Comparing measures of sample skewness and kurtosis. *J. R. Stat. Soc. Ser. D (The Statistician)* **47**, 183–189 (1998).
32. Mohammed, M., Khan, M. B. & Bashier, E. B. M. *Machine Learning: Algorithms and Applications* (CRC Press, 2016).
33. Bishop, C. M. *et al.* *Neural Networks for Pattern Recognition* (Oxford University Press, 1995).
34. Ayodele, T. O. Types of machine learning algorithms. *New Adv. Mach. Learn.* **3**, 19–48 (2010).
35. 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. *J. Magnet. Resonan. Imaging* **31**, 680–689 (2010).
36. Zacharaki, E. I. *et al.* Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnet. Resonan. Med.* **62**, 1609–1618 (2009).
37. Chen, T. *et al.* Detection and grading of gliomas using a novel two-phase machine learning method based on MRI images. *Front. Neurosci.* **15**, 650629 (2021).
38. Vijithananda, S. M. *et al.* Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques. *Biomed. Eng. Online* **21**, 52 (2022).
39. Vamvakas, A. *et al.* Imaging biomarker analysis of advanced multiparametric MRI for glioma grading. *Phys. Med.* **60**, 188–198 (2019).
40. Gupta, N., Bhatele, P. & Khanna, P. Glioma detection on brain MRIs using texture and morphological features with ensemble learning. *Biomed. Signal Process. Control* **47**, 115–125 (2019).
41. Zhang, X. *et al.* Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget* **8**, 47816 (2017).
42. Zulpe, N. & Pawar, V. GLCM textural features for brain tumor classification. *Int. J. Comput. Sci. Issues (IJCSI)* **9**, 354 (2012).
43. Qin, J.-B. *et al.* Grading of gliomas by using radiomic features on multiple magnetic resonance imaging (MRI) sequences. *Med. Sci. Monit.* **23**, 2168 (2017).
44. Rajagopal, R. Glioma brain tumor detection and segmentation using weighting random forest classifier with optimized ant colony features. *Int. J. Imaging Syst. Technol.* **29**, 353–359 (2019).
45. Reza, S. M., Samad, M. D., Shboul, Z. A., Jones, K. A. & Ifekharuddin, K. M. Glioma grading using structural magnetic resonance imaging and molecular data. *J. Med. Imaging* **6**, 024501–024501 (2019).



46. Alis, D. *et al.* Machine learning-based quantitative texture analysis of conventional MRI combined with ADC maps for assessment of idh1 mutation in high-grade gliomas. *Jpn. J. Radiol.* **38**, 135–143 (2020).
47. Han, J., Zhang, Y., Yu, X. & Wang, H. Glioma grading using texture features from diffusion-weighted imaging: A comparison study of machine learning methods. *Med. Sci. Monit.* **24**, 6883–6893 (2018).
48. Soliman, R. K., Essa, A. A., Elhakeem, A. A., Gamal, S. A. & Zaitoun, M. M. Texture analysis of apparent diffusion coefficient (ADC) map for glioma grading: Analysis of whole tumoral and peri-tumoral tissue. *Diagn. Intervent. Imaging* **102**, 287–295 (2021).
49. Citak-Er, F., Firat, Z., Kovanlikaya, I., Ture, U. & Ozturk-Isik, E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3t. *Comput. Biol. Med.* **99**, 154–160 (2018).
50. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
51. Fernández, A., Garcia, S., Herrera, F. & Chawla, N. V. Smote for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* **61**, 863–905 (2018).
52. Mujahid, M. *et al.* Sentiment analysis and topic modeling on tweets about online education during covid-19. *Appl. Sci.* **11**, 8438 (2021).
53. Douzas, G., Bacao, F., Fonseca, J. & Khudinyan, M. Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric smote algorithm. *Remote Sens.* **11**, 3040 (2019).
54. Yang, L. & Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **415**, 295–316 (2020).
55. Alksas, A. *et al.* A novel system for precise grading of glioma. *Bioengineering* **9**, 532 (2022).
56. Ryu, Y. J. *et al.* Glioma: Application of whole-tumor texture analysis of diffusion-weighted imaging for the evaluation of tumor heterogeneity. *PLoS one* **9**, e108335 (2014).

## 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.

## Author contributions

S.M.V., M.L.J., B.H., and T.G. contributed to the conceptualization and designing the study. S.M.V., B.H., K.D.D., and A.D.S. contributed to the data curation process. Formal analysis and interpretation of data were done by S.M.V., T.G., M.L.J., L.M.R., T.D.K., and B.S.W. S.M.V. and T.D.K. involved in designing and creating new image-processing software to extract data from the images, and S.M.V. wrote the main manuscript while all the other authors were involved in reviewing and editing process of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.L.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023