Deep Learning Approach for Glioma Grade Classification with 3D Wavelet Radiomics

CEENA MATHEWS[1]
ANUJ MOHAMED[2]

[1]Department of Computer Science, Prajyoti Niketan College, Pudukad, India
ceenamathews@prajyotiniketan.edu.in
[2]School of Computer Sciences, Mahatma Gandhi University, Kottayam, India
anujmohamed@mgu.ac.in

Abstract. Gliomas are classified as high-grade glioma and low-grade glioma based on the extent of spread. The tumor grade provides essential information about the tumor's aggressiveness and malignancy, assisting physicians in prescribing the appropriate dosage of radiation and chemotherapy. Histopathological analysis of tumor tissue samples, usually obtained through biopsy, is commonly required for brain tumor grading. However, these tissue samples may not completely represent the tumor's heterogeneity, leading to potential sampling bias. To overcome this limitation and avoid the negative impact of biopsy, it becomes crucial to assess the tumor grade directly from MRI scans. Hence in this study, we propose a method that leverages a deep neural network classifier on optimally selected 3D wavelet radiomic features, extracted automatically from multisequence 3D MRI to predict the tumor grade. The proposed method for classifying high-grade glioma and low-grade glioma is evaluated on BraTS 2019 3D MRI dataset using metrics such as F1-score, precision, recall, and accuracy. The proposed method outperforms the conventional machine learning algorithms and also outperforms the state-of-the-art tumor grade classification models.

Keywords: Glioma, Radiomics, Wavelet, Feature Selection, Deep Neural Network.

(Received December 12th, 2024 / Accepted June 28th, 2025)

1 Introduction

Glioma is the most prevalent form of malignant brain tumor, constituting around 80% of all malignant brain tumors [16]. Based on the extent of spread, glioma are classified as high-grade glioma (HGG) and low-grade glioma (LGG). Several treatment procedures such as microsurgical resection, radiotherapy and chemotherapy are used to treat gliomas. The glioma core tissues, are surgically removed and treated with radiation and chemotherapy. For treatments such as radiotherapy and chemotherapy, accurate determination of tumor grade is critical which helps the physician to prescribe the correct dosage for radiation and chemotherapy.

Traditionally, tumor grade assessment involves an invasive procedure known as biopsy, which carries the risk of potential long-term morbidity or even mortality [14]. The histopathology samples obtained through biopsy are utilised to predict the overall tumor grade, with particular focus on the high-grade component. However, due to the inherent heterogeneity of gliomas, histopathology samples taken from different areas of the same tumor may have varying tumor grades. As pathologists typically analyse only a small portion of the tumor for examination, there is a risk that the biopsy sample might not accurately represent the highest-grade component. This phenomenon, known as biopsy sampling error, can potentially lead to incorrect clinical management of the disease. In order to avoid biopsy sampling error and negative consequences of biopsy procedure, it is essential to determine tumor grade non-invasively.

In medical imaging, a paradigm shift has occurred where images are no longer viewed solely for visual interpretation. Brain tumors are normally diagnosed non-invasively from magnetic resonance images (MRI). The brain MRI can also serve as a powerful tool for non-invasive tumor grade assessment. Advanced techniques, such as radiomic feature extraction and deep learning-based methods, are extensively utilised to analyse these images with precision and efficiency. These techniques leverage the power of medical imaging data to extract relevant features and patterns associated with the disease. These non-invasive methods offer a safer and more patient friendly means of assessing tumor grade, contributing to improved treatment, decision-making and patient care. Each approach possesses unique strengths and weaknesses, and the choice between them depends on various factors, such as the research objective, data availability, and computational resources. A combination of both approaches can be used to harness the advantages of each method for feature extraction and analysis from MR images.

Accurate determination of tumor grade using radiomics plays a crucial role in tailoring personalised treatment strategies based on tumor aggressiveness and malignancy. Radiomic features provide quantitative characterisations of tumor tissues, capturing intricate details related to shape, texture, and intensity variations [8]. These features effectively reflect the heterogeneity and complexities of brain tumors, which are essential for accurate classification. These extracted features are then utilised with artificial intelligence for disease prediction, diagnosis, and prognosis.

The radiomic features can be categorised into different types based on their definitions and descriptive characteristics. These categories include statistics-based, model-based, transform-based, morphology-based, and sharpness-based features, each providing valuable insights into the underlying tissue properties. Statistics-based features in radiomics encompass a range of quantitative measures derived from the density histogram of medical images. Among these features are first-order statistics, run-length features, and matrices like neighbourhood gray tone difference matrix (NGTDM) and gray level co-occurrence matrix (GLCM).

First-order statistics features provide essential insights into tumor density distribution. Mean, standard deviation, skewness, and kurtosis are fundamental statistical measures used to describe the distribution of voxel intensities in the image. Run-length features are employed to characterise image coarseness. They count the number of maximum contiguous voxels with the same gray level along a line. Higher values indicate a coarser texture, while lower values indicate a finer texture. The statistics-based features provide useful information about tissue characteristics and identifies subtle changes in texture patterns that can indicate the presence of abnormal tissue. By extracting these texture features, machine learning algorithms can be trained to classify brain tumors into HGG and LGG. The combination of these different texture analysis methods help capture diverse textural characteristics, improving the accuracy of the classification process.

Model-based features provide fractal dimension features. The fractal dimension feature describes the relationship between change in a measuring scale and its resultant measurement value at that scale. The rougher the texture, the larger the fractal dimension. Transform-based features are extracted from transformed images. Gabor filters and wavelets decomposition are two important transformations in the field of digital image processing. Gabor filters are linear filters designed for edge detection. Wavelets decomposition can extract finer/coarser textures at multiple frequency scales. Morphology-based features contain information about tumor size and shape. Shape features can describe both global and local tumors [13]. In sharpness-based features, sigmoid curve fitting feature is used to quantify the density relationship between a tumor and its surrounding background, e.g., sharpness of the tumor margin.

By leveraging artificial intelligence, these quantitative features can be analysed to facilitate precise diagnosis and improve the prognosis of the disease. Utilising the extracted features, predictive models can be developed to forecast the dependent variable(s) of interest, further enhancing our understanding of brain tumor characteristics and contributing to advancements in medical research and patient care [3]. Radiomic features exhibit consistency across diverse datasets, enable rapid implementation for swift analysis, and provide transparent and interpretable outcomes, making them well-suited for medical applications. Moreover, radiomic feature extraction is computationally less demanding compared to deep learning.

The motivation of this study is to non-invasively determine the grade of tumor with accuracy from multisequence 3D MRI. Thus, in this study, we adopt radiomic feature extraction from multisequence 3D MRI as a valuable approach for accurate tumor grade classification. In this work, we automatically extract 3D wavelet filter-based and unfiltered radiomic features using the PyRadiomics 3.0.1 package. To identify the most informative subset of features in the extracted radiomic features, we compared and evaluated the performance of different feature selection methods and the feature selection method that performed well was chosen for feature selection. These selected features are then applied to a deep neural network (DNN) classifier for tumor grade classification into HGG and LGG.

The proposed approach does not require domain expertise for feature extraction and selection, making it accessible to researchers and practitioners. The use of DNNs bring several advantages to our model. By incorporating multiple layers, DNNs can detect complex relationships within the features, even those that might be subtle or hidden. DNNs possess the capability to learn hierarchical representations of features, enabling them to capture increasingly intricate and abstract patterns in the data. This hierarchical representation learning enhances the model's ability to discern important distinctions between tumor grades, ultimately improving the overall discriminative power of the classifier.

2 Related Studies

Deep learning and machine learning techniques are more dominantly used in the classification of brain tumor grade.

[6] proposed a neural network classifier for differentiating between normal and malignant brain MR images. [7] used fuzzy C-means clustering algorithm for classification of different types of tumor.

[10] used logistic regression (LR) to classify HGG and LGG. Four histogram moment features were utilised in this study to describe the global gray-scale distributions of glioma tissues, while 14 textural features were used to evaluate local correlations between neighbouring pixel values. The individual feature set and the combination of both feature sets were utilised to build the malignancy prediction model using LR algorithm. The Cancer Genome Atlas (TCGA) and the Cancer Imaging Archive (TCIA) dataset are used in this work.

[5] used LR to classify HGG and LGG through radiomic features extracted from BraTS 2015 training dataset. In this work, 45 radiomic features based on histogram, shape and GLCM were extracted from each MRI sequence. L1-norm regularisation was used to choose significant features from 180 features.

[20] demonstrated the utility of deep learning to preoperatively grade glioma by using conventional MRI images. The authors used AlexNet and GoogLeNet to classify tumor grade using private dataset containing 113 patients. GoogLeNet performed better than AlexNet with test accuracy of 90.9%.

[1] proposes an improved deep learning framework. In the approach, ResNet50 and DenseNet201, two pretrained deep learning models, are used after the first preprocessing phase. Transfer learning was used to fine tune and train both models. The features are then retrieved from the feature layers. The enhanced ant colony optimisation (EACO) method was used to optimise the retrieved characteristics. The selected characteristics of each network are combined using a serial-based technique, and then they are eventually categorised using multi-class SVM, which employs the cubic method.

[21] proposes 3D brain tumor segmentation based on a modification of the popular U-Net model and mask R-CNN for automatic, non-invasively distinguishing LGG and HGG on BraTS 2018 MRI dataset and TCGA LGG collection dataset.

Since radiomic features offer consistent, interpretable outcomes across datasets, and are computationally efficient compared to deep learning methods, in this study, we adopted automatic radiomic feature extraction from multisequence 3D MRI using PyRadiomics package,

3 Materials and Methods

3.1 Dataset

The BraTS 2019 training dataset includes pre-operative multimodal MRI scans of 335 patients, of which 259 are HGG and 76 are LGG cases. This work uses 3D MR images of 150 HGG patients from the BraTS 2019 training dataset. Each patient case has four MRI sequences such as T1-weighted (T1), T1-weighted with gadolinium contrast (T1Gd), T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR), and ground truth. The ground truths in these datasets are manually segmented and annotated by the experts as background (label 0), NET (NCR/NET) (label 1), ED (label 2), and ET (label 4). Label 3 is not used by the experts.

The dimension of each MRI is (240 x 240 x 155) where 240 x 240 indicates the height and width of a slice and 155 specifies the number of slices. These MRI scans were acquired with different clinical protocols and various scanners from multiple (n=19) institutions. Since the MR images are acquired using different scanners, they are of different resolution. The images are co-registered, skull-stripped, and re-sampled to 1mm3 [15][2].

3.2 Methods

The workflow of the proposed model consists of 4 steps

1. Feature extraction

2. Feature selection using optimal feature selection process
3. Enhancing classification with modified DNN Classifier

Feature Extraction using PyRadiomics

To avoid biopsy sampling error and the negative consequences of biopsy, it is necessary to determine the tumor grade non-invasively from MR images. Techniques such as radiomic feature analysis and deep learning are prevalently used to extract features from MR images. However, in this study, radiomic feature analysis approach is used since these features can be automatically extracted from 3D MRI using PyRadiomics 3.0.1 package, based on the segmented mask. Radiomic feature analysis provide useful information about tissue characteristics and identifies subtle changes in texture patterns that can indicate the presence of abnormal tissue.

Since wavelets decomposition can extract finer/coarser textures at multiple frequency scales, we apply 3D discrete wavelet transform (DWT) filter in PyRadiomics to extract features automatically. Wavelet transforms are useful in image processing to analyse the abrupt change in the image. A 3D DWT is a mathematical tool used to analyse and process signals and images. It decomposes the volumetric images into eight decomposed volumes of images such as LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH where L and H are low and high frequency signals respectively.

Some of the classes of extracted statistics-based features are as follows:

1. Gray Level Co-occurrence Matrix (GLCM) : It is a statistical texture analysis method that quantifies the co-occurence of gray levels between neighbouring pixels in an image. It provides information on the spatial relationships between pixels. GLCM can be used to extract features such as contrast, dissimilarity and homogeneity of an image. These features can be used to analyse the texture of an image. Variations in GLCM features reflect differences in textural patterns, which can help distinguish HGG from LGG.
2. Gray Level Run Length Matrix (GLRLM) : It analyzes the distribution of run lengths, which are consecutive pixels with the same intensity value, in an image. It quantifies the number and length of runs for each gray level, capturing information about the texture's uniformity, coarseness, and directionality. GLRLM features are commonly used for texture classification and segmentation tasks. Differences in GLRLM features can reflect differences in the coarseness, homogeneity, and complexity of the tumor texture [18]. It can be used to extract features such as gray level run-length entropy, short run emphasis, long run emphasis.
3. Gray Level Size Zone Matrix (GLSZM): It focuses on characterising the size and spatial distribution of homogeneous regions or zones within an image. To compute the GLSZM, an image is first converted into a grayscale representation. Then, a thresholding operation is applied to classify the image into different homogeneous regions based on intensity levels. Each region represents a distinct zone in the image. Then the GLSZM is constructed by counting the number of occurrences of each zone size and the corresponding gray level. The GLSZM is essentially a matrix where the rows represent different zone sizes, and the columns represent different gray levels. Common statistical measures derived from the GLSZM include the number of zones, zone size distribution, zone size entropy, and zone size non-uniformity. These measures describe the characteristics of the image texture. HGGs often have larger zones and a higher number of zones compared to LGGs, reflecting the more chaotic and disorganized nature of high-grade tumors.
4. Neighboring Gray Tone Difference Matrix (NGTDM): It quantifies the difference between a gray value and the average gray value of its neighbours within delta distance. It characterises the differences in gray level values between each pixel and its surrounding neighbors. It measures the local variations in intensity within an image. NGTDM features capture textural information related to the distribution of gray level differences, providing insights into the texture's roughness, homogeneity, and complexity.
5. Gray Level Dependence Matrix (GLDM) : It focuses on quantifying the differences between adjacent pixels in an image to extract texture information. Common statistical measures used in GLDM include the mean, variance, entropy, energy, and contrast [17]. They can capture subtle changes in texture patterns, such as variations in texture roughness or coarseness, which may indicate the presence of abnormal tissue or other distinctive features in medical imaging applications. HGGs generally exhibit higher heterogeneity and irregular texture patterns, resulting in different GLDM features compared to LGGs.

6. 3D Shape

Feature Selection

In radiomics analysis, the careful selection of relevant radiomic features is essential to avoid overfitting and enhance prediction accuracy. To mitigate overfitting, it is essential to reduce the number of extracted radiomic features from medical images to a manageable level. Feature selection methods play a critical role in identifying an optimised subset of features that have a significant relationship with the target variable [4]. There are three main types of feature selection methods: filter, wrapper, and embedded methods.

Filter methods evaluate feature relevance or importance based on their intrinsic characteristics, independent of any specific machine learning algorithm. Statistical measures like Pearson's Correlation, chi-square, and ANOVA are commonly used to rank or select features. Wrapper methods, on the other hand, employ a machine learning approach to generate a subset of features. Techniques like forward feature selection (FFS) and recursive feature elimination (RFE) are examples of wrapper methods. FFS is an iterative wrapper technique that builds a predictive model by sequentially adding features based on their importance or relevance [12][11]. Embedded feature selection is a technique that integrates feature selection into the process of training a machine learning algorithm. Unlike filter methods that evaluate features independently or wrapper methods that use an external evaluator, embedded methods perform feature selection as part of the model training process. Embedded method is a combination of filter and wrapper methods. LASSO (Least absolute shrinkage and selection operator) and RIDGE are examples of embedded methods.

In order to determine the best feature selection technique, in this study we first compare the performance of some of the feature selection techniques on the extracted radiomic features.

Enhancing Classification with Modified Deep Neural Network Classifier

In order to leverage the advantages of deep learning methods, we employ a DNN classifier to classify tumor grade based on optimal feature subset chosen from the radiomic features extracted from brain MR images. The architecture of the DNN, illustrated in Figure 1, comprises four hidden layers with 64, 32, 32, and 16 hidden units, respectively, in addition to the output layer. These hidden layers utilise the rectified linear unit (RELU) activation function, while the output layer employs the sigmoid activation.

4 Results and Discussion

In this work, we adopted a novel approach to classify brain tumor grade into HGG and LGG. In the proposed method, we use PyRadiomics 3.0.1 python package for extracting wavelet filtered radiomic features from multisequence 3D MR images based on the segmented task. It calculates single value per feature for the segmented mask. Features are extracted from T1, T1Gd, T2 and FLAIR MRI sequences of 259 HGG and 76 LGG patients (1340 samples) in BraTS 2019 dataset. Features are extracted automatically from both unfiltered original images and 3D wavelet based filtered images.

In total, 710 features were obtained based on eight decompositions per level (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH), of 3D wavelet filters and unfiltered images. Out of the 710 features, 22 were diagnostic characteristics such as such as version, applied filters, settings, and original image spacing etc. , 601 were wavelet-based features, and 89 were unfiltered original image features. The extracted radiomic features belong to the classes of unfiltered original image and 3D wavelet filters, such as GLCM, GLRLM, GLSZM, NGTDM and GLDM. All features of these classes are extracted. These features provide useful information about tissue characteristics and identifies subtle changes in texture patterns that can indicate the presence of abnormal tissue. By extracting these texture features, machine learning algorithms can be trained to classify brain tumors into HGG and LGG. The combination of these different texture analysis methods helps capture diverse textural characteristics, improving the accuracy of the classification process. The 22 diagnostic characteristics in the extracted feature set, are removed from the analysis since they do not offer relevant contextual information about the tumor. Subsequently, we included the target label "grade" with values HGG and LGG for further classification.

To fully harness the potential of deep learning, we utilise a DNN classifier to assess and classify tumor grade. The DNN is configured using the Adam optimiser with a learning rate of $1e^{-3}$ and binary cross-entropy loss function. To evaluate the performance of the tumor grade classifier model, various evaluation metrics such as F1-score, precision, recall, area under curve (AUC), and accuracy are used, providing comprehensive insights into the model's effectiveness and predictive capabilities.

The DNN classifier was trained and tested using the extracted features without applying any feature selection technique, and achieved a classification accuracy of

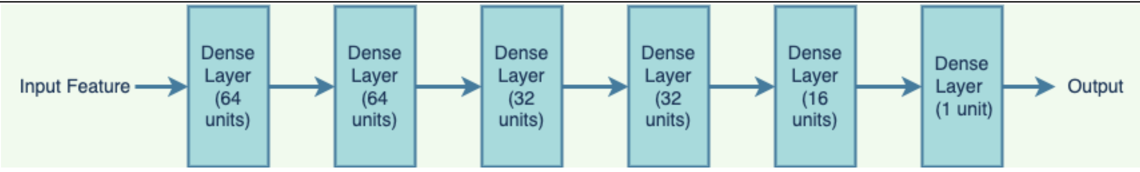

Figure 1: Modified Deep Neural Network (DNN) classifier

85.44%. To further improve classification accuracy, it is essential to reduce the number of extracted radiomic features from medical images to a manageable level. Therefore in this study to enhance the performance of tumor grade classification, radiomic features extracted using 3D wavelet-filter in PyRadiomics are subjected to feature selection. Feature selection techniques are used to choose the optimal subset of features based on the target variable.

In order to determine the best feature selection technique, in this study we first compare the performance of some of the feature selection techniques on the extracted radiomic features. The extracted features are preprocessed by normalisation such that each feature/variable will have mean = 0 and standard deviation = 1. The target labels are encoded with values 0 for HGG and 1 for LGG. In all feature selection techniques except for LASSO and RIDGE, we tried with different set of features (50, 100 and 150). The number of features to select cannot be specified in LASSO and RIDGE. In the experiment, we have used the value of the hyper parameter α as 0.0001 and 0.1 in LASSO and RIDGE respectively. We applied LR algorithm for classification. It is apparent from the Table 1 that FFS method with 150 feature selection, performs better except for precision metric, than the other feature selection techniques. Therefore we employed the FFS technique for feature selection.

FFS is an iterative method that begins with zero features and progressively adds the features that contribute the most to the model's accuracy. This iterative process continues until adding a new variable no longer improves the model's performance. FFS utilises an induction algorithm in conjunction with a statistical resampling technique, such as cross-validation, to estimate the final accuracy of the selected feature subsets. This approach ensures that the chosen feature subset is optimised for accurate tumor grade classification [9]. The process of FFS typically involves the following steps:

1. Start with an empty set of selected features.
2. Iterate through the remaining features that have not been selected yet.
3. Along with each candidate feature, evaluate the performance of the model using the current set of selected features .
4. Select the feature that improves the model's performance the most.
5. Add the selected feature to the set of selected features.
6. Repeat steps 2-5 until a stopping criterion is met (e.g., reaching a predetermined number of features or a certain level of performance improvement).

Further we explored several induction algorithms with FFS method as shown in Table 2, such as random forest (RF), support vector classifier (SVC), stochastic gradient descent (SGD), and LR, within the FFS method. In the experiment with FFS using RF, we specifically use an RF classifier with 100 decision trees, leveraging out-of-bag samples to estimate prediction error. Additionally, we experimented with other algorithms, including LR with SGD cost function, linear SVC with regularisation parameter C set to 1, and SGD classifier.

We also conducted experiments to determine the number of features to be selected, considering options like 100, 139, and 150 features, to achieve an improved classification result. However, due to the computational intensity of the RF and LR algorithms used in the FFS approach, we limit the exploration to 150 features. The RF-based FFS approach took approximately 17 hours to select the optimal subset of 150 features for tumor grade prediction. The results in Table 2 clearly indicates that the DNN model achieves superior classification outcomes when utilising 150 features selected with induction algorithm RF, in the FFS process. The selected subset of 150 features contains 14 shape-based and 136 texture-based features, obtained from both unfiltered and wavelet-filtered images. The texture-based features consist of 54 GLCM, 23 GLDM, 29 GLRLM, 23 GLSZM, and 7 NGTDM.

Figure 2 displays the AUC obtained by the DNN model using features selected with different induction algorithms (RF, LR, SVC, and SGD) in the FFS mechanism. The ROC curve illustrates the model's ability to differentiate between the HGG and LGG classes. A higher ROC curve (closer to the top-left corner of the

Table 1: Comparison of feature selection techniques on the 3D wavelet filter based radiomic features

Feature Selection	No. of Features selected	F1-score	Precision	Recall	Accuracy	AUC
SelectKBest	150	0.6434	0.7115	0.5873	0.8470	0.76
SelectKBest	100	0.5714	0.6530	0.5079	0.8208	0.71
RFE	150	0.6610	0.7090	0.6190	0.8507	0.77
RFE	100	0.6153	**0.7804**	0.5079	0.8507	0.73
SelectPercentile	172	0.6722	0.7142	0.6349	0.8544	0.78
FFS	150	**0.7301**	0.7301	**0.7301**	**0.8731**	**0.82**
LASSO	139	0.6611	0.7272	0.4782	0.8470	0.77
RIDGE	250	0.6557	0.6779	0.5942	0.8432	0.77

Table 2: Performance metrics of DNN classifier based on the 150 features selected by FFS technique using different algorithms.

Induction Algorithm used with FFS	F1-score	Precision	Recall	Accuracy	AUC
SGD	0.8811	0.8806	0.6825	0.8843	0.81
SVM	0.8784	0.8775	0.6984	0.8805	0.82
RF	**0.9041**	**0.9044**	**0.7301**	**0.9067**	**0.85**
LR	0.8589	0.8598	0.7142	0.8582	0.81

plot) indicates better discrimination power, suggesting a higher true positive rate while maintaining a low false positive rate. From the graph, it can be observed that the FFS with RF algorithm achieves the maximum AUC value of 0.85.

Figure 3 presents the confusion matrices predicted by the DNN classifier using various induction algorithms employed with FFS feature selection. When using induction algorithm RF, the DNN classifier correctly classifies 197 out of 205 HGG patient cases and 46 out of 63 LGG cases.

The selected subset of features is preprocessed through normalisation, ensuring that each feature/variable has a mean of 0 and a standard deviation of 1. Target labels are encoded with 0 for HGG and 1 for LGG. The dataset is split into an 80:20 ratio for training and testing the model. Preprocessed features are then trained using a DNN classifier with Adam optimiser and binary cross-entropy cost function, utilising a learning rate of $1e^{-3}$.

Additionally, the performance of DNN classifier is compared against conventional machine learning classifiers (LR, RF, SVM, SGD) using 150 features. Table 3 shows that the DNN classifier outperforms all conventional machine learning classifiers in all evaluation metrics.

Table 4 compares the performance of our proposed approach with state-of-the-art methods for classifying LGG and HGG tumors. State-of-the-art methods use different feature extraction methods such as radiomics and deep learning, and each of them are evaluated on different datasets and different image types such as volumetric, multimodal. The proposed method uses radiomic feature extraction method and are evaluated on BraTS 2019 multimodal dataset. The proposed approach achieves greater accuracy of 94% against all state-of-the-art methods.

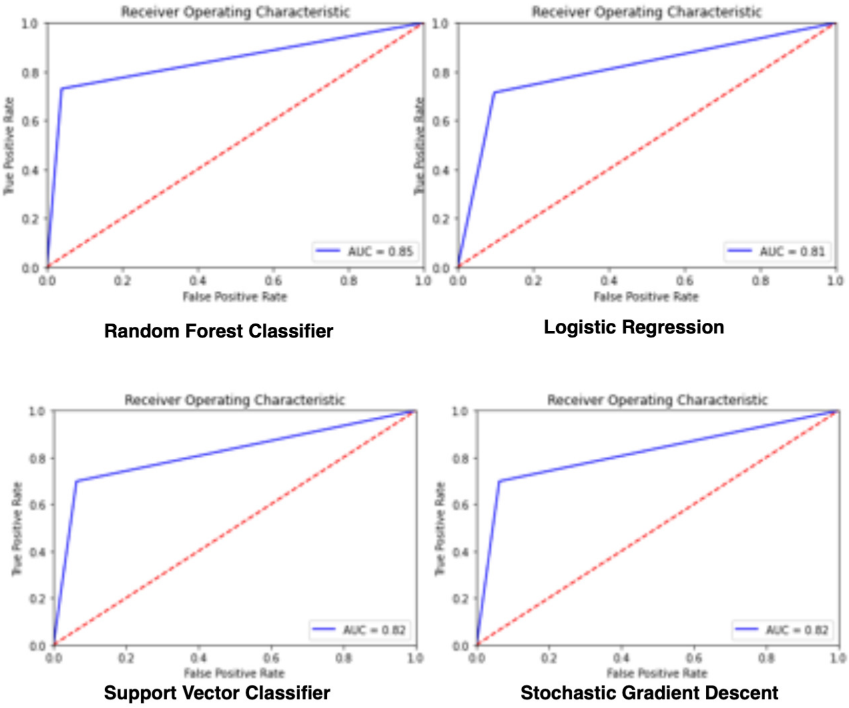

Figure 2: Area under Curve(AUC) of DNN model with RF, LR, SVC and SGD as induction algorithms in FFS

Table 3: Classification accuracy of DNN classifier against conventional machine learning classifiers

Classifier Model	F1-score	Precision	Accuracy	Recall	AUC
SGD	0.7448	0.7105	0.8619	0.7826	0.8360
LR	0.7580	0.8545	0.8880	0.6811	0.8204
SVC(kernel=linear)	0.8244	0.8709	0.9141	0.7826	0.8712
RF	0.6285	0.9166	0.8544	0.4782	0.7315
Proposed DNN	**0.8787**	**0.9206**	**0.9402**	**0.8405**	**0.9077**

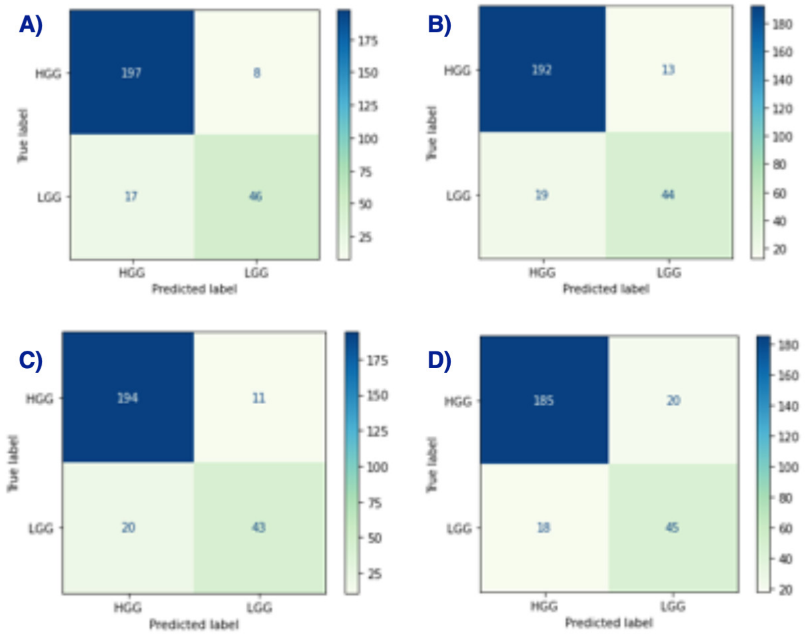

Figure 3: Confusion Matrix of the DNN classifier based on features selected using FFS with various induction algorithms.A) RF classifier, B) SVM Classifier, C) SGD classifier, and D) LR.

Table 4: Performance measure of DNN classifier against existing state-of-the-art machine learning classifiers on test dataset

Classifier	Dataset	Image	Feature Extraction	Accuracy (%)
[5]	BraTS 2017 (285 Patients)	Volumetric (Multimodal)	Radiomics	89
[10]	TCGA (107 Patients)	Sliced (T1-weighted)	Radiomics	88
[20]	Private (113 Patients)	Sliced T1-weighted	Deep Learning	**91**
[1]	BraTS 2019 (271 Patients)	Sliced (Multimodal)	Deep Learning	86.2
Proposed method	BraTS 2019 (335 Patients)	Volumetric (Multimodal)	Radiomics	**94**

5 Conclusion

In this study, we automatically extracted radiomic features from multisequence volumetric MR images in the BraTS 2019 benchmark dataset using PyRadiomics 3.0.1 with a 3D wavelet filter. Radiomic features are preferred due to their consistent, interpretable outcomes across datasets and computational efficiency compared to deep learning methods. The DNN classifier was trained and tested using these extracted features without feature selection, achieving a classification accuracy of 85.44%. To enhance classification accuracy, we applied feature selection methods and selected the forward feature selection (FFS) technique for our proposed method. We tested various induction methods (RF, SVC, SGD, LR) for selecting the optimal feature subset from the extracted features, trying different feature numbers (50, 100, 150). With 150 features selected using the RF-based FFS technique, the DNN classifier achieved better results. Due to class imbalance in the experimented dataset (76 LGG and 259 HGG instances), the classifier could achieve only a classification accuracy of 94%. The DNN classifier outperformed conventional machine learning classifiers and achieved better accuracy in comparison with the state-of-the-art methods. Training the model with more samples from both tumor grades may further enhance its performance.

References

[1] Aziz, A., Attique, M., Tariq, U., Nam, Y., Nazir, M., Jeong, C.-W., Mostafa, R. R., and Sakr, R. H. An ensemble of optimal deep learning features for brain tumor classification. *Computers, Materials & Continua*, 69(2), 2021.

[2] Bakas, S. et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. nature sci. data 4, 170117 (2017), 2017.

[3] Bi, Q., Goodman, K. E., Kaminsky, J., and Lessler, J. What is machine learning? a primer for the epidemiologist. *American journal of epidemiology*, 188(12):2222–2239, 2019.

[4] Chandrashekar, G. and Sahin, F. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

[5] Cho, H.-h., Lee, S.-h., Kim, J., and Park, H. Classification of the glioma grading using radiomics analysis. *PeerJ*, 6:e5982, 2018.

[6] Gadpayleand, P. and Mahajani, P. Detection and classification of brain tumor in mri images. *International Journal of Emerging Trends in Electrical and Electronics, IJETEE–ISSN*, pages 2320–9569, 2013.

[7] Ghosh, D. and Bandyopadhyay, S. K. Brain tumor detection from mri image: An approach. *IJAR*, 3(6):1152–1159, 2017.

[8] Gillies, R. J., Kinahan, P. E., and Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

[9] Hall, M. A. Correlation-based feature selection for machine learning. 1999.

[10] Hsieh, K. L.-C., Lo, C.-M., and Hsiao, C.-J. Computer-aided grading of gliomas based on local and global mri features. *Computer methods and programs in biomedicine*, 139:31–38, 2017.

[11] Kohavi, R. *Wrappers for performance enhancement and oblivious decision graphs*. PhD thesis, stanford university, 1996.

[12] Kohavi, R. and John, G. H. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997.

[13] Li, R., Xing, L., Napel, S., and Rubin, D. L. *Radiomic and Radiogenomics: technical basis and clinical applications*. CRC Press, 2019.

[14] Malone, H., Yang, J., Hershman, D. L., Wright, J. D., Bruce, J. N., and Neugut, A. I. Complications following stereotactic needle biopsy of intracranial tumors. *World neurosurgery*, 84(4):1084–1089, 2015.

[15] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[16] Ostrom, Q. T., Gittleman, H., Fulop, J., Liu, M., Blanda, R., Kromer, C., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. S. Cbtrus statistical report: primary brain and central nervous system tumors diagnosed in the united states in 2008-2012. *Neuro-oncology*, 17(suppl_4):iv1–iv62, 2015.

[17] Sun, C. and Wee, W. G. Neighboring gray level dependence matrix for texture classification.

Computer Vision, Graphics, and Image Processing, 23(3):341–352, 1983.

[18] Tustison, N. and Gee, J. Run-length matrices for texture analysis. *Insight J*, 1:1–6, 2008.

[19] Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J.-C., Pieper, S., and Aerts, H. J. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.

[20] Yang, Y., Yan, L.-F., Zhang, X., Han, Y., Nan, H.-Y., Hu, Y.-C., Hu, B., Yan, S.-L., Zhang, J., Cheng, D.-L., et al. Glioma grading on conventional mr images: a deep learning study with transfer learning. *Frontiers in neuroscience*, 12:804, 2018.

[21] Zhuge, Y., Ning, H., Mathen, P., Cheng, J. Y., Krauze, A. V., Camphausen, K., and Miller, R. W. Automated glioma grading on conventional mri images using deep convolutional neural networks. *Medical physics*, 47(7):3044–3053, 2020.