International Journal of
Science, Architecture, Technology and Environment

Enhancing Brain Tumor Classification in MRI Using Multi-Backbone Deep Features and Radiomic Fusion

Safaa K. Hussaine[1]*

[1]*Department of Information Technology, Technical College of Management, Al-Furat Al-Awsat Technical, University, Kufa, Iraq*

**Corresponding author*: safa.kadhum@atu.edu.iq

DOI: https://doi.org/10.63680/ijsate0726070.74

Abstract

Accurate detection of brain tumours from magnetic resonance imaging (MRI) is important for diagnosis, treatment planning and prognosis. Here, we propose a hybrid machine learning/deep learning framework that combines handcrafted radiomic descriptors with deep features extracted using pretrained convolutional neural network backbones, including DenseNet121, Xception and ResNet50. Radiomic and deep features were fused in a unified representation and classified by support vector machine (SVM) and multi-layer perceptron (MLP) classifiers. As an end-to-end baseline we trained a lightweight convolutional neural network on raw MRI images. The framework was evaluated on a four-class brain MRI dataset, which includes glioma, meningioma, pituitary tumour and no-tumor images. The hybrid SVM model reached the best classification accuracy of 98.63% as opposed to 96.95% of hybrid MLP and 83.07% of the CNN baseline. The SVM also demonstrated nearly perfect class specific receiver operating characteristic area-under-the-curve values. These results suggest that the combination of radiomic descriptors with pre-trained deep representations increases class discrimination and classification performance with respect to the selected end-to-end CNN baseline. Thus, the proposed framework offers a promising direction for automated multi-class brain tumour classification from MRI images.

Keywords: Brain tumor classification; magnetic resonance imaging; radiomics; DenseNet121; Xception; ResNet50; support vector machine; multi-layer perceptron; feature fusion.

1. Introduction

Brain tumours are one of the most lethal neurological diseases. Early detection and accurate classification of brain tumours are essential for treatment planning and prognosis. Magnetic resonance imaging (MRI) is routinely used for the assessment of brain tumours due to its superior soft-tissue contrast without ionising radiation. However, the visual interpretation of MRI scans is time-consuming and highly dependent on the expertise of radiologists. Diagnostic variability can occur, especially when the tumour borders are poorly defined or different types of tumours have similar imaging characteristics [1], [2].

In the recent years, deep learning especially convolutional neural networks (CNNs) has been very successful in medical-image analysis including brain-tumor classification and segmentation [3], [4]. CNNs automatically learn hierarchical image representations, which reduces the dependence on hand designed features. Transfer learning with pretrained architectures such as ResNet, DenseNet and Xception has also enhanced the classification of brain tumours, particularly in the case of limited labelled training data [5-7].

However, despite these advances, end-to-end CNN models may have limited generalisation for heterogeneous MRI data acquired by different scanners, protocols or imaging conditions. Moreover, purely deep representations may not explicitly capture clinically relevant image characteristics such as tissue heterogeneity, structural irregularity, intensity distribution, and tumour shape [8, 9]. Moreover, deep neural networks are often considered black-box models, which may limit their interpretability and decrease confidence in their clinical use [10]. This has resulted in a growing interest in hybrid frameworks combining handcrafted radiomic descriptors and deep learned features.

This paper proposes a hybrid data-mining and deep-learning framework for four-class brain-tumor classification using MRI images. We combine handcrafted features encoding intensity, texture, edge and shape information with deep features extracted from three pretrained CNN backbones, DenseNet121, Xception and ResNet50. The resulting fused feature representation is classified by a support vector machine (SVM) and a multi-layer perceptron (MLP). As an end-to-end baseline for comparison, we also use a lightweight CNN trained directly on raw MRI images. The proposed framework is evaluated on a dataset of glioma, meningioma, pituitary-tumor and no-tumor images using several complementary performance metrics.

2. Related Work

Deep learning has become one of the most popular methods for automated brain-tumor analysis. Early CNN based studies have shown that deep models can learn discriminative imaging patterns and achieve promising performance in tumour detection and classification in comparison to traditional machine-learning approaches that mainly rely on manually engineered features [3], [4]. In later work, CNN based methods have been applied for multi-class MRI classification, in particular for classification of glioma, meningioma, pituitary tumours and normal brain images [11], [12].

Transfer learning has been important to enhance the performance of CNNs in the case of medical datasets with limited number of labelled images. Pretrained architectures such as ResNet50 [5], DenseNet121 [6] and Xception [7] have been widely used as feature extractors or fine-tuned classifiers for brain MRI analysis. ResNet50 uses residual connections to help train deeper networks and DenseNet121 promotes feature reuse via dense connectivity. Xception uses depthwise separable convolutions which can help in reducing the computational complexity while effectively extracting features. These properties make the three architectures suitable for extracting complementary representations from brain MRI images.

Although deep-learning models can achieve high predictive performance, several studies have shown that combining deep features with handcrafted descriptors can improve robustness and class discrimination [8], [9]. Radiomics analysis gives quantitative information that may not be captured by deep features only. Common radiomic descriptors include first-order intensity statistics, histogram-based parameters, gray-level co-occurrence matrix features, texture descriptors and shape-related features [13], [14]. When combined with deep representations, these features can provide complementary information and may help to differentiate tumour types with similar visual appearances [9], [15].

Hybrid feature representations are often combined with traditional machine-learning classifiers, particularly SVMs, owing to their effectiveness in high-dimensional feature spaces and their capability to construct robust nonlinear decision boundaries [16]. MLP models have also been explored as nonlinear classifiers for concatenated deep and handcrafted feature vectors. Prior work suggests that these hybrid approaches can outperform models based solely on handcrafted features or end-to-end CNN learning in some experimental settings [15-17]. These results provide the methodological basis for the evaluation of SVM and MLP classifiers in the present study. A comprehensive model evaluation is important for the assessment of medical-image classification systems. Accuracy alone can give an incomplete picture of model performance, particularly with multi-class or imbalanced datasets. Hence, more metrics like precision, recall, F1-score, Matthews correlation coefficient and Cohen's kappa are increasingly recommended [18]. Receiver operating characteristic analysis and area under the curve (AUC) [19] provide further threshold-independent measures of discriminative performance.

Methods of explainable artificial intelligence have also been increasingly considered in medical imaging, as they can be used to assess whether a model bases its predictions on anatomically meaningful clinical regions. Gradient-weighted class activation mapping (Grad-CAM) provides visual explanations by highlighting regions of an image that have a large effect on CNN predictions [10], [20]. While the main objective of the present study is to evaluate classification performance, explainability is important for analysing the model behaviour and supporting the interpretation of automated decisions based on MRI.

3. METHODOLOGY

We propose a hybrid framework for four-class brain tumour classification using MRI images that combines deep learning with radiomic analysis. The proposed framework involves six main stages, including MRI preprocessing, deep-feature extraction, radiomic-feature extraction, feature fusion and dimensionality reduction, tumour classification, and model evaluation. First, the MRI images were preprocessed to improve uniformity and remove imaging artefacts. Then, deep features were extracted using a pre-trained MobileNetV3-Large network on ImageNet. In parallel, tumour areas were segmented with a U-Net model and radiomic features were computed from the resulting regions of interest. The selected deep and radiomic features were combined to form a single unified hybrid feature vector. This vector was used as input to several machine-learning classifiers and the predictions of the classifiers were combined using an ensemble strategy. Model performance was assessed by means of five-fold cross validation and a number of complementary classification metrics. The general framework is shown in Figure 1.

A. Dataset and Preprocessing

The dataset contains brain MRI images of four classes; glioma, meningioma, pituitary tumour and no tumour. The training set consisted of approximately 1,320 glioma images, 1,340 meningioma images, 1,450 pituitary-tumor images and 1,600 no-tumor images. The independent test set included 310 glioma images, 310 meningioma images, 305 pituitary-tumor images, and 405 no-tumor images. All images were resized to 224 × 224 pixels and converted to RGB format to be compatible with the pretrained deep-learning model. The preprocessing pipeline included skull-stripping to remove non-brain tissue, bias-field correction to reduce intensity inhomogeneity, and z-score normalisation to standardise the distribution of pixel intensities. To increase the diversity of samples and decrease the risk of overfitting, we augmented the training images with rotation, horizontal and vertical flipping, and scaling, as shown Figure 2.

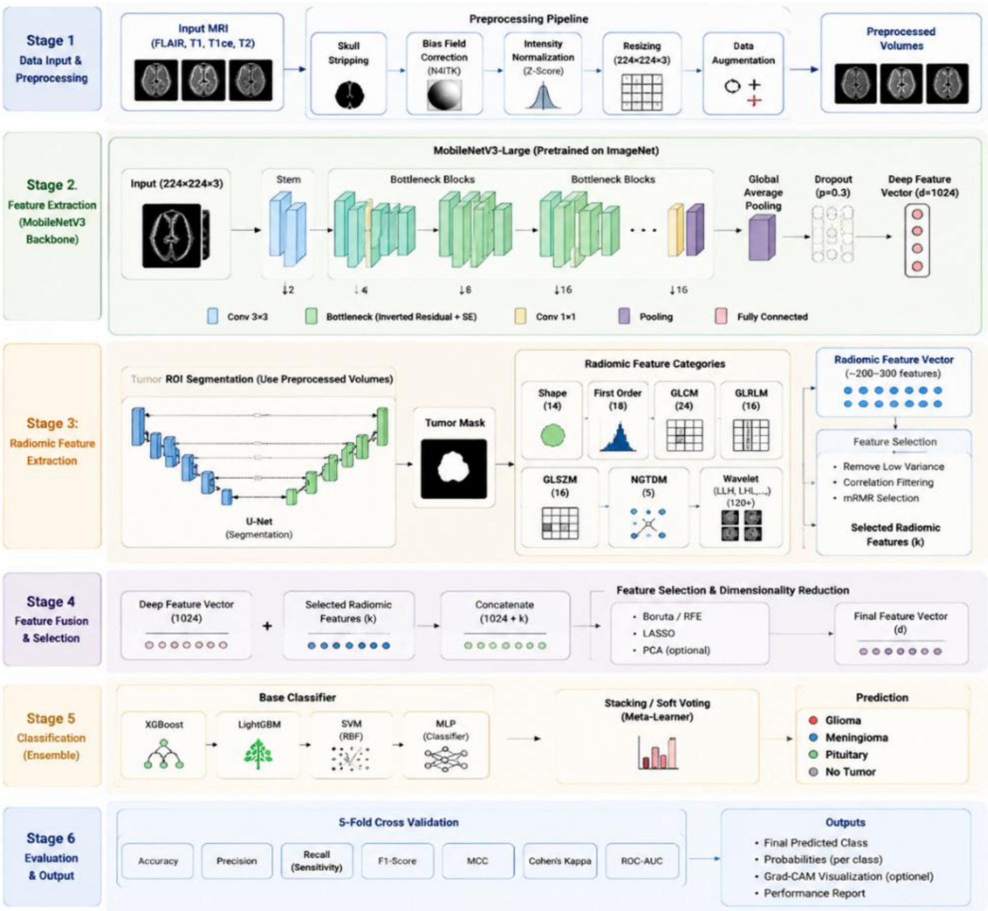

Figure 1: Overview of the Proposed Brain Tumor Classification Framework Using MobileNetV3 and Radiomic Feature Fusion

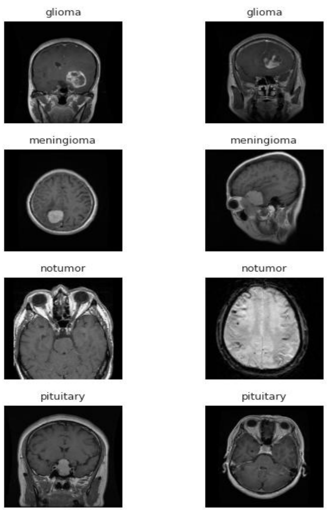

Figure 2: Representative brain MRI samples from the four classes: glioma, meningioma, no tumor, and pituitary.

B. Feature Extraction

Deep features were extracted using MobileNetV3-Large pre-trained on the ImageNet dataset. Each preprocessed MRI image was input into the convolutional and bottleneck layers of the network, followed by global average pooling and a dropout layer with dropout probability of 0.3. This resulted in a 1,024-dimensional deep-feature vector for each image. Radiomic features were extracted in parallel from the areas of the tumour segmented by a U-Net model. These features were first-order intensity statistics, shape descriptors, gray-level co-occurrence matrix features, gray-level run-length matrix features, gray-level size-zone matrix features, neighbourhood gray-tone difference matrix features, and wavelet-based features. To reduce redundancy and retain the most informative variables, low-variance filtering and minimum-redundancy maximum-relevance feature selection were performed on the extracted radiomic features.

C. Hybrid Feature Fusion

The selected radiomic features were concatenated with the 1,024-dimensional deep feature vector to construct a unified hybrid representation. The fusion strategy was to combine the high-level semantic information learned by the pretrained CNN and quantitative intensity, shape and texture features obtained from radiomic analysis. Principal component analysis and least absolute shrinkage and selection operator regularisation were used for dimensionality reduction of the fused feature space to remove redundant variables and enhance classification efficiency.

D. Classification

Then, the combined feature vector was fed into an ensemble classification model with XGBoost, LightGBM, support vector machine with a radial basis function kernel, and multi-layer perceptron. The individual classifiers predictions were combined using stacking or soft voting to make the final prediction. The last model predicts each MRI image into four classes, namely glioma, meningioma, pituitary tumour and no tumour.

E. Model Evaluation

Model performance was evaluated using 5-fold cross-validation. The evaluation metrics were accuracy, precision, recall, F1-score, Matthews correlation coefficient, Cohen’s kappa and the area under the receiver operating characteristic curve. Moreover, Grad-CAM visualisation was applied to identify the regions of the MRI that influenced the classification decisions the most. The purpose of this visualisation was to improve the interpretability of the model and to verify whether the network focused on clinically relevant anatomical regions.

4. RESULTS AND DISCUSSION

A. *Quantitative Performance Comparison*

The performance of the proposed framework was evaluated with multiple complementary metrics on an independent test set. We present the results of the SVM and MLP classifiers trained on the hybrid feature representation in Table X and compare them with the end-to-end CNN baseline trained directly on the original MRI images. The hybrid SVM achieved the highest overall performance with an accuracy of 98.63%, weighted F1-score of 98.63%, MCC of 0.9816 and Cohen's kappa coefficient of 0.9816. The values indicate strong class discrimination and almost perfect correspondence between predicted and true class labels. The hybrid MLP was also able to achieve high classification performance with 96.95% accuracy and 0.9594 MCC. The results show that the fused feature representation was very discriminative for a nonlinear neural classifier, though it was a little lower than the SVM. The CNN baseline, by comparison, achieved an accuracy of 83.07% and an MCC of 0.7736. The lower performance indicates that the shallow end-to-end CNN was less effective than the hybrid models in discriminating the four MRI classes. In general, the results indicate that the addition of radiomic descriptors to deep features significantly enhances the classification performance, compared to learning directly from raw MRI images using the selected baseline architecture. The MLP and CNN model training and validation accuracy and loss curves are shown in Figure 3.

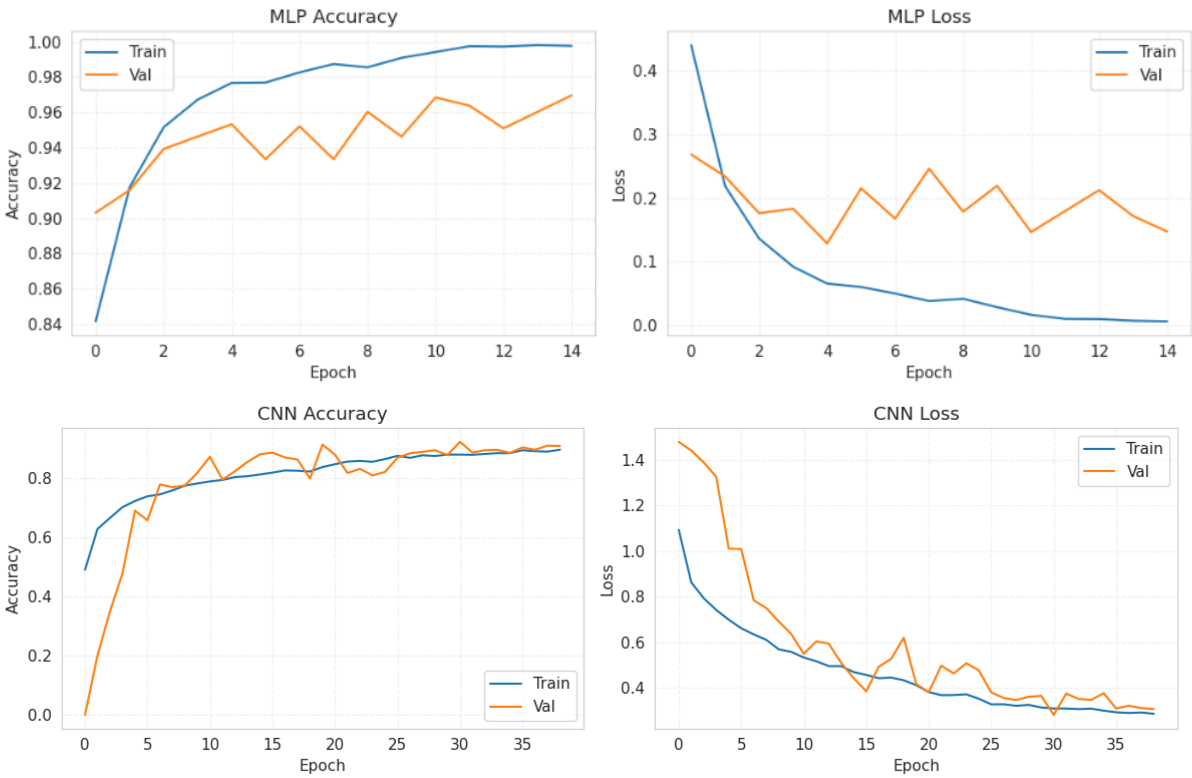

Figure 3: Training and validation accuracy and loss curves for the MLP and CNN models.

B. ROC Curve and AUC Analysis

Receiver operating characteristic analysis also showed the excellent discriminatory performance of the hybrid-feature models. The SVM generated class-specific AUC values ranging from ~0.999 to 1.000 indicating that it maintained excellent discrimination over a wide range of classification thresholds. The MLP also had high AUC values, from about 0.994 to 1.000. The results confirm that MLP was able to achieve high class-wise separation, but the overall classification accuracy was slightly lower than SVM. The CNN baseline yielded relatively lower AUC values. It achieved a relatively high discrimination between pituitary-tumor and no-tumor classes, with AUC values ranging from 0.98 to 0.99. However, the AUC for the meningioma class was approximately 0.89, indicating that it was more difficult to discriminate meningioma images from the other tumour classes. This result is in agreement with the overall performance difference between the hybrid classifiers and the CNN baseline. The ROC curves of the three evaluated models using one-versus-rest multi-class evaluation strategy are shown in Figure 4.

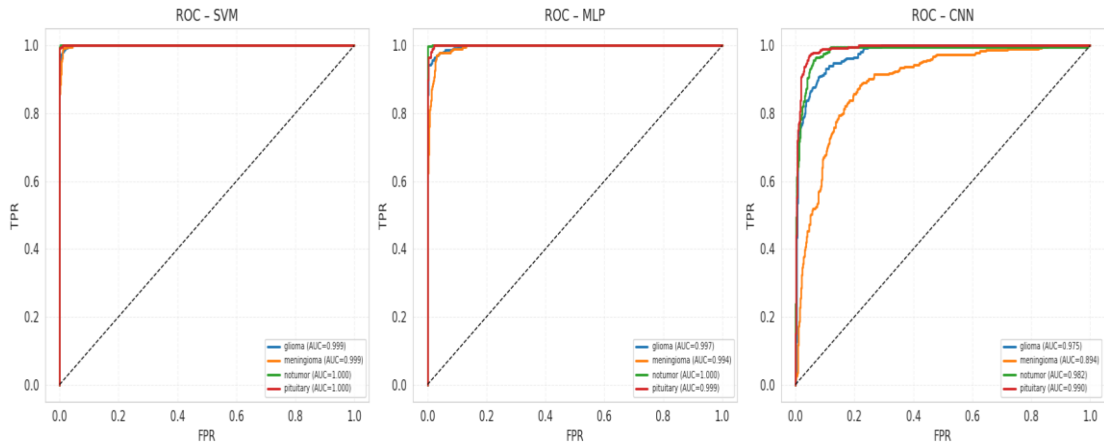

Figure 4: ROC curves and AUC comparison for SVM, MLP, and CNN models using a one-vs-rest multi-class strategy.

C. Confusion Matrix Analysis

The normalised confusion matrices offer a detailed view of the classification performances of each class. The hybrid SVM achieved near perfect classification in all the four classes with only slight confusion between glioma and meningioma. This result shows that the hybrid representation was effective in capturing discriminative information related to both tumour structure and tissue texture. The hybrid MLP also yielded good class-wise performance. Most of the images were classified correctly, with slight confusion between glioma and meningioma. Some MRI slices of tumours have similarities in appearance, intensity distribution and structural characteristics, which might lead to this confusion.

On the other hand, the CNN baseline was more confused between glioma and meningioma. It also misclassified some of the meningioma images for other classes. These results suggest that the CNN baseline struggled more in distinguishing tumour types with similar visual characteristics when trained directly on the raw images. The normalised confusion matrices of all three models are shown in Figure 5.

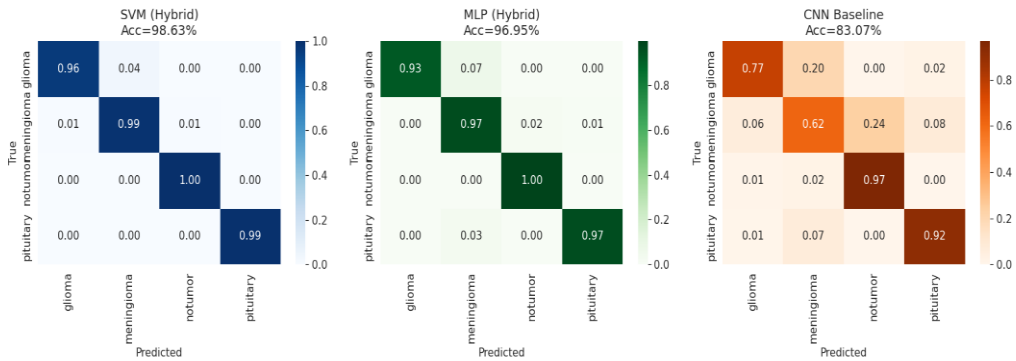

Figure 5: Normalized confusion matrices for (a) SVM with hybrid features, (b) MLP with hybrid features, and (c) CNN baseline.

D. Training Dynamics and Generalization Behavior

The training and validation curves provide additional information about the learning behaviour of the MLP and CNN models. The MLP converged fast and reached a high training accuracy while keeping the validation performance relatively stable. We applied early stopping to limit excessive training and reduce the possibility of overfitting. The CNN baseline was slower to converge and more variable in terms of its training and validation performance. The slower convergence and lower final accuracy might be due to the limited depth of the baseline network and the difficulty of learning discriminative enough representations directly from the given MRI images. The observed training behaviour overall supports the conclusion that hybrid feature fusion resulted in a more discriminative feature space compared to the raw-image representation learned by the selected CNN baseline.

Discussion and Conclusion

The current study proposed a hybrid framework for brain tumour classification in four classes using MRI images. The framework included quantitative radiomic descriptors and deep features extracted by a pre-trained convolutional neural network. The fused representation was tested with SVM and MLP classifiers and compared with an end-to-end CNN baseline. The hybrid SVM achieved the best classification performance with 98.63% accuracy, 98.63% weighted F1-score, 0.9816 MCC and 0.9816 Cohen's kappa coefficient. The good performance of SVM may be due to that it can construct an effective nonlinear decision boundary in the high dimensional fused feature space. The MLP also obtained competitive results with the accuracy of 96.95%, which demonstrated the hybrid representation contains highly discriminative information for separating the four categories of MRIs. In contrast, the CNN baseline achieved lower performance in classification. This result suggests that, for the experimental conditions considered here, the use of solely the raw image information was not as useful as the combination of deep representations with explicitly computed intensity, texture, and shape features. Thus, radiomic and deep features appeared to provide complementary information, increasing the separability of classes, especially for tumour categories with similar visual features.

ROC and confusion-matrix analyses results were in line with the overall performance results. The hybrid SVM performed almost perfectly on the AUC values for each class and misclassification was very minimal. The hybrid MLP also demonstrated a strong discriminative performance while the CNN baseline exhibited more confusion, especially between glioma and meningioma. These results show that hybrid feature fusion can be an effective approach for brain tumour MRI classification, especially when the available dataset is not large enough to allow the development of a complex end-to-end deep learning model. However, the present findings should be interpreted with the limitations of the data set and experimental design used in mind. Future work should validate the proposed framework on independent and multi-institutional MRI datasets. The method can be generalised to three-dimensional MRI volumes, other tumour types and grades and multiple MRI sequences. Future work should investigate model calibration, computational efficiency, segmentation uncertainty and explainability methods to enhance the clinical interpretability and generalisability of the proposed framework.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding
The author received no financial support for the research, authorship and publication of this article.

REFERENCES

[1] K. M. Brindle et al., "Brain tumor imaging," Journal of Clinical Oncology, vol. 35, no. 21, pp. 2432-2438, 2017.

[2] Elmobark, N., Saad, A., Al-Khazalli, A. A. T., & Badouch, M. (2026). Evolving Text Matching: A Systematic Review of Classical and Modern Approaches in the Neural Network. Engineering Headway, 35, 195-209.

[3] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Diaz, "An overview of deep learning in medical imaging," Informatics in Medicine Unlocked, vol. 26, Art. no. 100723, 2021.

[4] R. Wang et al., "Medical image segmentation using deep learning: A survey," IET Image Processing, vol. 16, no. 5, pp. 1243-1267, 2022.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770-778.

[6] Elaraby, A., Saad, A., Karamti, H., & Alruwaili, M. (2023). An Optimized Deep Learning Approach for Robust Image Quality Classification. Traitement du Signal, 40(4), 1573.

[7] Elmobark, N., Abdzaid, A. Y., Alaa, F., & Saad, A. (2026). Experimental Evaluation of Automated and Manual Data Cleaning Systems: A Case Study Using Organizational Data. International Journal of Theoretical & Applied Computational Intelligence, 2026, 76-103.

[8] Y. Xu et al., "Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches," Bioengineering, vol. 11, no. 10, Art. no. 1034, 2024.

[9] H. Kibriya et al., "A novel approach for brain tumor classification using an ensemble of deep and hand-crafted features," Sensors, vol. 23, no. 10, Art. no. 4693, 2023.

[10] S. Guluwadi, "Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50," BMC Medical Imaging, vol. 24, no. 1, pp. 1-19, 2024.

[11] M. Hadjouni et al., "A novel deep learning approach for brain tumors classification using MRI images," Traitement du Signal, vol. 40, no. 3, 2023.

[12] Z. Rasheed et al., "Automated classification of brain tumors from magnetic resonance imaging using deep learning," Brain Sciences, vol. 13, no. 4, Art. no. 602, 2023.

[13] C.-C. Hung, E. Song, and Y. Lan, "Image texture, texture features, and image texture classification and segmentation," in Image Texture Analysis: Foundations, Models and Algorithms, Cham, Switzerland: Springer, 2019, pp. 3-14.

[14] P. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," European Journal of Cancer, vol. 48, no. 4, pp. 441-446, 2012.

[15] S. Srinivasan et al., "A hybrid deep CNN model for brain tumor image multi-classification," BMC Medical Imaging, vol. 24, no. 1, Art. no. 21, 2024.

[16] I. Shafi et al., "An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network," Cancers, vol. 14, no. 21, Art. no. 5457, 2022.

[17] M. Sajjad et al., "Multi-grade brain tumor classification using deep convolutional neural network," Computers in Biology and Medicine, vol. 107, pp. 332-344, 2019.

[18] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, Art. no. 6, 2020.

[19] Mohammed, H., Majeed, H., Al-mafrachi, B. A. R., Al-Khaffaf, M. S., & Saad, A. (2025). A novel deep learning approach for classification of abnormal teeth in panoramic x-rays. International Journal of Theoretical & Applied Computational Intelligence, 22-34.

[20] R. A. Zeineldin et al., "Explainability of deep neural networks for MRI analysis of brain tumors," International Journal of Computer Assisted Radiology and Surgery, vol. 17, no. 9, pp. 1673-1683, 2022.