SciEn

Proceedings of 8th International Conference on Mechanical Engineering and Renewable Energy 2025
10-12 December, 2025, Chattogram, BANGLADESH

SciEn Conference Series: Engineering Vol. 4, 2026, pp 265-271 https://doi.org/10.38032/scse.2026.4.131

Integration of Radiomics and Deep Learning for Advanced MRI-Based Brain Tumor Characterization

Tonmoy Kanti Shaha[1], Sampad Ghosh[1,], Shamsul Enam Faruquei[1], Md Jafrul Hasan[2]*

[1]Department of Electrical and Electronic Engineering, Chittagong University of Engineering & Technology, Chattogram-4349, Bangladesh

[2]Higher Degree by Research (HDR), Torrens University, Melbourne, Australia

ABSTRACT

Accurate brain tumor classification is critical for early diagnosis and effective treatment planning. This study proposes a hybrid framework that integrates handcrafted radiomics descriptors with deep learning features extracted from MRI scans using MobileNetV2. Three models were systematically evaluated: a radiomics-only model, a deep-only model, and a fusion model combining both feature sets. The radiomics-only approach achieved moderate performance with 84.5% accuracy, an F1-score of 0.84, and an AUC of 0.97 on the test set. The deep learning model markedly improved results, attaining 95.9% accuracy, an F1-score of 0.96, and an AUC of 0.996. The proposed fusion model delivered the best outcomes, with 97.1% accuracy, an F1-score of 0.97, and an AUC of 0.998, confirming its robustness and superior discriminative power. These findings demonstrate that combining radiomics with deep representations enhances classification stability and reliability compared to single-modality models. Future research should focus on multi-center validation, integration of clinical data, and explainable AI techniques to advance clinical adoption.

Keywords: Brain tumor classification, Radiomics-deep learning fusion, MRI-based diagnosis.

1. Introduction

Brain tumors represent a significant clinical challenge, accounting for considerable morbidity and mortality across all age groups. Magnetic resonance imaging (MRI) remains the primary modality for non-invasive brain tumor assessment, offering superior contrast resolution and soft-tissue delineation compared with other imaging techniques. Radiomics employs computational techniques to derive large sets of quantitative descriptors from medical images, supporting improved disease characterization, prognostic assessment, and forecasting of therapeutic outcomes [1]. In parallel, deep learning methods, particularly convolutional neural networks (CNNs), have achieved state-of-the-art performance in automated image classification by learning hierarchical feature representations directly from imaging data. Radiomic descriptors are capable of quantifying tumor heterogeneity and subtle textural patterns that often remain imperceptible to visual inspection by clinicians [2]. This study investigates such a hybrid fusion framework, evaluating its potential to advance reliable brain tumor classification using MRI. This automated approach significantly improves diagnostic efficiency compared to traditional clinical assessments [3]. In this study, we present a hybrid classification framework that integrates radiomics features with deep learning representations for brain tumor detection in MRI. Through improved risk stratification, such a model can assist clinicians in optimizing patient management by guiding surveillance planning and tailoring therapeutic interventions [4]. This work examines two fusion strategies that integrate either feature representations or predictive outputs from distinct AI models [5]. An approach employed to assess the practical clinical value of the models [6]. This investigation outlines specific directions for future work, offering guidance for continued advancements in the field [7]. By generating more reliable preoperative predictions, the model may enhance surgical planning and support more effective patient management [8]. The incorporation of radiomic features into deep learning frameworks has broadened the utility of conventional imaging methods across diverse diseases, facilitating the construction of predictive models with enhanced accuracy and stability [9-11]. The central objective of this work is to demonstrate that feature-level fusion not only enhances classification performance but also provides a more stable and interpretable basis for clinical translation.

must be prepared according the guidelines given in this template. This MS-Word document can also be used as a template. Prepare the article with the paper size A4, i.e., 210 mm by 297 mm. The top margin is 22 mm, and the left, right, and bottom margins are of 15 mm. The main text should be written in two-column format with a font of 10pt Times New Roman. The spacings between columns should be 10 mm.

2. Methodology

The study followed a structured experimental framework consisting of dataset preparation, feature extraction, model development, and performance evaluation. A publicly available brain tumor MRI dataset was employed, containing four diagnostic categories: glioma, meningioma, pituitary tumor, and normal brain scans. The dataset was organized into two parts: an augmented set, used for model training and validation, and a Raw set, reserved exclusively for

*Corresponding Author Email Address: sampad@cuet.ac.bd

independent testing. The training subset was further divided into training and validation splits (85:15) using stratified sampling to preserve class balance.

Three classification approaches were investigated: (i) a Radiomics-only model, trained on handcrafted features; (ii) a Deep-only model, using CNN-derived embedding; and (iii) a Fusion model, combining radiomics and deep features at the feature level. For the fusion approach, feature vectors were concatenated, normalized, and reduced using principal component analysis (PCA) prior to classification with a multinomial logistic regression model.

Fig. 1 presents the proposed hybrid classification pipeline. Input MRI scans are divided into training and validation sets from the Augmented dataset, while the Raw dataset is retained for independent testing. Two feature extraction branches operate in parallel: the radiomics branch derives handcrafted descriptors including intensity statistics, GLCM, LBP, Gabor filters, and Hu moments, whereas the deep branch extracts high-level embedding from a pre-trained MobileNetV2 using global average pooling. The resulting features are concatenated, normalized, and reduced via PCA in the fusion stage before classification with a multinomial logistic regression model into four categories: glioma, meningioma, pituitary tumor, and normal brain. The framework outputs predictions alongside evaluation metrics such as confusion matrices, ROC and PR curves, accuracy, macro F1-score, and macro AUC (Area under curve), enabling systematic comparison of radiomics-only, deep-only, and fused approaches.

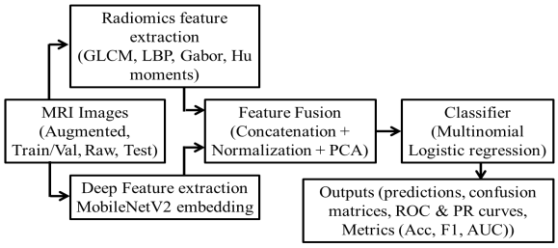

Fig. 1 Proposed hybrid methodology integrating radiomics and deep features.

2.1 Data Acquisition

The study utilized a publicly accessible brain MRI dataset comprising four diagnostic categories: glioma, meningioma, pituitary tumor, and normal brain scans, as shown in Fig. 2. The dataset was distributed in two distinct subsets. The Augmented set, which includes pre-processed and artificially expanded images, was employed for training and validation, while the Raw set was reserved exclusively for independent testing to ensure unbiased evaluation. Each image was standardized to a fixed resolution and formatted consistently before analysis. The dataset structure allowed the development of a training-validation split while maintaining class balance, and it provided a dedicated test cohort to assess model generalization under conditions resembling real-world clinical variability.

2.2 Data Preprocessing

To prepare the raw vibration signals for deep learning analysis, several preprocessing steps were applied to improve signal quality and ensure dataset consistency.

Image Standardization: All MRI scans were resized to a fixed resolution (224 × 224 pixels) and converted to a consistent format to ensure uniformity across radiomics and deep learning pipelines.

Normalization and Enhancement: Pixel intensities were scaled to the [0,1] range for deep learning inputs, while grayscale conversions were applied for radiomics feature extraction. Augmented images were further inspected to confirm quality and eliminate corrupted samples.

Dataset Partitioning: The Augmented dataset was stratified into training (85%) and validation (15%) subsets to preserve class balance, whereas the Raw dataset was reserved entirely for testing to enable unbiased evaluation of model generalization.

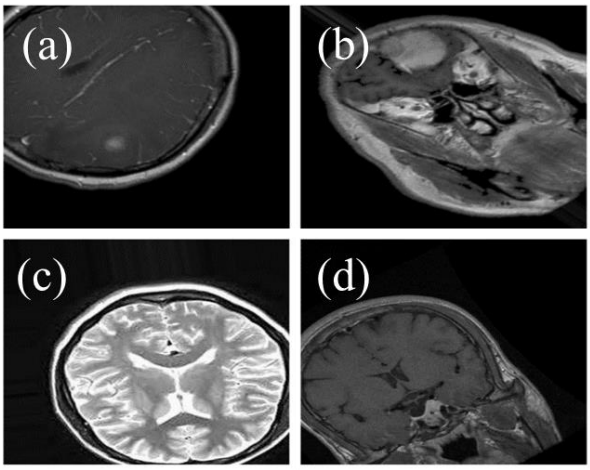

Fig. 2 Dataset for four categories: (a) Glioma, (b) Meningioma, (c) Pituitary tumor, and (d) Normal brain.

2.3 Feature Extraction

To extract meaningful characteristics from raw vibration signals, both time-domain and frequency-domain features were computed for each segmented window. These features help distinguish between healthy and faulty gearbox conditions.

2.3.1 Radiomics Features

First-order statistics: It is computed from the raw pixel intensity distribution of the MRI images, reflecting basic gray-level properties such as mean, variance, skewness, and kurtosis:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} xi \quad (1)$$

$$\sigma^2 = \frac{1}{N}\sum_{1}^{N} (xi - \mu)^2 \quad (2)$$

Texture-based features (frequency-domain features): Texture heterogeneity was captured using descriptors derived from the gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), and Gabor filters.

$$GLCM\ Contrast = \sum_{i,j} (i-j)^2 P(i,j) \quad (3)$$

Where P(i,j) is the normalized co-occurrence probability. The Gabor filter response can be written as

$$g(x,y) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) cos\left(2\frac{x'}{\lambda}+\varphi\right) \quad (4)$$

Where λ is the wavelength, ϕ the phase offset, γ the aspect ratio, and σ the Gaussian spread.

Shape features: Geometric properties were captured using Hu moments, which are invariant to translation, scale, and rotation, ensuring robustness across different MRI orientations.

2.3.2 Deep Features

Deep features were extracted using MobileNetV2 pre-trained on ImageNet. The final classification layer was removed, and a global average pooling (GAP) operation was applied to generate a fixed-length feature embedding from the convolutional layers. These embeddings capture high-level hierarchical patterns such as tumor boundaries, texture irregularities, and global context.

$$F_k = \frac{1}{HW}\sum_{x=1}^{H} \quad \sum_{y=1}^{W} \quad f_k(x,y) \qquad (5)$$

Where H and W denote the spatial dimensions of the feature map.

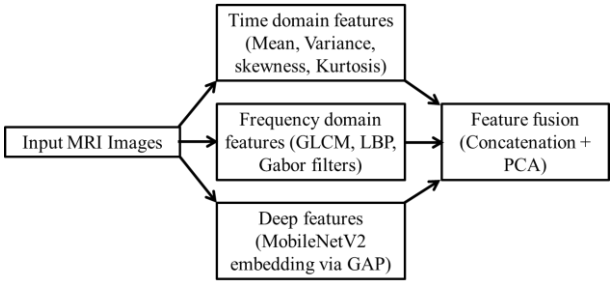

Fig. 3 Feature Extraction Process.

2.4 Data Splitting and Partitioning

To ensure systematic model development and unbiased performance evaluation, the dataset was divided into training, validation, and test subsets. The Augmented images were stratified into training (85%) and validation (15%) splits, preserving class balance during partitioning. These two subsets were used for model training and hyperparameter tuning. The Raw dataset was kept entirely separate and employed as an independent test set. This separation simulated a real-world clinical scenario, where models are trained on augmented data but evaluated on unseen, original scans. Such a partitioning strategy reduces the risk of data leakage and provides a more reliable assessment of model generalizability.

2.5 Model Architectures

This section describes the architecture and operational principles of the deep learning model, radiomics model, and fusion model used in this study for the classification of brain tumor characterization.

2.5.1 Radiomics-Only Model Architecture

The Radiomics-Only Model is constructed entirely on handcrafted features derived from MRI scans. Initially, the images are converted to grayscale and analyzed to generate quantitative descriptors. These encompass first-order intensity measures (mean, variance, skewness, and kurtosis), texture-based attributes obtained from gray-level co-occurrence matrices (GLCM), local binary patterns (LBP), and Gabor filter responses, along with shape-related features extracted through Hu moments.

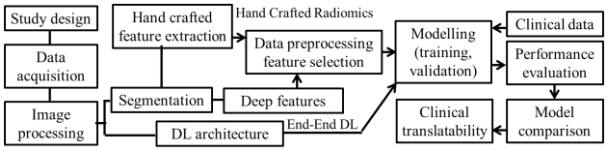

Fig. 4 Basic architecture of the Radiomics-Only Model.

Input Layer: MRI scans are provided in grayscale format to ensure consistent feature extraction.

Radiomics Feature Extraction Layer: Compute descriptors such as mean, variance, skewness, and kurtosis to capture basic intensity distributions. Extracted using methods like gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), and Gabor filters to represent heterogeneity and spatial frequency content. Quantified through Hu moments, which are invariant to rotation, scale, and translation.

Preprocessing Layer (Normalization + PCA): Radiomics features are standardized to a common scale. PCA is applied to reduce dimensionality, retaining the most informative components while minimizing redundancy.

Classification Layer: A multinomial logistic regression model serves as the final classifier. It maps the reduced feature set to four possible diagnostic categories: glioma, meningioma, pituitary tumor, or normal brain.

2.5.2 Deep Learning Model Architecture

Input Layer: MRI scans resized to 224×224×3 were used as inputs. Images were standardized using ImageNet preprocessing to match the pre-trained model requirements.

Initial Convolution and Activation: A 3×3 convolution with 32 filters captures basic image features such as edges and simple textures. This is followed by batch normalization for stable training and a ReLU6 activation function, which constrains activation values between 0 and 6 to improve robustness.

Inverted Residual Blocks: Linear shortcut connections are added when input and output dimensions match, allowing efficient feature reuse. Each block consists of: 1×1 expansion convolution (increases channel depth), 3×3 depth-wise convolution (applies filters separately to each channel, reducing complexity), and 1×1 projection convolution (reduces dimensions back to compact form).

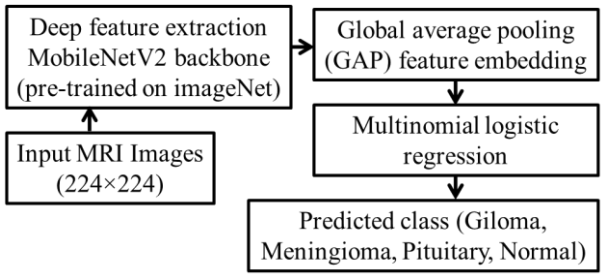

Fig. 5 Basic Architecture of Bi-LSTM Model.

Progressive Feature Extraction: The network stacks multiple inverted residual blocks, gradually increasing the number of filters (16□24□32□64□96□160□320). This progression enables learning from low-level patterns (edges, textures) to high-level semantic features (tumor shapes and structures).

Final Convolutional Stage: A concluding 1×1 convolution aggregates features across all channels, preparing them for global summarization.

Global Average Pooling (GAP): Transforms 2D feature maps into a 1D vector by averaging each feature map. Produces a compact embedding that represents the entire image.

External Classifier: This classifier outputs predictions across the four diagnostic categories: glioma, meningioma, pituitary tumor, and normal brain.

2.5.3 Proposed Fusion Model Architecture

A hybrid deep learning model combining CNN and Bi-LSTM was used to analyze vibration signals. The CNN extracts local features, while the Bi-LSTM captures long-term patterns, making the model more effective at identifying subtle gearbox faults.

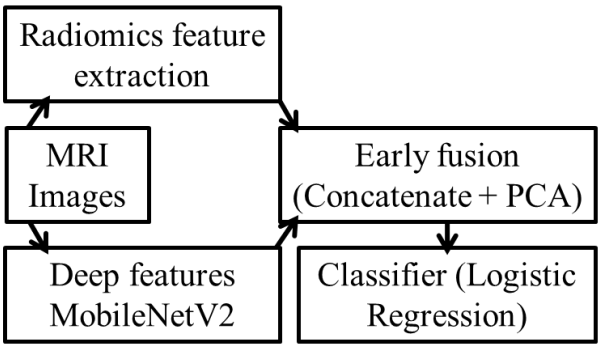

Fig. 6 Basic Architecture of Fusion (Radiomics+ MobileNetV2) Model.

Input MRI Images: Preprocessed brain MRI scans resized to 224×224 pixels and normalized for consistency.

Radiomics Feature Extraction: Handcrafted descriptors computed, including first-order intensity statistics, Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Gabor filters, and shape descriptors.

Deep Feature Extraction: A pre-trained MobileNetV2 network (with ImageNet weights) processes the MRI images.

Feature Fusion: Radiomics and deep feature vectors are concatenated into a single hybrid feature set.

Classification Layer: The fused features are passed into a multinomial logistic regression classifier. The model outputs one of four diagnostic categories: glioma, meningioma, pituitary tumor, or normal brain.

Fully Connecting Layers: Refine features and prepare for classification.

Output Layer: Performance was measured using accuracy, macro F1-score, ROC-AUC, and precision–recall curves.

The fusion model offers a clear advantage by combining handcrafted radiomics features with deep learning embedding, allowing it to capture both interpretable texture-based patterns and complex hierarchical representations. This complementary integration reduces misclassification, improves robustness, and enhances overall predictive accuracy compared to single-modality models. The approach ensures more reliable tumor characterization, as reflected in its superior accuracy, F1-score, and AUC values. The full model, which combines radiomics and deep learning features, is compared against the above two variants. This analysis quantifies the added value of feature fusion, illustrating the synergy between traditional radiomics and modern deep learning techniques in improving classification performance.

2.6 Evaluation Metrics

Accuracy: It shows the overall proportion of correct predictions and can be calculated by dividing (TP+TN) by (TP+TN+FP+FN).

Precision: It tells how many of the predicted faults were actually correct. High precision means fewer false alarms. It is the ratio of TP and (TP+FP).

Recall (Sensitivity): It measures how well the model finds all actual faults. Critical in safety cases where missing a fault is risky. It is the ratio of TP and (TP+FN).

F1-Score: It balances precision and recall into one score. Helpful when the dataset is imbalanced.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Confusion Matrix: The confusion matrix is a tabular representation of predicted versus actual class labels. It provides a breakdown of TP, TN, FP, and FN, helping to identify misclassification patterns and assess class-wise performance. Where TP, TN, FP, and FN indicate True positives, True negatives, False positives, and False negatives, respectively.

3. Results and Discussion

3.1 Radiomics Model Performance

The confusion matrix for the radiomics model was generated by comparing the predicted class labels from the logistic regression classifier, trained on handcrafted radiomics features, against the actual ground truth labels in the independent test set. This provided a class-wise summary of correct predictions and misclassifications.

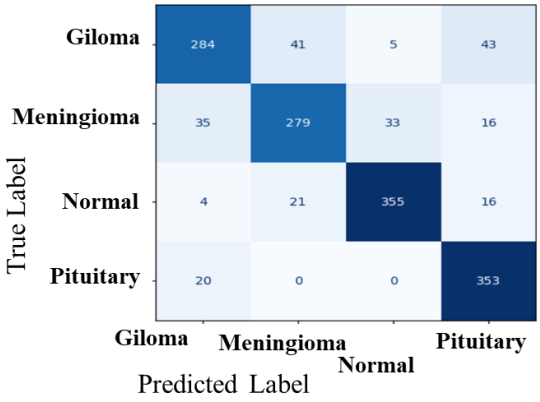

Fig. 7 Confusion matrix of the Radiomics Model.

The Radiomics-only model showed moderate classification performance, with strong accuracy in identifying normal brain and pituitary cases but notable confusion between glioma and meningioma. Specifically, most normal (355/396) and pituitary (353/373) samples were correctly recognized, while glioma (284/373) and meningioma (279/363) exhibited higher misclassification rates, often being mistaken for each other or for pituitary tumors. These results highlight that although radiomics features capture useful image characteristics, they may be insufficient to fully discriminate between tumor subtypes with overlapping appearances.

The Radiomics model achieved 84.45% validation accuracy by the 30th epoch, with steadily decreasing loss and minimal overfitting. This confirms strong learning and good generalization to unseen data.

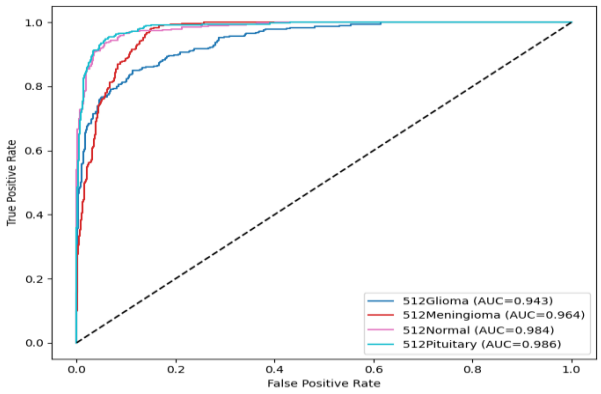

Fig. 8 ROC Graph of Radiomics Model.

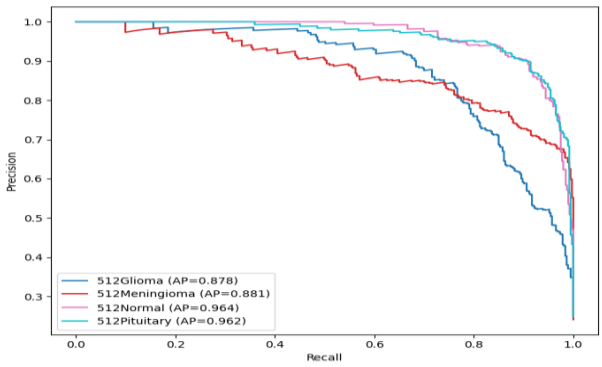

Fig. 9 Precision-Recall Graph of Radiomics Model.

3.2 DL (MobileNetV2) Model Performance

The confusion matrix for this model was obtained by evaluating the predictions of the MobileNetV2-based classifier on the independent test set and comparing them with the ground truth labels, thereby summarizing correct identifications and misclassifications for each class.

The deep-only model demonstrated strong performance across all tumor classes, with high accuracy for glioma (350/373), meningioma (341/363), normal (384/396), and pituitary (368/373). Misclassifications were minimal, with small overlaps primarily between glioma and meningioma, as well as a few errors in normal and pituitary predictions. Overall, the deep learning approach captured discriminative features more effectively.

The deep-only model achieved high accuracy with minimal misclassifications, showing strong discriminative capability across all tumor classes.

3.3 Proposed Fusion Model Performance

The confusion matrix was generated by comparing predicted labels from the logistic regression classifier-trained on the fused feature set obtained by concatenating radiomics and deep embedding (with PCA applied for dimensionality reduction)-against the ground truth test labels, thereby summarizing class-wise correct and incorrect predictions.

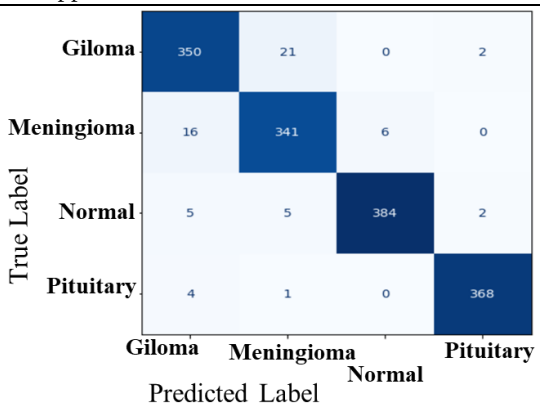

Fig. 10 Confusion matrix of MobileNetV2 Model.

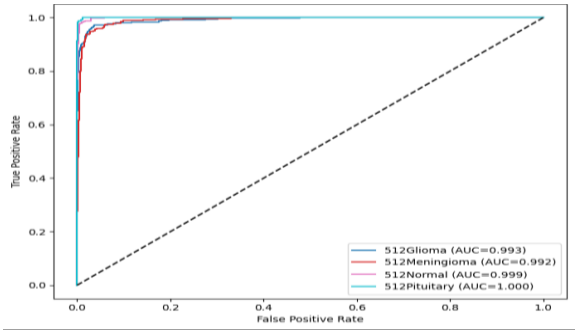

Fig. 11 ROC Graph of MobileNetV2 Model.

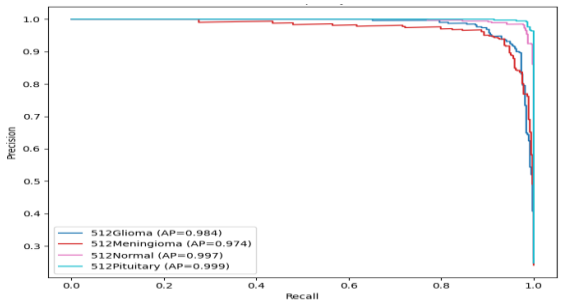

Fig. 12 Precision-Recall Graph of Deep Learning Model.

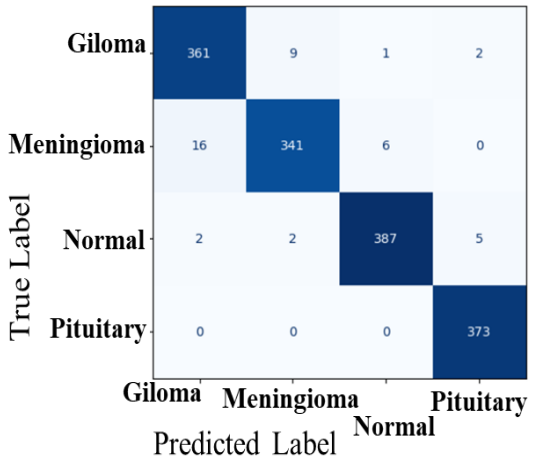

Fig. 13 Confusion matrix of CNN-BiLSTM Model.

The fusion model integrating radiomics with MobileNetV2 achieved excellent classification accuracy, correctly identifying the majority of cases across all categories, including 361/373 gliomas, 341/363 meningiomas, 387/396 normals, and all 373 pituitaries. The ROC curves yielded a macro AUC of 0.998, while the precision–recall analysis reported a macro AP of 0.993, confirming the robustness and near-perfect discrimination ability of the fused approach.

The fusion model delivered near-perfect classification with outstanding ROC and precision-recall performance, confirming its superiority over single-modality approaches.

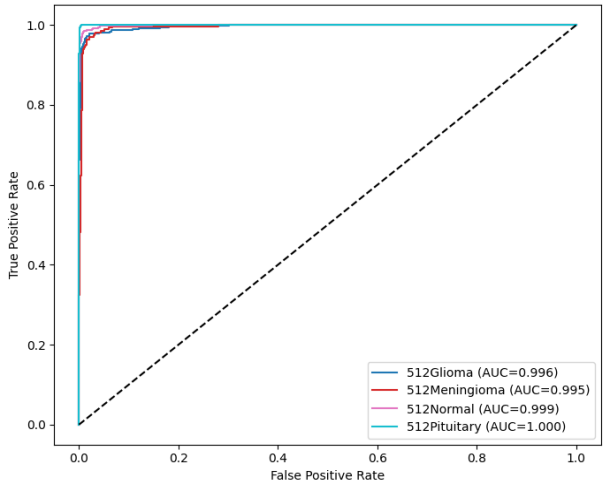

Fig. 14 ROC Graph of Proposed Fusion (Radiomics + Deep Learning) Model.

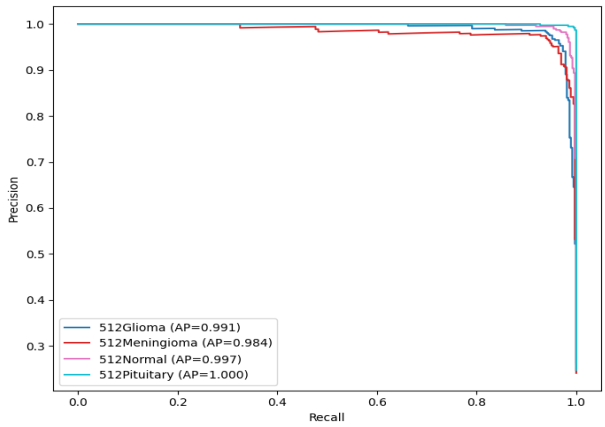

Fig. 15 Precision-Recall Graph of Proposed Fusion (Radiomics + Deep Learning) Model.

3.4 Evaluation Metrics Comparison

The comparative evaluation of all three models shows that radiomics alone achieved moderate performance with a test accuracy of 84.5%, macro F1-score of 0.84, and AUC of 0.97. The deep learning model significantly improved results, reaching 95.9% accuracy, 0.96 macro F1, and 0.996 AUC on the test set, reflecting its stronger feature learning capacity. The fusion approach combining radiomics and deep embedding delivered the best outcomes, attaining 97.1% accuracy, 0.97 macro F1, and 0.998 AUC, demonstrating that integrating handcrafted and deep features yields superior robustness and predictive reliability compared to single-modality models.

Table 1 Evaluation Metrics Comparison of CNN, Bi-LSTM, and CNN-BiLSTM Hybrid Models.

Model	*Accuracy*	*AUC*	*F1 score*
Radiomics (Test)	0.844518	0.969340	0.841795
Radiomics (Val)	0.856667	0.978446	0.856000
DL Model (Test)	0.958804	0.995691	0.958454
DL Model (Val)	0.893333	0.981442	0.891815
Fusion (Test)	0.971429	0.997571	0.971101
Fusion (Val)	0.911111	0.988198	0.910406

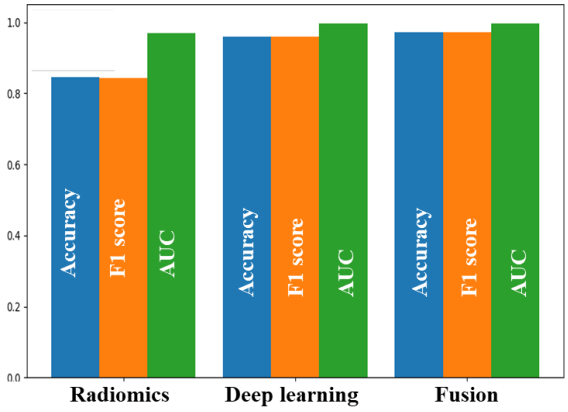

Fig. 16 Comparison of evaluation metrics across Radiomics, Deep Learning and Fusion model.

4. Conclusions

This study demonstrates that integrating radiomics with deep learning features substantially enhances brain tumor classification performance compared to either approach alone, achieving high accuracy, F1-score, and AUC across all tumor subtypes. The results highlight the complementary strengths of handcrafted and learned features in capturing tumor heterogeneity and improving diagnostic reliability. While the fusion model shows strong generalization on the independent test set, future work could extend this framework to multi-center datasets, incorporate additional clinical variables, and explore methods to improve clinical trust. Further research on real-time deployment and cross-modality validation may strengthen its applicability in clinical decision support systems.

5. References

[1] X. Huang, X. Huang, K. Wang, H. Bai, X. Lu, and G. Jin, 2025. 2.5D deep learning radiomics and clinical data for predicting occult lymph node metastasis in lung adenocarcinoma. *BMC Med Imaging, 25(1)*.

[2] W. Wang, H. Yang, W. Hu, G. Shan, W. Ding, C. Yu, B. Wang, X. Wang, Q. Xu, 2010. Clinical Study of the Necessity of Replanning Before the 25th Fraction During the Course of Intensity-Modulated Radiotherapy for Patients with Nasopharyngeal Carcinoma. *International Journal of Radiation Oncology*Biology*Physics, 77(2).*

[3] D. Hui, X. Wang, L. Xie, F. Chen, Y. Guo, Y. Luo, and X. He, 2025. Automatic differentiation of Parkinson's disease motor subtypes based on deep learning and radiomics. *Frontiers in Neurology, 16.*

[4] https://www.researchgate.net/ publication/ 394524847 [Date of access: 18 September 2025]

[5] R. Wang, X. Tian, Z. Wei, Q. Sun and P. Wang, 2025. Fully distributed energy management strategy for DC bus charging stations with three charging modes. *Scientific Reports, 15(1009)*.

[6] Z. Lu, B. Zhu, H. Ling and X. Chen, 2025. Deep learning radiomics based on MRI for differentiating tongue cancer T – staging. *BMC Cancer, 25(1358)*.

[7] Y. Liang, M. He, W. Chen, L. Li, Y. Dong, G. Liang, H. Huangfu, Z. Jiang and S. He, 2025. Deep learning radiomics nomogram predicts lymph node metastasis in laryngeal squamous cell carcinoma *Frontiers in Oncology, 15*.

[8] L. Zhong, L. Shi, X. Liu, Y. Zhao, L. Gu, W. Bai and Y. Zheng, 2025. Development and validation of a prediction model for lymph node metastasis in thyroid cancer: integrating deep learning and radiomics features from intra- and peri-tumoral regions. *Gland Surgery, 14(7)*, pp. 1272–1282.

[9] L. Zhang, L. Hu, L. Tan, Z. Zhang, M. Chen, W. Gan, L. Chen, Y. Zou, S. Wang, Y. Pang, Z. Fan and J. Liu, 2025. The dysbiosis of gut microbiota and dysregulation of metabolites in IgA nephropathy and membranous nephropathy. Frontiers in Medicine, 12.

[10] B. Yu, Y. Luo, S. Tan, L. Cui, X. Liu and Q. Zhou, 2025. Postoperative Anxiety in Childbearing-Age Women with PTMC: Comparison Between Traditional Surgery and Thermal Ablation. *Cancer Management and Research, 17*, pp. 2143-2152.

[11] A. M. Bejar et al., 2025. Multi-center evaluation of radiomics and deep learning to stratify malignancy risk of IPMNs. *Research Square*.