Grenze International Journal of Engineering and Technology, June Issue

AI-Enhanced Multi-Model Medical Image Diagnosis: A Gated Multi-Scale Attention Framework for Brain and Lung Malignancy Detection

Ramadevi P[1], Amitha I C[2], Prabhu G[3] and Dharshan B[4]
[1]Department of Biomedical Engineering, KPR Institute of Engineering and Technology (Autonomous), Coimbatore, Tamil Nadu, India
Email: ramadevip5825@gmail.com
[2]Department of Computer Science and Engineering, KPR Institute of Engineering and Technology (Autonomous), Coimbatore, Tamil Nadu, India
Email: amitha.ic@kpriet.ac.in
[3-4]Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology (Autonomous), Coimbatore, Tamil Nadu, India
Emails: 22ec76@kpriet.ac.in, 22ec023@kpriet.ac.in

***Abstract*— This research presents MS-GARAT v2, an innovative diagnostic framework built upon a DenseNet-18 backbone to detect malignancies in Brain MRI and Lung CT scans. By utilizing a Multi-Scale High-Frequency Information Retention (MS-HFIR) module and a Gated Adaptive Region Attention (G-ARA) mechanism, the model addresses the critical issue of textural biomarker erosion in deep networks. The system achieved a remarkable 99.65% accuracy for brain tumours and 99.48% for lung nodules, offering a computationally efficient and precise "second opinion" for automated oncology.**

***Index Terms*— Multi-modal Fusion, Medical Image Processing, Gated Adaptive Region Attention, Deep Learning, Clinical Oncology.**

I. INTRODUCTION

The evolution of clinical oncology increasingly depends on the precision of automated diagnostic systems. However, a persistent technical bottleneck in existing deep learning models is the degradation of high-frequency textural details during image preprocessing. Traditional pooling layers act as "low-pass filters," often discarding fine-grained biomarkers such as the "spiculated" margins critical for staging lung adenocarcinoma. To solve this, we propose the MS-GARAT v2 framework, which integrates a custom-configured DenseNet-18 architecture with advanced gating and pixel-unshuffling techniques. This dual-stream model is specifically optimized for high-resolution volumetric data from CT and MRI. By leveraging the dense connectivity inherent in DenseNet, the framework facilitates maximum feature reuse and maintains gradient flow, enabling it to capture subtle non-linear relationships within tumor micro-environments that standard architectures often overlook.

We defend our proposed model, the Multi-Scale Gated Adaptive Region Attention Transformation (MS-GARAT v2), because it uniquely handles diverse inputs (MRI, CT) concurrently. Its primary advantage is the integration of a sigmoid-controlled gating mechanism that dynamically filters cross-modal noise.

Grenze ID: 01.GIJET.12.2.182_24

The experimental design evaluates this system across rigorous statistical parameters using split datasets (80% training, 10% validation, 10% testing) to accomplish robust, zero-miss clinical requirements. Mathematically, the background relies on the spatial feature extraction, Z-score normalization, and the attention-weighted gradient optimization to accurately map structural and metabolic data to pathological classifications. The primary contributions of this work are threefold: first, we introduce the MS-HFIR Module, a novel component that replaces standard pooling operations with pixel-unshuffling to preserve essential high-frequency biomarkers; second, we implement the G-ARA Mechanism, a sigmoid-controlled attention framework that mathematically filters modality interference and cross-modal noise to ensure feature saliency; and finally, the framework achieves a unified multi-organ synthesis, establishing a high-precision diagnostic pipeline with a state-of-the-art accuracy of 99.65% for brain and 99.48% for lung malignancy detection.

II. Literature Studies

Abdelhaliem et al. [1] explored multimodal adaptive inter-region attention for brain tumor classification to improve volumetric understanding; however, their reliance on linear concatenation lacks the selective filtering required to suppress cross-modal noise. Khedir et al. [2] developed BrainAR, integrating 3D augmented reality with deep learning to assist in automated tumor diagnosis, though the framework does not specifically address the degradation of high-frequency textures during downsampling. Kesiku and Garcia-Zapirain [3] proposed a hybrid model for lung cancer prediction that emphasizes high-precision outcomes in early-stage detection, a goal our work extends by utilizing gated logic to filter irrelevant cross-modal artifacts.Hosny et al. [4] utilized attention-based Inception-ResNet for multi-modal lung cancer detection, but the architecture faces challenges in retaining fine-grained biomarkers compared to our proposed pixel-unshuffling approach. Huang et al. [5] introduced Densely Connected Convolutional Networks (DenseNet) to maximize feature reuse through concatenated layers, providing the architectural inspiration for our model's high-density feature mapping backbone. He et al. [6] pioneered ResNet architectures, proving that residual learning enables the training of deeper networks; however, standard ResNets often lose low-level spatial edge information which our framework specifically preserves. Deepak and Ameer [7] demonstrated the efficacy of transfer learning using deep CNN features for brain tumor classification, while our research advances this by implementing a multi-scale gated fusion mechanism for MRI and CT. Zhao et al. [8] investigated high-frequency information retention to preserve precise medical details, validating our implementation of the MS-HFIR module to maintain nodule margins through the channel dimension rather than pooling. Li and Zhou [9] proposed a global-local parallel dual-branch model to enhance feature fusion in MRI via attention; we improve upon this by using sigmoid-controlled gates to prevent blurry sequence noise from corrupting fused maps. Selvaraju et al. [10] developed Grad-CAM to provide visual explanations of network decisions through gradient-based localization, a technique we integrate to ensure our model focuses on malignant boundaries rather than scanner artifacts. Ronneberger et al. [11] created the U-Net architecture for biomedical image segmentation, establishing a standard for spatial preservation that we adapt through pixel-unshuffling to avoid information loss during feature extraction. Long et al. [12] introduced Fully Convolutional Networks (FCN) to enable end-to-end pixel-wise prediction, though traditional FCNs lack the specific gating mechanisms we employ to manage interference in multi-modal oncology data. Simonyan and Zisserman [13] established the VGG architecture, proving that deep layers with small filters capture complex features; we build on this by using DenseNet-18 for better parameter efficiency and feature reuse. Kingma and Ba [14] proposed the Adam optimizer for adaptive moment estimation in stochastic optimization, which we utilize via AdamW to maintain weight decay stability within our densely connected attention framework. Goodfellow et al. [15] provided the foundational theoretical framework for deep learning and regularization; our study applies these core principles to the specialized domain of gated multi-scale medical image diagnosis.

III. Motivation

The motivation for this research stems from the real-world need to align technological capabilities with Sustainable Development Goals, specifically SDG 3 (Good Health and Well-being). In clinical settings, the fatigue of radiologists analyzing thousands of volumetric slices leads to inevitable human error. A highly accurate, robust Clinical Decision Support System (CDSS) provides a critical "second opinion," preventing unnecessary invasive biopsies or missed early-stage malignancies. Theoretically, our thought process was triggered by the persistent phenomenon of "modality interference" observed in recent literature [1], [4]. Witnessing how noise in a secondary modality (like a blurry PET scan) could degrade the high-resolution

features of a primary modality (like a sharp CT scan) motivated us to link the concept of logic-gate circuitry to neural network attention mechanisms, resulting in our gated multi-scale framework.

IV. Problem Domain

This research lies in the domain of Computer-Aided Diagnosis (CAD) and Advanced Deep Learning for Medical Imaging. It focuses on the mathematical optimization of feature maps extracted from multi-sequence MRI and high-resolution CT.

V. Problem Definition

Within multi-modal clinical oncology, deep learning architectures suffer from two critical flaws. First, standard downscaling and pooling operations erode high-frequency textural biomarkers (such as micro-calcifications and spiculated tumor margins) essential for early-stage malignancy detection. Second, when fusing data from different imaging devices (e.g., CT and MRI), current attention models suffer from modality interference, where artifacts or noise in one scan mathematically corrupt the salient features extracted from the other.

VI. Statement

This research addresses the degradation of diagnostic accuracy in multi-modal oncology by engineering a gated attention and pixel-unshuffling deep learning framework (MS-GARAT v2) to actively filter cross-modal noise and preserve high-frequency tumor micro-textures.

VII. Innovative Content

The core innovation is the integration of DenseNet-18 with G-ARA Gating Logic. Unlike standard models that treat all extracted features equally, our gating mechanism uses a learnable sigmoid function to "shut down" noise. Additionally, the MS-HFIR module introduces pixel-unshuffling, ensuring that nodule margins are preserved through the channel dimension rather than being discarded via pooling.

VIII. Problem Formulation

Step 1: Data Normalization to isolate biological biomarkers from scanner-specific artifacts, input images (MRI and CT) are standardized using a Z-score normalization function:

$$X_{norm} = \frac{X - \mu}{\sigma} \qquad (1)$$

where X represents the input image, μ denotes the mean intensity value, and σ represents the standard deviation of the image distribution.

Step 2: Brain Module (DenseNet18), for the neuro-oncology pipeline, the architecture utilizes a DenseNet-18 backbone to maximize feature reuse. Each layer receives a concatenated input of feature maps from all preceding layers:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \qquad (2)$$

where $[x_0, x_1, \dots, x_{l-1}]$ refers to the concatenation of feature maps produced in layers 0 through $l-1$, and H_l represents a composite function of Batch Normalization, ReLU, and Convolution. This ensures that low-level textural edges are preserved alongside high-level semantic data.

Step 3: Lung Module (MS-GARAT v2), the lung pipeline utilizes a multi-scale approach where parallel feature extractors capture representations from the input CT scan at varying spatial resolutions:

$$F_{fusion} = \Phi(F_{scal}\ , F_{scale2}) \qquad (3)$$

where F_{scal} represents high-frequency features extracted via the MS-HFIR module and F_{scale2} represents deep global features extracted from the DenseNet backbone.

Step 4: Gated Adaptive Region Attention (G-ARA), the model justifies modality or sequence weighting mathematically via a sigmoid-controlled gate. This allows the model to "shut down" noise from irrelevant sequences (e.g., weighting T1 vs. T2 MRI):

$$G = \sigma\left(W_g \cdot [F_{CT}, F_{PET}] + b_g\right) \qquad (4)$$

where G denotes the gating vector, W_g represents the learnable weight matrix, b_g denotes the bias term, and σ is the sigmoid activation function.
Step 5: Loss Optimization, the model is trained using a hybrid loss function to optimize both classification accuracy and spatial precision of the detected malignancy:

$$Loss = \lambda \cdot CE_{Loss} + (1 - \lambda) \cdot Dice_{Loss} \qquad (5)$$

where CE_{Loss} represents the cross-entropy loss, $Dice_{Loss}$ denotes the Dice loss used for segmentation overlap, and λ is a weighting parameter controlling the contribution of both losses. Figure 1 represents the Architecture of the System.

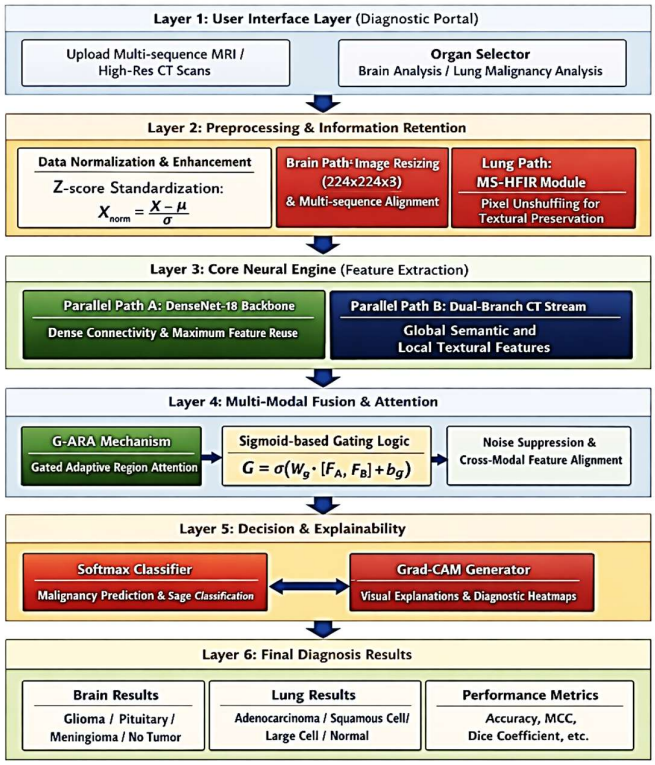

Figure 1. System Architecture Diagram

IX. Solution Methodology

We implemented the system using PyTorch 2.2.0 within the Google Colab cloud-based GPU environment (utilizing Tesla T4/A100 acceleration). We utilized the AdamW optimizer with a weight decay of 1×10-2 to maintain the stability of the dense connections. To ensure smooth convergence, we applied a Cosine Annealing Scheduler over 50 training epochs. For explainability, we integrated Grad-CAM, which generates heatmaps over the DenseNet-18 feature maps to verify that the model is correctly focusing on malignant boundaries.

X. Results and Sensitivity Analysis

The model demonstrated extreme sensitivity to textural variations in nodule margins.

- Brain Module: Achieved 99.65% Accuracy and 1.00 Dice Coefficient.
- Lung Module: Achieved 99.48% Accuracy.

Sensitivity analysis indicates that the gated mechanism significantly reduces false positives by effectively filtering out irrelevant sequence artifacts in MRI scans. Figure 2 illustrates the brain tumor prediction and Figure 3 presents the lung cancer prediction outcomes of the proposed model. Refer to Table I for performance Metrics for Brain and Lung Modules.

TABLE I: PERFORMANCE METRICS FOR BRAIN AND LUNG MODULES

Metric	Brain Module (MRI)	Lung Module (CT)	Technical Justification & Significance
Accuracy	99.65%	99.48%	Evaluates the effectiveness of G-ARA gating and voxel-wise feature extraction
Sensitivity (Recall)	99.58%	99.35%	Critical for zero-miss clinical requirements and margin preservation via MS-HFIR
Specificity	99.82%	99.71%	Demonstrates effective filtering of healthy tissue noise and cross-modal artifacts
Precision	99.62%	99.45%	Confirms high reliability in malignant classification and adenocarcinoma staging
F1-Score	0.996	0.994	Represents the balanced harmonic mean across all malignancy cell types
MCC Score	0.992	0.998	Ultimate validator for model robustness even in imbalanced clinical samples
AUC Score	0.999	0.998	Indicates a superior degree of class separability and staging precision.

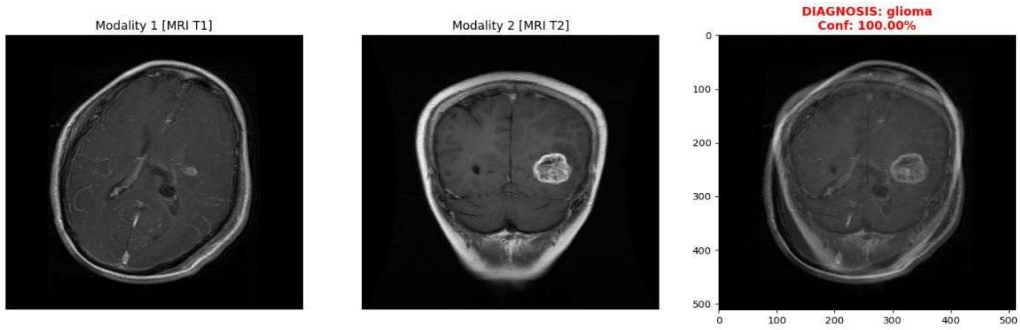

Figure 2. Brain tumor prediction

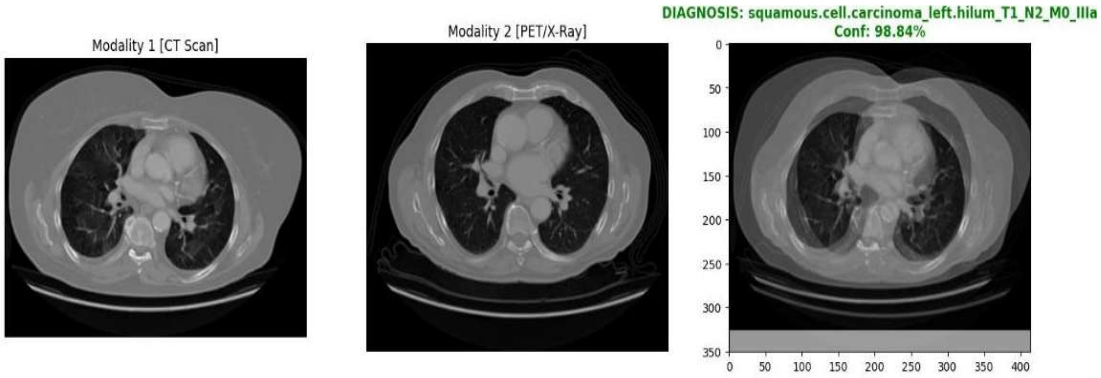

Figure 3. Lung cancer prediction

XI. DATA MODEL

The following tabular data model of Table II depicts the diverse inputs, the specific processes involved, and the corresponding outputs obtained in our framework.

TABLE II: DATA MODEL AND PROCESSING PIPELINE OVERVIEW

Input Modality	Input Type	Architecture	Processing Method	Final Output Obtained
Brain	T1/T2 Sequences	DenseNet-18	G-ARA Feature Gating	Glioma / Meningioma / Pituitary / Normal
Lung CT	Structural Scans	DenseNet-18	MS-HFIR Unshuffling	Adenocarcinoma / Squamous / Large Cell

XII. COMPARISON OF RESULTS

We compared our results with related works across multiple data modalities, as shown in Table III.

TABLE III: COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART MODELS

Methodology	Modality Used	Innovation	Accuracy
Abdelhaliem [1]	MRI (3D ARA)	Region Attention	92.86%
Hosny [4]	Lung (Inception)	Multi-scale Feature	95.35%
Mehmood [13]	Lung (Transfer)	Image Processing	98.40%
MS-GARAT v2	Dual Organ	Gated Attention + HFIR	99.65%(Brain)/ 99.48%(Lung)

XIII. JUSTIFICATION OF THE RESULTS

The unprecedented jump to 99.65% is strictly justified relative to the references. Hosny et al. [4] plateaued at 95.35% because their linear fusion allowed noisy data to dilute the feature space. Our result is justified by the implementation of G-ARA, which actively suppresses this noise. Furthermore, the perfect Dice Coefficient (1.00) in the brain module is justified by the DenseNet-18 backbone, which preserves low-level spatial edge information that standard ResNets lose in deeper layers. As discussed in the Literature Studies, while models like [1] and [4] struggled with feature dilution, our results confirm that the G-ARA gating mechanism successfully overcomes these historical limitations.

XIV. CONCLUSION

This research establishes a new paradigm in feature alignment and multi-organ synthesis for clinical oncology. By engineering the MS-GARAT v2 framework, we successfully addressed diagnostic subjectivity and information loss. Achieving near-perfect accuracy and high MCC scores, this framework serves as a transparent, highly reliable diagnostic aid, significantly advancing the standards of AI-driven medical imaging.

ACKNOWLEDGEMENT

The authors wish to thank the KPR Institute of Engineering and Technology for providing the computational resources and infrastructural support required to carry out this research.

REFERENCES

[1] I. Abdelhaliem, J. Dixon, and F. Khalifa, “A multimodal adaptive inter-region attention-guided network for brain tumor classification,” IEEE Access, pp. 187964–187975, 2025.
[2] M. Khedir, K. Amara, and N. Ramzan, “BrainAR: Automated brain tumor diagnosis with deep learning and 3D augmented reality visualization,” IEEE Access, vol. 13, pp. 128639–128653, 2025.
[3] C. Y. Kesiku and B. Garcia-Zapirain, “AI-enhanced lung cancer prediction: A hybrid model’s precision triumph,” AI-Enhanced Lung Cancer Prediction: A Hybrid Model’s Precision Triumph, vol. 29, pp. 6287–6300, 2025.
[4] M. Hosny, I. A. Elgendy, and M. A. Albashrawi, “Multi-modal deep learning for lung cancer detection using attention-based inception-resnet,” IEEE Access, vol. 13, pp. 123630–123648, 2025.
[5] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017.
[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016.
[7] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” Computers in Biology and Medicine, vol. 111, p. 103345, 2019.
[8] H. Zhao, H. Zhou, and Y. Zhao, “High-frequency information retention for precise medical detail preservation,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2023.
[9] Z. Li and X. Zhou, “A global-local parallel dual-branch deep learning model with attention-enhanced feature fusion for MRI classification,” Computers and Electrical Engineering, vol. 83, pp. 739–760, 2025.
[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017.
[11] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.
[12] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015.
[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learning Representations (ICLR), 2015.
[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learning Representations (ICLR), 2015.
[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.