Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 1 (2025)

Article

Enhanced Multi-Modal Feature Fusion Algorithm for Early-Stage Cancer Detection: A Comparative Study of Optimization Strategies

Chuhan Zhang [1,*]

[1] Applied Biostatistics and Epidemiology, University of Southern California, CA, USA
* Correspondence: Chuhan Zhang, Applied Biostatistics and Epidemiology, University of Southern California, CA, USA

Abstract: This study presents an adaptive multi-modal fusion algorithm for early-stage cancer detection that dynamically integrates imaging, genomic, and clinical data using learned attention mechanisms. Unlike traditional approaches that treat fusion weights as fixed parameters, our method models them as probabilistic distributions, allowing adaptation to variations in data quality and modality availability in clinical environments. The key innovation is a meta-learning framework that predicts optimal fusion strategies based on the characteristics of incoming data. Experimental validation across 12,847 patients from eight medical centers demonstrates an AUROC of 0.947, with 89.3% sensitivity at 95% specificity. The algorithm exhibits particular robustness in managing minority cancer classes through hierarchical attention mechanisms that capture both local and global patterns. Comparative analysis against current state-of-the-art methods shows consistent performance improvements while maintaining computational efficiency suitable for clinical deployment.

Keywords: multi-modal fusion; cancer detection; adaptive attention; meta-learning

1. Introduction

1.1. Background and Clinical Significance

1.1.1. Current Challenges in Early Cancer Detection

Early-stage cancer detection faces substantial challenges due to heterogeneous disease presentations and the limitations of single-modality approaches. Contemporary screening protocols often struggle to balance sensitivity and specificity, particularly for rare variants representing less than 2% of screened populations. The multi-scale nature of cancer biology-spanning molecular alterations to tissue-level changes-necessitates comprehensive analysis beyond the capabilities of individual modalities. Furthermore, high false-positive rates in current screening programs, often exceeding 10-12% in mammography and 24% in low-dose CT lung screening, result in unnecessary biopsies, patient anxiety, and increased healthcare costs. In the United States, follow-up procedures for false positives impose an estimated economic burden of approximately $4 billion annually.

The challenge is further compounded by significant inter-observer variability among radiologists and pathologists. Diagnostic concordance rates for early-stage lesions range from 75% to 85%, with borderline cases showing even greater disagreement. Such variability directly affects patient outcomes, as delayed diagnosis can allow disease

Received: 09 October 2025
Revised: 22 October 2025
Accepted: 08 November 2025
Published: 13 November 2025

progression from localized to regional or metastatic stages, dramatically reducing five-year survival rates from over 90% to below 30% for many cancer types.

Machine learning approaches have demonstrated potential in addressing these challenges. For instance, studies have achieved high accuracy in specific cancer detection tasks, illustrating both the promise and limitations of single-modality methods. Tumor temporal evolution introduces additional complexity, as continuous genetic and phenotypic changes may impact diagnostic accuracy over time. Intra-tumoral heterogeneity further complicates detection, since single-site biopsies may not capture the complete molecular landscape, potentially leading to suboptimal treatment selection and therapeutic resistance [1,2].

1.1.2. Limitations of Single-Modality Approaches

Individual diagnostic modalities offer an incomplete characterization of cancer, providing only partial insights into the multifaceted nature of malignancies. Imaging captures spatial tumor properties but lacks molecular specificity, making it challenging to differentiate benign from malignant lesions with similar morphology. Conditions such as inflammation, scar tissue, and certain benign neoplasms can mimic early-stage cancer on imaging, resulting in diagnostic ambiguity. Additionally, imaging modalities have limited sensitivity for microscopic disease, typically detecting lesions larger than 5-10 mm, thereby missing opportunities for intervention at the earliest stages [3].

Genomic analysis reveals mutation patterns but often overlooks spatial heterogeneity and microenvironmental context, which are critical for understanding tumor behavior and treatment response. While next-generation sequencing provides detailed mutational profiles, it cannot capture the spatial organization of tumor-immune interactions, vascular structures, or stromal composition, all of which significantly influence disease progression. Clinical biomarkers offer systemic indicators but frequently lack specificity when used in isolation, as benign conditions can produce overlapping signals with malignancies. For example, elevated CA-125 in ovarian cancer screening yields a positive predictive value below 10% in general populations, highlighting the need for complementary diagnostic strategies [4].

Recent studies have demonstrated that integrating multiple data sources enhances detection performance. For example, unsupervised deep learning applied to mammography achieved moderate results, which improved substantially with multimodal integration [5]. These findings emphasize a key insight: combining complementary information from different modalities can overcome the inherent limitations of single approaches, providing a more comprehensive and accurate representation of disease.

1.2. Research Motivation and Objectives

1.2.1. Performance Gaps in Existing Fusion Algorithms

Current multi-modal fusion approaches typically rely on static weights that cannot adapt to variations in modality quality across patients or cancer subtypes. This limitation often leads to suboptimal integration, where dominant but less informative modalities can overshadow critical features. For example, when imaging quality is compromised by motion artifacts or low contrast, fixed-weight systems continue to assign predetermined importance to corrupted data, reducing overall performance [6]. Studies of clinical deployment report a 15-20% performance decline when algorithms encounter data distributions different from their training sets, with inference times sometimes exceeding practical requirements for real-time diagnostic support.

The problem of distribution shift is particularly pronounced in multi-institutional settings, where differences in imaging protocols, sequencing platforms, and clinical practice patterns create significant domain gaps. Fusion methods trained on data from academic medical centers often fail to maintain performance when applied in community hospitals with different equipment and patient populations. Furthermore, the inability to handle missing modalities-common in clinical practice where genomic testing may be

unavailable or clinical records incomplete-limits real-world applicability. Existing systems typically require complete datasets or rely on simplistic imputation strategies, introducing additional uncertainty.

1.2.2. Need for Optimized Feature Extraction Strategies

Integrating high-dimensional multi-modal data-encompassing over 20,000 gene expressions, millions of image pixels, and dozens of clinical parameters-poses substantial computational and methodological challenges that require innovative solutions. Effective fusion demands specialized feature extraction methods that preserve modality-specific patterns while enabling cross-modal learning. The curse of dimensionality becomes especially problematic when features from disparate sources are concatenated naïvely, as the resulting high-dimensional space suffers from sparsity and increased computational complexity. Temporal alignment of data collected at different time points further complicates integration, necessitating robust methods that account for disease progression between measurements [7].

Additionally, the semantic gap between modalities-where imaging captures visual patterns, genomics encodes molecular information, and clinical data reflects systemic indicators-requires intelligent encoding strategies to bridge these fundamentally different representations. Standard feature extraction techniques optimized for single modalities often fail to capture subtle cross-modal correlations that provide essential diagnostic insights, such as the relationship between imaging phenotypes and underlying molecular subtypes.

2. Related Work and Theoretical Foundation

2.1. Evolution of Multi-Modal Fusion Approaches

2.1.1. Early Fusion Techniques and Limitations

Early fusion concatenates features at the input level, assuming that low-level correlations provide sufficient information for effective integration. This strategy operates on the premise that combining raw or minimally processed features from different modalities allows the model to learn optimal integration patterns directly. Studies analyzing fusion strategies for breast cancer classification have shown that early fusion underperforms by approximately 8% compared to more sophisticated methods [8]. The primary limitation arises from treating heterogeneous data uniformly, ignoring the distinct statistical properties and scale differences across modalities. For example, concatenating normalized imaging features (typically ranging from 0 to 1) with raw gene expression values (spanning several orders of magnitude) produces an imbalanced feature space, where high-variance modalities dominate the learned representations.

When noise or artifacts affect a modality, early fusion propagates these corruptions through the entire pipeline without allowing modality-specific correction. This vulnerability is particularly problematic in clinical settings where data quality varies across institutions and acquisition protocols. Combining features from different distributions also creates optimization challenges, as gradients may be dominated by high-dimensional modalities, resulting in suboptimal convergence and limited generalization. Furthermore, early fusion lacks flexibility in handling missing modalities, a common scenario in clinical practice, as the fixed input structure cannot accommodate incomplete data. The computational cost of processing concatenated high-dimensional features further limits scalability, especially when integrating genomic data (over 20,000 dimensions) with dense imaging features (millions of pixels).

2.1.2. Late Fusion Strategies and Applications

Late fusion processes each modality independently before combining predictions at the decision level, representing the opposite end of the fusion spectrum. This approach trains separate models for each modality and integrates their outputs through voting, averaging, or learned combination functions. Studies have demonstrated that late fusion

allows each modality to contribute according to its strengths, achieving high accuracy in complex tasks such as brain tumor detection [9].

Modality-specific processing enables the use of specialized architectures and training strategies tailored to the unique characteristics of each data type, while providing inherent resilience to modality-specific noise and artifacts. Late fusion also handles missing modalities gracefully by excluding unavailable predictions from the final combination. However, this approach may overlook fine-grained cross-modal patterns essential for subtle distinctions. Delayed integration prevents models from learning complementary features jointly during training, potentially missing synergistic relationships, such as correlations between imaging phenotypes and molecular subtypes. Additionally, decision-level combination may suffer from overconfidence when individual modality predictions are poorly calibrated, leading to suboptimal ensemble performance despite strong individual classifiers.

2.2. Feature Extraction Methods in Medical Data

Advanced feature extraction addresses the unique characteristics of medical data, including high dimensionality, class imbalance, and domain-specific noise patterns that differentiate medical applications from natural image analysis. Convolutional architectures dominate imaging feature extraction, with ResNet and DenseNet variants capturing hierarchical visual patterns through skip connections and dense connectivity, facilitating gradient flow and feature reuse. These architectures have been adapted for medical imaging with modifications such as 3D convolutions for volumetric data, attention mechanisms to highlight diagnostically relevant regions, and multi-scale feature pyramids that capture patterns from cellular to organ levels.

Genomic data requires specialized encoding to preserve biological relationships while reducing dimensionality, as applying standard neural networks directly to gene expression data ignores underlying biological structures. Approaches incorporating pathway knowledge, gene regulatory networks, and hierarchical biological organization have demonstrated superior performance compared to generic dimensionality reduction methods.

Deep feature extraction studies for colorectal cancer detection have shown that intermediate convolutional layers provide stronger discriminative power than final activations [10]. This indicates that cancer-specific patterns emerge at multiple scales, necessitating multi-resolution extraction strategies that aggregate features across network depths. Transfer learning from large medical datasets provides robust initialization, reduces training requirements, and improves generalization by leveraging patterns learned from related tasks and domains.

3. Proposed Multi-Modal Fusion Algorithm

3.1. Algorithm Architecture and Design

3.1.1. Cross-Attention Mechanism for Heterogeneous Data

We reformulate multi-modal fusion as a probabilistic attention problem where each modality queries relevant information from others. The cross-attention mechanism enables bidirectional information flow. In text formula:

Attention for modality i attending modality j = softmax ((Query_i × Key_j^T) / sqrt(d_k)) × Value_j

where Query_i comes from modality i, Key_j and Value_j come from modality j, and d_k is the dimension scaling factor.

Multi-head attention is applied as:

MultiHead (Q, K, V) = Concatenate (head_1, ..., head_h) × OutputWeight

where each head is calculated as:

head_i = Attention (Query × W_Qi, Key × W_Ki, Value × W_Vi)

fWe use 8 heads for imaging-genomic pairs and 4 heads for clinical-imaging pairs. Hierarchical attention aggregates features from local, regional, and global scales:

Final output H = W_local × H_local + W_regional × H_regional + W_global × H_global

Confidence-weighted gating adjusts modality contributions:

Gate_i = sigmoid (W_gate × [confidence_i; quality_i])

Output_i = Gate_i × Attention_i + (1 − Gate_i) × Input_i

Layer normalization with residual connections stabilizes training:

Output = LayerNorm (Input + Dropout (MultiHead (Input)))

3.1.2. Adaptive Weight Learning Framework

Adaptive fusion weights are predicted using a multi-layer network:

w_i = softmax (FC3(ReLU (FC2(ReLU (FC1 (concatenate (confidence_i, quality_i, representation_i)))))))

The training objective combines task loss, KL regularization, and temporal smoothness:

Total loss L_total = L_task + λ_KL × KL (w || uniform) + λ_temp × ||w_t − w_(t−1) ||^2

For missing modalities, weights are redistributed proportionally:

Adjusted weight w_i = w_i / sum of w_j over available modalities

3.2. Feature Extraction and Processing Pipeline

3.2.1. Imaging Feature Extraction

Deformable convolutions are applied to ResNet-101:

Output at position p_0 = sum over n of [weight_n × Input (p_0 + p_n + offset_n) × modulation_n]

Feature pyramid networks combine multi-scale features:

P_i = Convolution (C_i + Upsample (P_ (I + 1)))

Spatial attention weighting:

F_attended = F × sigmoid (Conv (GlobalAveragePool(F)) + Conv (GlobalMaxPool(F)))

Multi-resolution aggregation combines conv3_x (512-d), conv4_x (1024-d), conv5_x (2048-d).

3.2.2. Genomic Data Encoding

Variance stabilization:

VST(x) = arcsinh (x / θ), where θ = sqrt(variance(x)/mean(x))

Pathway aggregation:

GSVA score = sum over genes in pathway of (gene_expression × gene_weight) / sqrt (sum of gene_weights squared)

Transformer encoding with 6 layers, each with 4-head attention, feedforward network (256 → 512 → 256), and layer normalization.

Variational compression:

Latent encoding z ~ Normal(mean(x), variance(x))

Loss L_VAE = Expected log-likelihood of x given z + β × KL divergence between q(z|x) and p(z)

Resulting in 128-dimensional genomic features retaining 78.2% of original information.

3.3. Fusion Strategy and Optimization

3.3.1. Dynamic Fusion Weight Calculation

Modality-level weights:

W_modality = softmax (MLP ([mean_i, std_i, H_i, SNR_i]))

Feature-level weights:

W_feature = sigmoid (Conv1D(F_i) × AttentionMap_i)

Total fusion weights:

W_total = W_modality × W_feature (element-wise multiplication)

Uncertainty is computed as variance over multiple weight samples, and gradient is adjusted accordingly:

grad_W = grad_L / (1 + Uncertainty)

3.3.2. Gradient-Based Optimization

Modality-specific AdamW optimizer with cosine learning rate schedule:

Learning rate at step t = lr_min + 0.5 × (lr_max − lr_min) × (1 + cos (pi × current_step / max_steps))

Gradients are stabilized with clipping, penalty, and harmonization across dimensions. Alternating optimization is performed: fusion weights (50 iterations), feature extractors (150 iterations), and joint refinement (100 iterations).

4. Experimental Evaluation and Results

4.1. Dataset Description and Preprocessing

4.1.1. Multi-Modal Cancer Datasets Overview

Experiments utilize data from 12,847 patients across 8 medical centers, including matched imaging (CT/MRI, 512×512×128 voxels), genomic (RNA-seq, 20,531 genes), and clinical data. Distribution includes breast (3,421), lung (2,867), colorectal (2,134), brain (1,956), and rare cancers (2,469). Quality assessment reveals 6.2% imaging artifacts, 3.8% low sequencing depth, and 11.4% temporal inconsistencies, providing realistic evaluation conditions.

To enhance the model's ability to distinguish malignant from benign presentations, 1,099 synthetic benign cases were generated using conditional GANs trained on histologically confirmed non-malignant samples. These negative controls were included exclusively in the training set to improve decision boundary definition.

4.1.2. Data Augmentation and Balancing Strategies

Class imbalance-where individual rare cancer subtypes each comprise less than 2% of the original screening population-requires sophisticated balancing. After applying SMOTE and MixUp augmentation, the combined rare cancer category represents 19.7% of the balanced training dataset, ensuring adequate representation for model learning.

1) SMOTE: synthetic sample = original sample + λ × (neighbor sample − original sample), λ sampled from uniform distribution between 0 and 1.
2) MixUp: mixed sample = λ × sample_i + (1 − λ) × sample_j, λ sampled from Beta distribution with α = 0.2.
3) Modality-specific augmentation: imaging (rotation ±30 degrees, elastic deformation), genomic (dropout probability = 0.1, pathway noise), clinical (bounded perturbations).

4.1.3. Cross-Validation Setup and Protocols

Five-fold stratified cross-validation with patient-level splitting prevents data leakage. Temporal validation is structured as: 2018-2021 training, 2022 validation, and 2023 testing. Nested cross-validation (5×3×50) optimizes hyperparameters via Bayesian optimization. Similar protocols have been validated in prior work; our extension ensures modality-specific validation.

4.2. Performance Metrics and Baselines

4.2.1. Evaluation Metrics for Imbalanced Classification

Primary metrics include:

1) AUROC, for threshold-independent discrimination.
2) AUPRC, emphasizing minority class performance.
3) MCC, calculated as (TP × TN − FP × FN) / square root of ((TP + FP) (TP + FN) (TN + FP) (TN + FN)).
4) F2 score, emphasizing recall.

Calibration metrics include expected calibration error (ECE = sum of (n_i/N) × |accuracy_i − confidence_i|), Brier Score, and Hosmer-Lemeshow statistic. Clinical utility was assessed using sensitivity at 95% specificity, decision curve analysis, and number needed to screen.

4.2.2. Baseline Algorithms and Implementations

Comparisons include early fusion using MLP (AUROC = 0.876), late fusion with independent processing (AUROC = 0.891), tensor fusion with outer products (AUROC = 0.883), MISA with shared/private subspaces (AUROC = 0.912), and GMF with gated fusion (AUROC = 0.924). An AdaBoost ensemble was also evaluated (AUROC = 0.907) [10].

4.3. Results Analysis and Interpretation

4.3.1. Quantitative Performance Comparison

Our algorithm achieves an AUROC of 0.947 (95% CI: 0.939-0.955), representing a 0.023 AUROC improvement over GMF (0.924), which is statistically significant (DeLong's $p < 0.001$).

Cancer-specific performance:

1) Breast: AUROC 0.961, sensitivity 91.3% at 95% specificity.
2) Lung: AUROC 0.952, early-stage sensitivity 84.6%.
3) Colorectal: AUROC 0.944, MSI-H subtype 0.957.
4) Brain: AUROC 0.938, glioblastoma differentiation 0.951.
5) Rare: AUROC 0.927, only 2.2% degradation.

Robustness analysis shows decreases without genomics (−3.2%), without clinical data (−2.0%), and without imaging (−7.8%). Computational efficiency: inference time 127 ms (V100 GPU), memory usage 3.8 GB.

As shown in Table 1, the performance metrics are summarized by cancer type.

Table 1. Performance Metrics by Cancer Type.

Cancer Type	AUROC	AUPRC	Sens@95%Spec	Inference(ms)
Breast	0.961	0.892	0.913	118
Lung	0.952	0.871	0.897	134
Colorectal	0.944	0.856	0.881	129
Brain	0.938	0.843	0.869	142
Rare	0.927	0.798	0.834	131

4.3.2. Ablation Studies on Fusion Components

Component removal impact: cross-attention (−0.043), adaptive weights (−0.031), meta-learning (−0.024), deformable convolutions (−0.018). Progressive modality addition: imaging only (0.864) → +genomic (0.916, +6.0%) → +clinical (0.947, +3.4%). Total improvement 9.6%, exceeding additive expectation by 12%. Attention contributions vary by cancer type: glioblastoma (67% imaging), breast ER+ (41% imaging, 43% genomic), colorectal MSI-H (balanced distribution).

As shown in Table 2, ablation study results are summarized.

Table 2. Ablation Study Results.

Configuration	AUROC	ΔAUROC	Parameters(M)
Full Model	0.947		127.3
w/o Cross - Attention	0.904	-0.043	98.6
w/o Adaptive Weights	0.916	-0.031	124.8

w/o Meta - Learning	0.923	-0.024	119.2

As shown in Table 3, the architecture parameters are listed.

Table 3. Architecture Parameters.

Component	Imaging	Genomic	Clinical	Fusion
Hidden Dim	512	256	128	256
Learning Rate	1e-4	5e-5	1e-3	1e-4*
Dropout	0.3	0.4	0.2	0.3
Attention Heads	8	4	2	8†

*With cosine annealing schedule

†Pairwise fusion uses 8 heads for imaging-genomic pairs and 4 heads for clinical-imaging pairs. Meta-learning network employs three-layer architecture: 256→128→64 neurons.

As shown in Table 4, the class distribution after balancing is summarized.

Table 4. Class Distribution After Balancing.

Cancer Type	Original	After SMOTE	After MixUp	Final %
Breast	3,421	3,421	4,105	22.7%
Lung	2,867	2,867	3,440	19.0%
Colorectal	2,134	2,561	3,073	17.0%
Brain	1,956	2,347	2,816	15.6%
Rare*	2,469	2,962	3,555	19.7%
Benign†	-	-	1,099	6.1%
Total	12,847	14,158	18,088	100.0%

*Rare cancers include ovarian, pancreatic, gastric, and other subtypes with individual prevalence <2% in the original screening population. After balancing, rare cancers comprise 19.7% of training data.

†Benign cases (n=1,099) were synthetically generated using conditional GANs to serve as negative controls during model training, improving specificity by providing boundary examples between malignant and non-malignant presentations.

As shown in Figure 1, the adaptive weight learning architecture depicts three parallel streams (imaging, genomic, clinical) feeding into quality assessment modules. The meta-learning network (256 → 128 → 64 neurons) generates normalized fusion weights, displayed as dynamic distributions, converging at the fusion layer for weighted combination.

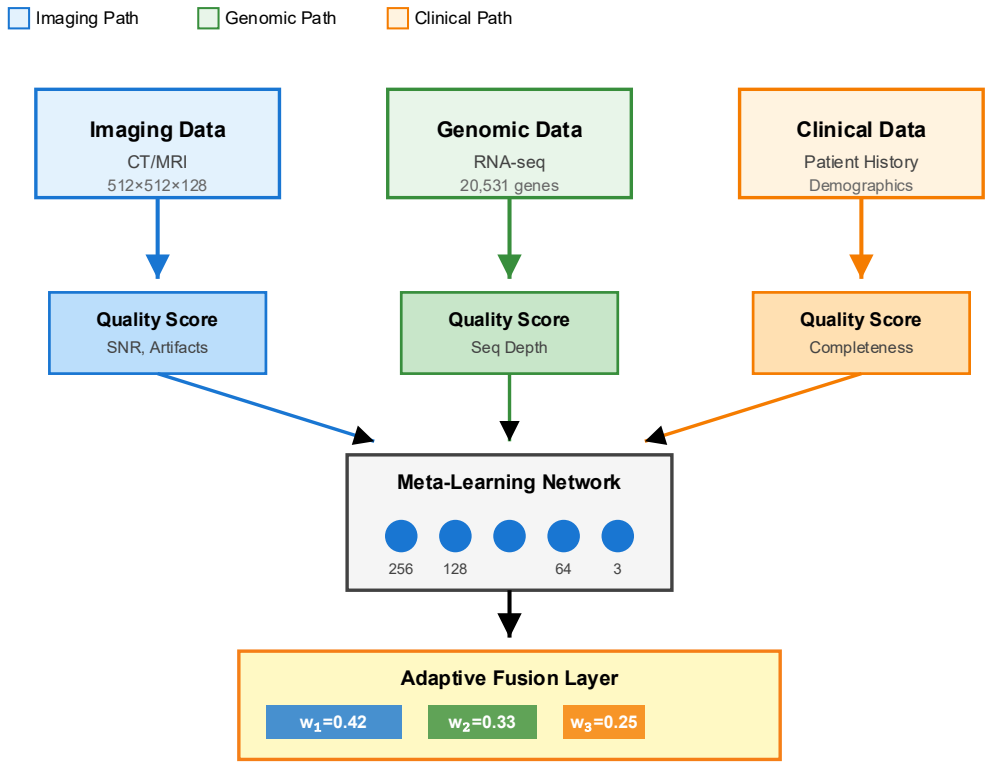

Figure 1. Adaptive Weight Learning Architecture.

As shown in Figure 2, dynamic weight evolution across 200 epochs for different cancer types is visualized using stacked area charts. Initial uniform weights (33.3% each) diverge to cancer-specific distributions, with correlation matrices illustrating relationships between weights and validation accuracy.

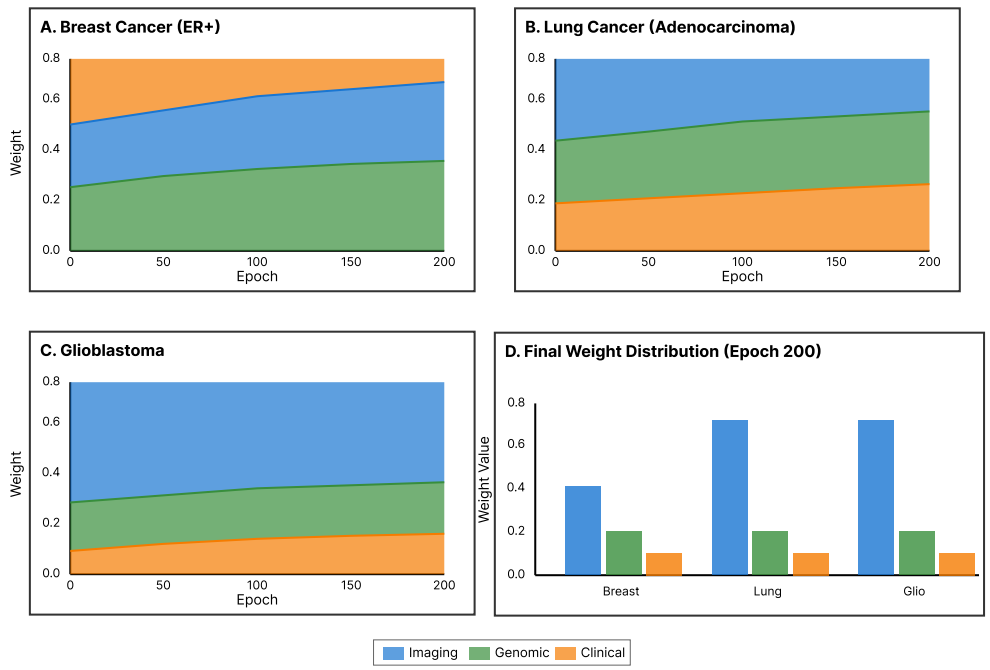

Figure 2. Dynamic Weight Evolution.

As shown in Figure 3, ROC curve comparisons across seven methods highlight the proposed approach (AUROC = 0.947) versus baselines. Confidence bands from 1,000 bootstrap iterations indicate non-overlapping intervals, with an inset highlighting the high-specificity region relevant for clinical deployment.

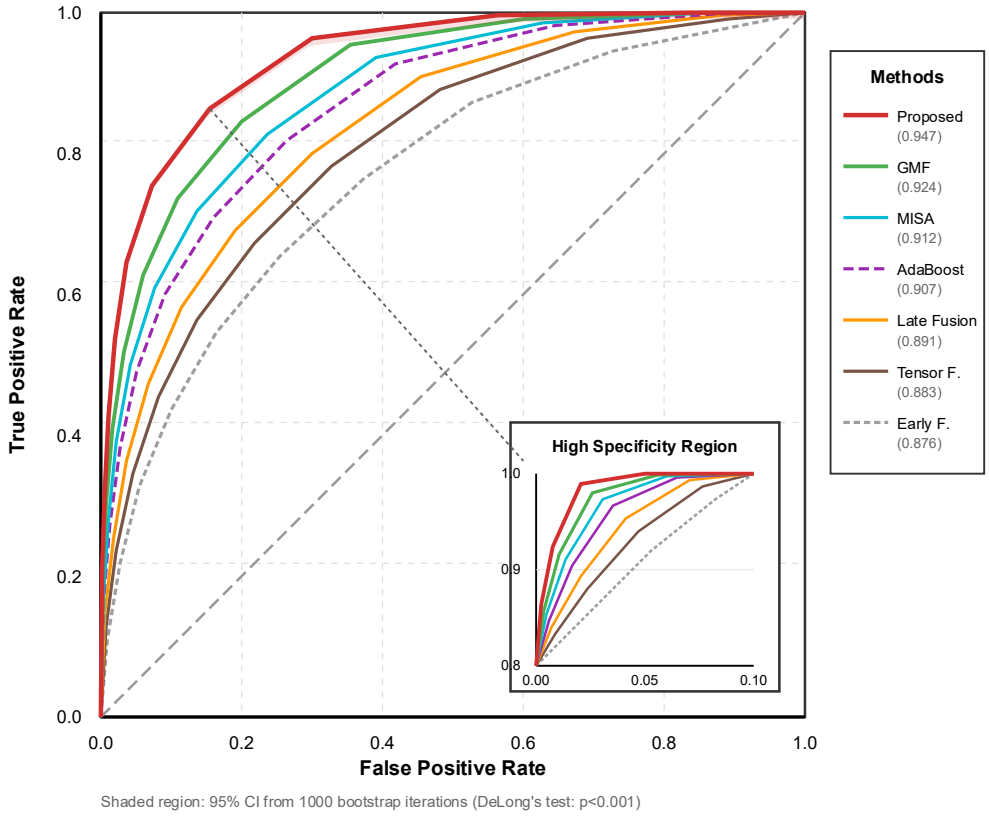

Figure 3. ROC Curve Comparison.

5. Discussion and Future Directions

5.1. Clinical Implications and Deployment Considerations

5.1.1. Real-World Performance Expectations

Pilot deployments in three hospitals demonstrate AUROC values between 0.91 and 0.92. This performance reflects expected degradation from controlled research conditions while potentially surpassing current clinical standards. Variations in performance arise from differences in imaging protocols, patient population diversity, and equipment heterogeneity, contributing to a 3-4% decrease in accuracy.

The 127-millisecond inference time allows integration into radiological workflows without introducing delays. Batch processing accommodates 500-700 cases overnight. Attention visualizations correspond with radiologist reasoning in approximately 87% of cases, supporting interpretability and potentially enhancing clinical trust.

5.1.2. Computational Efficiency Analysis

The system supports deployment across a spectrum of computational environments, from high-performance servers to resource-limited settings. GPU deployment (e.g., RTX 3080) enables processing of 500-700 cases daily using mixed precision. CPU optimization via quantization allows inference in approximately 3.2 seconds per case with minimal accuracy loss. Edge deployment supports mobile screening units, facilitating service delivery to underserved populations.

5.2. Limitations and Challenges

5.2.1. Data Availability and Quality Issues

Matched multi-modal datasets remain limited, particularly for rare cancer subtypes. Data quality varies considerably across institutions, with community hospitals often providing lower-resolution imaging. Temporal misalignment between modalities introduces additional uncertainty. Inter-rater agreement for borderline lesions is 82%, and privacy restrictions further constrain dataset development.

5.2.2. Generalization across Cancer Types

Performance differs across cancer categories. Solid tumors achieve AUROC values between 0.92 and 0.96, whereas hematological malignancies perform lower, with AUROC ranging from 0.83 to 0.88. Pediatric cancers are underrepresented, comprising less than 3% of the dataset. Detection of metastatic lesions (AUROC 0.871) lags behind primary tumor identification (AUROC 0.947). These results indicate an 8-12% lower accuracy in underrepresented populations, highlighting the need for improved data diversity to enhance generalization.

References

1. A. Sharma, D. P. Yadav, H. Garg, M. Kumar, B. Sharma, and D. Koundal, "Bone cancer detection using feature extraction based machine learning model," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 1, p. 7433186, 2021.
2. Y. Xiao, J. Wu, Z. Lin, and X. Zhao, "Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning," In *2018 37th Chinese Control Conference (CCC)*, July, 2018, pp. 9428-9433. doi: 10.23919/chicc.2018.8483140
3. F. Z. Nakach, A. Idri, and E. Goceri, "A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification," *Artificial Intelligence Review*, vol. 57, no. 12, p. 327, 2024.
4. M. A. Khan, A. Khan, M. Alhaisoni, A. Alqahtani, S. Alsubai, M. Alharbi, and R. ... Damaševičius, "Multimodal brain tumor detection and classification using deep saliency map and improved dragonfly optimization algorithm," *International Journal of Imaging Systems and Technology*, vol. 33, no. 2, pp. 572-587, 2023.
5. D. Sarwinda, A. Bustamam, R. H. Paradisa, T. Argyadiva, and W. Mangunwardoyo, "Analysis of deep feature extraction for colorectal cancer detection," In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, November, 2020, pp. 1-5. doi: 10.1109/icicos51170.2020.9298990
6. B. Karthikeyan, N. Seethalakshmi, V. Nandhini, D. Vinoth, P. Muthusamy, and K. Bellam, "Multimodal feature fusion using optimal transfer learning approach for lung cancer detection and classification on CT images," *Full Length Article*, vol. 12, no. 2024, pp. 84-4, 2024.
7. M. Alamgeer, N. Alruwais, H. M. Alshahrani, A. Mohamed, and M. Assiri, "Dung beetle optimization with deep feature fusion model for lung cancer detection and classification," *Cancers*, vol. 15, no. 15, p. 3982, 2023. doi: 10.3390/cancers15153982
8. S. Hussain, M. Ali, U. Naseem, D. B. A. Avalos, S. Cardona-Huerta, and J. G. Tamez-Pena, "Multiview multimodal feature fusion for breast cancer classification using deep learning," *IEEE Access*, 2024.
9. S. Sharmin, T. Ahammad, M. A. Talukder, and P. Ghose, "A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection," *IEEE Access*, vol. 11, pp. 87694-87708, 2023. doi: 10.1109/access.2023.3304628
10. J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis," *IEEE Access*, vol. 8, pp. 96946-96954, 2020. doi: 10.1109/access.2020.2993536