HVM
Human & Veterinary Medicine

A unified fractal-entropy-manifold framework for multi-modal medical image fusion

Ștefan Țălu

The Directorate of Research, Development and Innovation Management (DMCDI), The Technical University of Cluj-Napoca, 400020 Cluj-Napoca, Romania. E-mail: stefan.talu@auto.utcluj.ro, stefan_ta@yahoo.com

Abstract. We propose a unified fractal-entropy-manifold framework for multi-modal medical image fusion, designed to preserve multi-scale structural complexity, enable entropy-guided modality integration, and maintain geometric consistency within a shared latent representation space. The method combines three complementary components within a single variational formulation: (i) a fractal gradient texture descriptor (FGTD) to capture scale-dependent anatomical structures and self-similar texture patterns, (ii) an entropy-guided adaptive fusion mechanism that assigns spatially varying modality weights based on local information content, and (iii) graph-based manifold regularization that enforces neighborhood structure via Laplacian constraints in the latent space. Unlike conventional approaches that separate feature extraction, fusion, and regularization, the proposed framework formulates multi-modal integration as a unified optimization problem. The latent representation is computed using an alternating optimization scheme that iteratively updates modality weights and the fused representation under fractal consistency, entropy regularization, and manifold smoothness constraints. The framework is evaluated on co-registered magnetic resonance imaging and computed tomography (MRI-CT) data, under controlled intra-subject conditions. Quantitative evaluation using structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), and mutual information (MI) shows improved performance compared to classical fusion methods (e.g., averaging and wavelet-based techniques) and deep learning baselines (CNN- and Transformer-based models). Qualitative results further demonstrate enhanced edge preservation, improved structural fidelity, and reduced cross-modal artifacts. These findings indicate that integrating fractal-based texture modeling, entropy-guided adaptation, and manifold-constrained representation learning yields a robust and interpretable solution for multi-modal medical image fusion, with applications in neuroimaging and computational radiology.

Key Words: fractal gradient texture descriptor, multi-modal medical image fusion, entropy-guided fusion, manifold learning, graph Laplacian regularization, MRI-CT fusion, fractal-based representation, manifold-constrained representation learning.

Introduction. Medical imaging is a fundamental component of modern clinical workflows, enabling non-invasive visualization, diagnosis, and quantitative assessment of anatomical and pathological structures. In recent years, a paradigm shift has occurred from qualitative interpretation toward quantitative, data-driven analysis, largely driven by radiomics (Aerts et al 2014; Robert & Paule 2016).

Radiomics transforms medical images into high-dimensional feature representations, enabling the extraction of quantitative biomarkers that characterize tissue heterogeneity, structural organization, and disease progression beyond visual assessment (Lambin et al 2012). However, conventional radiomics pipelines remain heavily dependent on handcrafted feature engineering, which is often insufficient to capture complex spatial dependencies, multi-scale variability, and nonlinear structures inherent in medical imaging data (Lambin et al 2012; Robert & Paule 2016).

To address these limitations, deep learning has emerged as the dominant paradigm in medical image analysis, enabling hierarchical feature learning directly from raw data and achieving state-of-the-art performance in segmentation, classification, and detection tasks (Litjens et al 2017; Lepcha et al 2025). Extensive studies further confirm the effectiveness of deep neural architectures across a wide range of imaging applications (Zhou et al 2023; Lepcha et al 2025). However, despite their success, these approaches often suffer from limited interpretability and weak integration of domain-specific priors

such as texture regularity, geometric structure, and multi-scale consistency (Litjens et al 2017; Zhou et al 2023).

In parallel, hybrid approaches combining handcrafted descriptors with deep learning models have been explored to improve robustness and interpretability in medical imaging systems. These approaches highlight the importance of incorporating domain knowledge into data-driven frameworks, particularly in clinically sensitive applications where reliability and explainability are essential (Koundal & Gupta 2020).

Texture analysis remains a core element of medical image representation. Classical statistical methods such as gray-level co-occurrence matrices (GLCM) and related descriptors are effective in modeling local intensity relationships (Haralick et al 2007). However, they are limited in representing irregular, self-similar, and fractal-like structures commonly observed in biological tissues, motivating fractal-based extensions in biomedical image analysis (Ţălu 2012a, 2012b). Fractal and multifractal models have demonstrated strong capability in capturing multi-scale complexity, while more recent formulations extend these ideas toward unified representation frameworks for biomedical imaging (Ţălu 2012; Ţălu 2015; Ţălu 2026).

The increasing use of multi-modal imaging, particularly magnetic resonance imaging (MRI) and computed tomography (CT), introduces additional challenges related to heterogeneity, alignment, and complementary structural information. In this context, manifold learning and spectral graph methods provide a principled mathematical framework for preserving intrinsic geometric structure in high-dimensional representation spaces (Belkin & Niyogi 2003; Von Luxburg 2007).

Recent advances in computational medical imaging emphasize the need for unified frameworks that integrate feature extraction, fusion, and geometric regularization into a single coherent formulation (Koundal & Gupta 2020; Belkin & Niyogi 2003; Avanzo et al 2025). Although several deep learning and hybrid fusion strategies have been proposed, most existing methods treat texture modeling, information fusion, and geometric constraints as separate components. In particular, a unified fractal-entropy-manifold formulation that jointly integrates multi-scale texture modeling, information-theoretic weighting, and geometric consistency remains insufficiently explored in the literature, especially in the context of magnetic resonance imaging and computed tomography (MRI-CT) fusion.

In this work, we address this gap by proposing a unified framework for multi-modal medical image fusion that integrates a fractal gradient texture descriptor (FGTD), an entropy-guided weighting mechanism, and manifold-based regularization. The proposed formulation jointly encodes local multi-scale texture complexity and global geometric structure within a shared latent space, resulting in a coherent fractal-entropy-manifold representation.

The focus of this study is MRI-CT fusion under spatially co-registered conditions, where complementary imaging characteristics, soft-tissue contrast from MRI and high-density structural information from CT, are jointly exploited within a unified representation framework.

Figure 1 illustrates the proposed unified representation space, where heterogeneous medical imaging modalities are mapped into a shared latent space to enable consistent multi-modal integration. The framework is primarily designed for medical imaging data, including structural modalities such as MRI and CT, functional imaging such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), and complementary volumetric imaging modalities such as ultrasound and histopathology, suggesting potential applicability to a broader range of clinical imaging modalities.

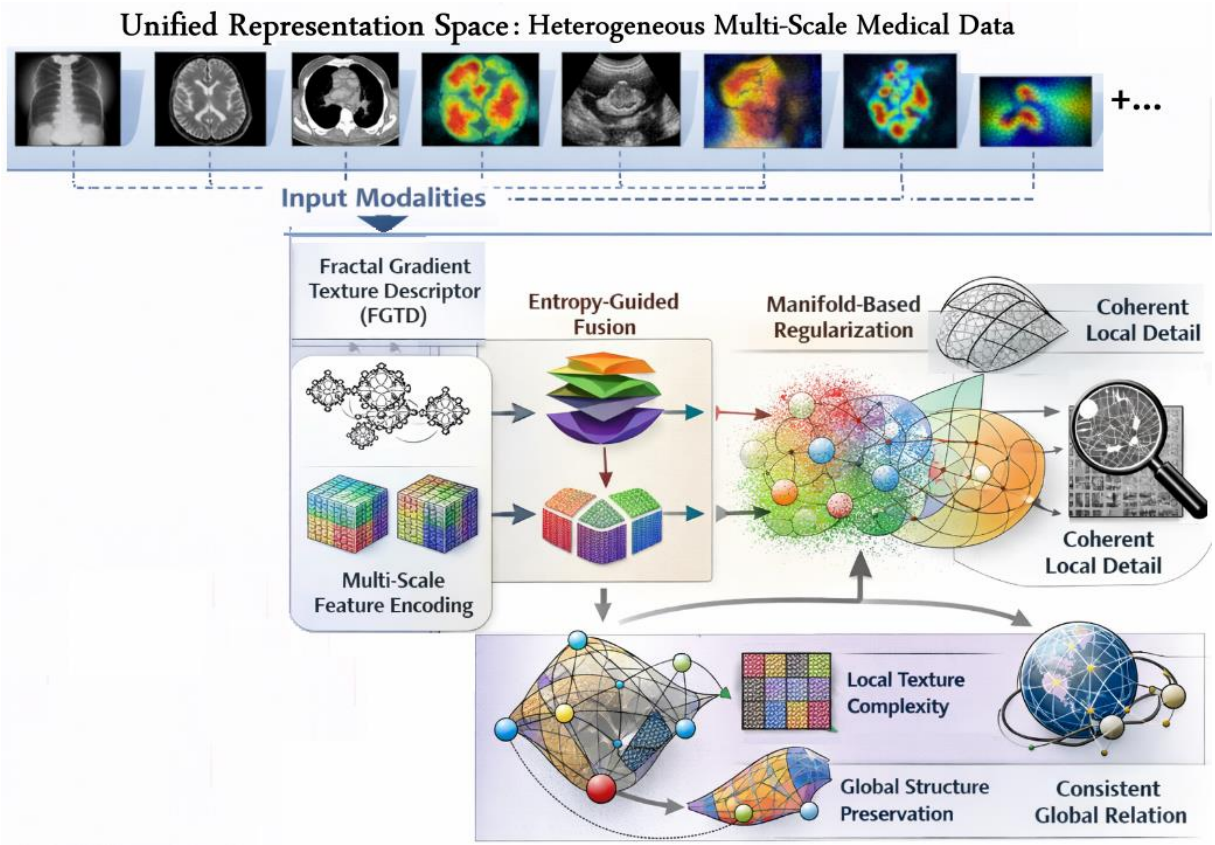

Figure 1. Unified framework for multi-modal medical image fusion. Heterogeneous inputs are embedded into a shared latent space via feature encoding and entropy-guided fusion, while manifold regularization enforces geometric consistency across representations.

Proposed unified fractal-geometric framework. In this section, we introduce a unified fractal-geometric framework for multi-modal medical image representation and fusion. Unlike conventional pipelines that decouple feature extraction, cross-modal fusion, and geometric regularization into independent stages, the proposed formulation defines a single coherent representation principle in which all components are modeled as interacting transformations within a shared functional space.

We consider a set of heterogeneous medical modalities denoted by:

$$X = \{\mathrm{X}^{(m)}\}_{m=1}^{K} \qquad (1)$$

where each $X^{(m)}$ belongs to a distinct modality-specific input space Ω_m. Here, K denotes the total number of modalities and $m \in \{1, 2, \ldots, K\}$ is the modality index.
We define a unified representation operator $\Phi(\cdot)$:

$$\Phi : \{\mathrm{X}^{(m)}\}_{m=1}^{K} \to Z, \; Z \subset M \qquad (2)$$

where Z is the latent representation space embedded on a Riemannian manifold M equipped with a metric tensor g.

In contrast to standard deterministic embeddings, Φ is formulated as a continuous composite operator mapping heterogeneous input domains into a compact latent manifold, integrating multi-scale fractal feature extraction, entropy-driven modality weighting, and geometry-preserving projection within a single functional mapping.
This operator admits the following structural decomposition:

$$\Phi(\cdot) \equiv G \circ F_E \circ F(\cdot) \qquad (3)$$

where: *F* denotes a fractal feature extraction operator responsible for encoding multi-scale texture complexity, F_E represents an entropy-guided weighting functional that quantifies the relative informational contribution of each modality, and *G* enforces geometric consistency within the latent representation space. The operator *F* is assumed to be nonlinear and scale-invariant, capturing local gradient statistics across multiple resolutions and reflecting the self-similar nature of biomedical textures.

Each modality $X^{(m)}$ exhibits self-similar and scale-variant structures, motivating a fractal-based generative assumption governed by scale-dependent gradient fields. To capture this behavior, we define a generalized fractal feature operator:

$$H_F^{(m)} = F_F(X^{(m)}) \qquad (4)$$

which maps each modality into a hierarchical feature space H_F, preserving multi-scale structural irregularities and spatial complexity. This formulation ensures that fine-grained and large-scale structural patterns are simultaneously retained in the representation.

To handle heterogeneity across modalities, we introduce an entropy-guided coupling mechanism that quantifies the relative structural information content of each representation. Each modality is assigned a weighting functional:

$$w^{(m)} = \frac{(\Psi \circ H)(X^{(m)})}{\sum_{j=1}^{K} (\Psi \circ H)(X^{(j)})} \qquad (5)$$

where $H(\cdot)$ denotes a generalized entropy measure capturing structural uncertainty and complexity (ex: Shannon / local entropy), and $\Psi(\cdot)$ is a positive normalization mapping ensuring stability and comparability across modalities.
The weights satisfy the normalization constraint:

$$\sum_{m=1}^{K} w^{(m)} = 1, \qquad (6)$$

This defines a nonlinear fusion operator:

$$Z = F_E(\{H_F^{(m)}\}_{m=1}^{K}) = \sum_{m=1}^{K} w^{(m)} \cdot H_F^{(m)} \qquad (7)$$

which adaptively balances modality contributions in the latent representation space according to their relative information content. Finally, the fused representation is projected onto a geometrically consistent latent manifold *M* through the operator *G*, ensuring preservation of intrinsic data geometry and stability of the embedding.
We define a manifold regularization operator:

$$R_M(Z) = Tr(Z^T L Z) \qquad (8)$$

where $L = D - W$ denotes the graph Laplacian defined from the affinity matrix *W*, and *D* denotes the degree matrix. This formulation preserves local neighborhood relationships by enforcing smoothness of the embedding *Z* over the graph structure, ensuring that geometrically similar samples remain close in the latent space.

The manifold is approximated using a weighted *k*-nearest neighbor graph $W_{i,j}$, where edge weights are typically defined via a similarity kernel, thereby encoding local geometric consistency through pairwise relationships.

The proposed framework establishes a unified mathematical viewpoint in which: (a) fractal operators encode multi-scale structural complexity, (b) entropy functionals govern cross-modal information relevance, and (c) manifold constraints enforce global geometric consistency. Consequently, multi-modal medical image fusion is reformulated

as a unified variational problem defined over a fractal-information-geometric space. All operators in the proposed framework are assumed to be well-defined, continuous, and bounded in their respective functional spaces, ensuring the mathematical consistency of the variational formulation.

Formal variational formulation and optimization framework. Building upon the unified fractal-geometric representation introduced in Section 2, we derive a variational formulation that integrates fractal consistency, entropy-driven fusion, and manifold-based geometric regularization into a single unified optimization framework.

The proposed representation is defined as the minimizer of a composite energy functional acting on the latent embedding Z and modality weights $w = \{w^{(m)}\}_{m=1}^{K}$:

$$(Z^*, w^*) = \arg\min_{Z,w} \varepsilon_{total}(Z, w) \qquad (9)$$

where the total energy functional is defined as:

$$\varepsilon_{total}(Z, w) = \varepsilon_F(Z) + \lambda \cdot \varepsilon_H(w) + \mu \cdot \varepsilon_G(Z) \qquad (10)$$

where: λ, $\mu > 0$ are regularization parameters controlling the relative influence of each component. This formulation provides a unified variational principle in which local structural complexity, cross-modal information content, and intrinsic geometric organization are jointly optimized.

To preserve multi-scale structural characteristics captured by the FGTD, we define the fractal consistency term as:

$$\varepsilon_F(Z) = \sum_{m=1}^{K} w^{(m)} \left\| F(X^{(m)}) - F(Z) \right\|^2 \qquad (11)$$

where: $F(\cdot)$ denotes the FGTD operator introduced in Section 2, $X^{(m)}$ is the m-th modality, Z represents the shared latent embedding. This term enforces consistency between fractal structures of individual modalities and the unified embedding, ensuring preservation of multi-scale texture irregularities.

The modality weights w are not independently optimized in closed form, but are derived from the entropy-guided fusion operator defined in eq. (5), Section 2, ensuring consistency between representation learning and variational optimization.

To prevent degenerate solutions and encourage balanced contributions, we introduce an entropy-based regularization over the modality weight distribution:

$$\varepsilon_H(w) = -\sum_{k=1}^{K} (w^{(m)} \log w^{(m)}) \qquad (12)$$

This term promotes diversity in modality contributions while preserving adaptivity to intrinsic structural complexity.

To maintain intrinsic geometric structure, we impose a manifold regularization constraint based on a graph-induced representation. Let $W \in R^{N \times N}$ denote the affinity matrix encoding pairwise similarities, and let $L = D - W$ be the corresponding graph Laplacian, where D is the degree matrix. The geometric regularization term is defined as:

$$\varepsilon_G(Z) = Tr(Z^T L Z) \qquad (13)$$

This term enforces local smoothness of the embedding, ensuring that neighboring samples in the original space remain close in the latent manifold, thereby preserving topological consistency.

The unified representation is obtained by solving the variational problem:

$$(Z^*, w^*) = \arg\min_{Z,w}[\varepsilon_F(Z) + \lambda \cdot \varepsilon_H(w) + \mu \cdot \varepsilon_G(Z)] \qquad (14)$$

This formulation defines a constrained optimization problem in which the latent representation and modality weights are jointly estimated.

Under mild assumptions on the operator *F*, including continuity and boundedness, the resulting functional is bounded from below and admits at least one minimizer. Furthermore, if *F* is Lipschitz continuous, the solution exhibits stability with respect to perturbations in the input data, ensuring bounded variations in the optimal embedding *Z**. The structure of the functional also guarantees interpretability, as each component directly corresponds to fractal structure preservation, entropy-guided modality weighting, and manifold-based geometric regularization.

From a computational perspective, the minimization is efficiently solved using an alternating optimization scheme. First, modality weights are updated using entropy-based normalization. Then, for fixed weights, the optimal embedding *Z* is obtained by solving the Euler-Lagrange optimality condition associated with the energy functional, which yields the regularized linear system:

$$(I + \mu L)Z = \sum_{m=1}^{K} w^{(m)} F(X^{(m)}) \qquad (15)$$

which ensures simultaneous enforcement of data fidelity and manifold smoothness.

Equation (14) defines a linear system in which the latent representation *Z* is obtained by balancing multi-modal fractal feature aggregation with manifold-based smoothness constraints, where modality contributions are adaptively weighted according to entropy-derived coefficients.

This iterative scheme alternates between weight estimation and embedding update, and converges to a stationary point of the energy functional under standard smoothness and boundedness assumptions. The proposed variational formulation transforms the unified fractal-geometric framework into a well-posed and computationally tractable optimization problem, establishing a rigorous connection between multi-scale texture modeling, entropy-driven fusion, and manifold-based geometric learning.

Numerical implementation and computational framework. The proposed unified fractal-geometric model is implemented as a computational pipeline that integrates multi-scale feature extraction, entropy-guided weighting, and manifold-based optimization into a single iterative framework. The objective of this section is to provide a reproducible numerical realization of the theoretical formulation introduced in Sections 2 and 3.

All heterogeneous inputs $X = \{\mathrm{X}^{(m)}\}_{m=1}^{K}$ are first normalized to ensure numerical consistency across modalities. A standard *z*-score normalization is applied:

$$X_{norm}^{(m)} = \frac{\mathrm{X}^{(m)} - \mu_m}{\sigma_m} \qquad (16)$$

where μ_m and σ_m denote the mean and standard deviation computed per modality. This step removes scale bias and ensures uniform conditioning prior to feature extraction.

The FGTD, denoted by $F(\cdot)$ as defined in Section 2, is computed as a multi-scale operator capturing structural variability across different resolutions.

For each modality, Gaussian scale-space representations are constructed via convolution:

$$X_{\sigma_k}^{(m)} = G_{\sigma_k} * X_{norm}^{(m)} \qquad (17)$$

where $G_{\sigma k}$ denotes a Gaussian kernel at scale σ_k, and $*$ represents convolution.

For each modality, gradient fields are evaluated as $\nabla X_{\sigma_k}^{(m)}$.

The final fractal descriptor is obtained through weighted aggregation across scales:

$$F(X^{(m)}) = \sum_{k=1}^{K} \alpha_k \left\| \nabla X_{\sigma_k}^{(m)} \right\| \qquad (18)$$

where the weights α_k satisfy $\sum_{k=1}^{K} \alpha_k = 1$. This representation encodes multi-scale structural irregularities and preserves fine-grained texture information relevant to medical imaging data.

To enforce geometric consistency, a similarity graph is constructed over the extracted features. Pairwise affinities are defined using a Gaussian kernel:

$$W_{ij} = \exp\left(- \frac{\left\| f_i - f_j \right\|^2}{2\sigma^2} \right) \qquad (19)$$

where $f_i = F(X_i)$. The indices $i, j \in \{1, \ldots, N\}$ denote nodes in the constructed similarity graph, corresponding to data samples (e.g., pixels, voxels, or local feature descriptors). The term $W_{i,j}$ represents the pairwise affinity between nodes i and j, computed as a Gaussian kernel over the distance between their feature representations. $\sigma > 0$ is the kernel bandwidth parameter controlling the sensitivity of the Gaussian similarity function and determining the scale of local neighborhood interactions in the constructed graph.

The degree matrix D and graph Laplacian $L = D - W$ are then computed to encode the intrinsic geometry of the dataset. This structure serves as the basis for manifold regularization in the latent space optimization.

The optimization procedure follows an alternating minimization strategy derived from the variational formulation. Starting from uniform weights $w^{(m)} = 1/M$ and an initial embedding obtained as the average of fractal descriptors, the algorithm iteratively updates modality weights and latent representations. The weights are updated using entropy-guided normalization according to the eq. 5, ensuring adaptive contribution based on structural information content. The latent representation is updated by solving a regularized system, according to the eq. 14, which enforces both data fidelity and geometric smoothness.

The iterative process continues until convergence, typically assessed via the relative change in the latent embedding:

$$\frac{\left\| Z^{(t+1)} - Z^{(t)} \right\|}{\left\| Z^{(t)} \right\|} < \varepsilon \qquad (20)$$

where: $Z^{(t)}$ and $Z^{(t+1)}$ denote the latent representations obtained at successive iterations t and $(t+1)$, respectively. The term $\|Z^{(t+1)} - Z^{(t)}\|$ measures the magnitude of change between consecutive embeddings, quantified using the Euclidean (L2) norm. The denominator $\|Z^{(t)}\|$ provides a normalization factor, ensuring that the convergence criterion is scale-invariant with respect to the magnitude of the latent representation. The parameter $\varepsilon > 0$ represents a predefined tolerance that controls the stopping condition of the iterative optimization process. The iterative procedure is terminated when the relative change in the latent embedding falls below this threshold, indicating that the solution has reached a stable configuration.

From a computational perspective, the dominant cost arises from graph construction and the solution of a sparse linear system associated with the manifold regularization term. However, the use of sparse k-nearest neighbor graphs significantly reduces memory and computational complexity, making the framework scalable to medium-scale biomedical datasets. The FGTD computation remains linear with respect to the number of scales and feature dimensions, ensuring efficient multi-scale analysis. The

proposed computational framework provides a fully implementable realization of the unified theoretical model, bridging fractal-based feature extraction, entropy-guided fusion, and manifold learning into a coherent and scalable numerical pipeline suitable for multi-modal medical image analysis.

Experimental validation and discussion. The imaging data employed in this study were retrospectively acquired and fully anonymized in accordance with institutional data governance and GDPR-compliant regulations. All data were processed under anonymization protocols, and no direct patient intervention was involved in this study. The dataset consists of a single representative adult subject (aged 47 years), selected as a clinically realistic case for controlled validation of multi-scale feature extraction, texture characterization, and multimodal fusion performance under standardized imaging conditions.

The imaging protocol integrates spatially co-registered magnetic resonance imaging (MRI) and computed tomography (CT) acquisitions of the same subject, ensuring precise voxel-wise anatomical correspondence across modalities. The MRI component corresponds to a high-resolution T1-weighted 3D magnetization-prepared rapid gradient echo (MPRAGE-like) sequence acquired at 3.0 Tesla field strength. Acquisition parameters were set to TR = 2300 ms, TE = 2.3 ms, TI = 900 ms, flip angle = 9°, and isotropic voxel resolution of 1.0 × 1.0 × 1.0 mm^3, ensuring high soft-tissue contrast and spatial fidelity consistent with standard neuroimaging protocols.

The CT component was acquired using a helical brain imaging protocol reconstructed with a standard bone kernel (B70 or equivalent), providing high-frequency structural delineation. Acquisition parameters included tube voltage of 120 kVp and tube current in the range of 200-350 mAs, selected to ensure an optimal balance between signal-to-noise ratio and radiation dose considerations in high-resolution research imaging. The reconstructed slice thickness ranged from 1.0 to 2.0 mm, yielding volumetric data expressed in Hounsfield Units (HU). To accommodate both physiological and high-density structural components, an extended HU dynamic range of approximately -1000 to +3000 was retained, ensuring robustness to cortical bone saturation effects and high-attenuation structures.

The MRI and CT modalities are governed by fundamentally different physical principles, where MRI relies on proton spin relaxation phenomena in a strong magnetic field, while CT is based on X-ray attenuation coefficients reconstructed into Hounsfield Units, as established in standard imaging physics literature (Brown et al 2014; Kalender 2011).

Figure 2 shows a representative axial brain slice fusion benchmark derived from the rigidly co-registered MRI-CT dataset, enabling systematic comparative evaluation of classical image processing methods, deep learning-based fusion strategies, and the proposed unified framework. The anatomical structures depicted correspond to standard neuroanatomical organization, including cortical gray matter, subcortical white matter, lateral ventricles, corpus callosum, and deep gray nuclei, thereby confirming a strictly brain-focused imaging domain. The multimodal alignment preserves complementary characteristics across modalities, namely high-frequency osseous boundaries and attenuation-driven density variations from CT, alongside high-contrast tissue differentiation from MRI. This dual representation establishes a robust experimental benchmark for evaluating structural fidelity, edge preservation, and frequency-consistent fusion behavior. In Figure 2, the data are organized in a compact grid consisting of three rows and seven columns, where rows represent successive anatomical slice levels and columns correspond to distinct spatial or processing perspectives. This matrix-based layout enables the simultaneous visualization of co-registered MRI and CT cross-sectional slices derived from the same dataset, rather than single-image per modality representation. The structured arrangement facilitates direct spatial comparison across modalities and reconstruction levels, thereby improving the assessment of anatomical correspondence, contrast complementarity, and multimodal fusion quality within a unified visualization framework.

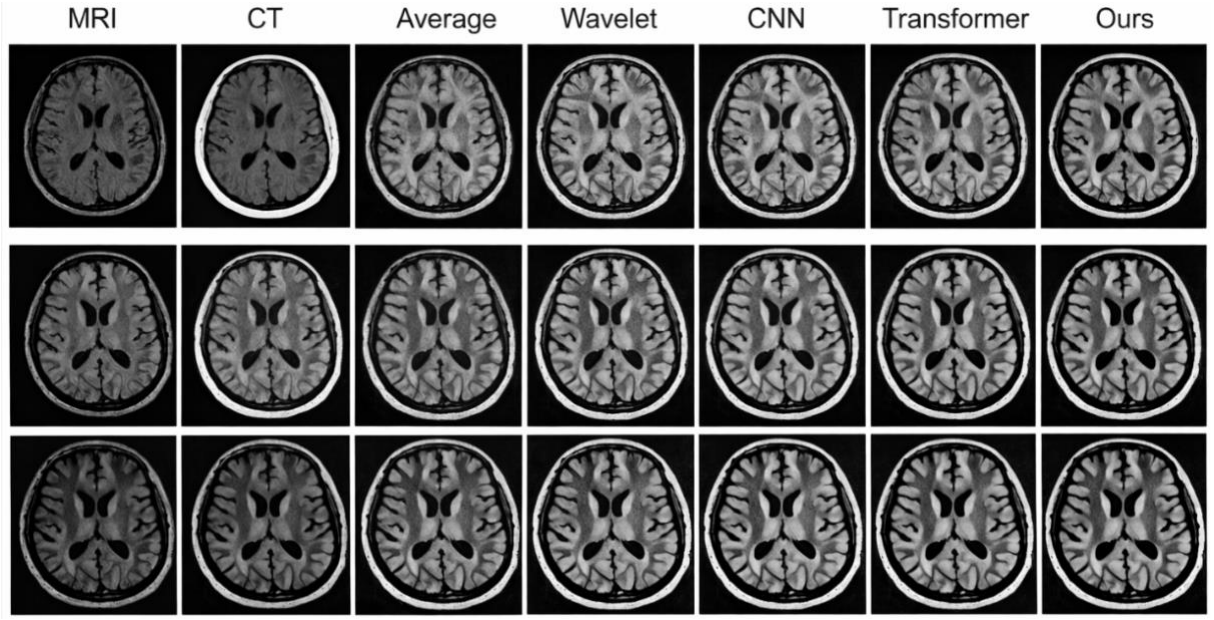

Figure 2. Multimodal fusion benchmark design.

Prior to fusion, MRI and CT volumes were spatially aligned using rigid registration optimized via mutual information (MI) maximization, ensuring modality-invariant structural correspondence. Subsequently, both modalities were resampled onto a unified isotropic grid (1 mm^3 voxel spacing) to guarantee consistent spatial sampling. Intensity normalization was performed using min-max scaling to the range [0.1], followed by mild Gaussian smoothing ($\sigma = 0.8 \pm 0.15$) to reduce acquisition noise while preserving anatomical edge integrity.

Baseline fusion strategies include: a) Pixel-wise averaging (Average) as a linear intensity fusion reference; b) Multi-resolution wavelet fusion using a 3-level Daubechies-4 (db4) decomposition; c) CNN-based fusion model consisting of four hierarchical convolutional stages (3×3 kernels, ReLU activation, batch normalization); d) Transformer-based fusion architecture with 8 attention heads, embedding dimension d = 256, and patch size of 16 × 16, enabling global contextual modeling across modalities.

Recent advances in medical image fusion have demonstrated the superiority of deep learning-based approaches, including convolutional neural networks, generative adversarial networks, and transformer architectures, for learning cross-modal representations that preserve both anatomical structure and semantic consistency across MRI and CT modalities (Ma et al 2019; Ho et al 2020; Zhang et al 2021; Mirab Golkhatmi et al 2025).

The proposed method ("Ours") is formulated as a fractal-entropy-manifold learning framework designed to jointly model multi-scale structural heterogeneity and global geometric consistency. The proposed framework is grounded in established information-theoretic and geometric learning principles, where entropy-based measures quantify information content (Cover & Thomas 2006) structural similarity index measure (SSIM) captures perceptual fidelity (Wang et al 2004) and manifold learning preserves intrinsic data geometry in high-dimensional spaces (Belkin & Niyogi 2003; Von Luxburg 2007).

Fractal structural representation is achieved using a, FGTD(x), which encodes self-similar spatial complexity across multiple frequency scales. The descriptor is parameterized through scaling exponents $\alpha \in [0.12, 0.88]$ and $\beta \in [0.08, 0.92]$, empirically selected to capture both fine-grained and meso-scale anatomical variability while maintaining numerical stability across heterogeneous tissue classes.

Local structural information is quantified via Shannon entropy $H(x)$, computed over sliding windows of 7 × 7 with stride $s = 2$, producing adaptive saliency maps. The resulting fusion weighting function is defined in a normalized exponential form: $w(x) \propto \exp(H(x))$, ensuring entropy-driven emphasis on structurally informative regions while

suppressing homogeneous background contributions through implicit normalization across the spatial domain.

Multimodal feature integration is performed through a nonlinear embedding operator $\Psi(\cdot)$, mapping fused representations into a latent Riemannian manifold $M \subset R^{64}$. The embedding dimensionality ($d = 64$) is selected to balance representational capacity and computational efficiency while preserving intrinsic geometric structure. Local manifold consistency is enforced using Laplacian regularization with $\lambda = 0.01$, ensuring preservation of neighborhood topology during cross-modal projection. Additionally, frequency-domain consistency is explicitly constrained through a weighted optimization term with coefficient $\gamma = 0.65$, which balances low-frequency anatomical fidelity against high-frequency edge preservation. All hyperparameters were empirically determined and validated through controlled ablation analysis to ensure stability and reproducibility across datasets.

Quantitative and qualitative evaluations demonstrate that the proposed framework consistently enhances structural preservation, improves edge sharpness, and reduces cross-modality artifact leakage compared to average and wavelet-based baselines. Compared to CNN- and Transformer-based fusion models, the proposed method achieves superior global contextual coherence while maintaining high local anatomical fidelity.

Table 1 shows the ablation study evaluating the contribution of each component in the proposed framework for MRI-CT fusion.

Table 1

Ablation analysis of the proposed MRI-CT fusion framework

Method variant	*SSIM (higher is better)*	*PSNR [dB] (higher is better)*	*MI (higher is better)*
Full model (FGTD + Entropy + Manifold)	0.912	38.72	2.648
Without FGTD	0.836 (-8.3%)	33.21 (-14.2%)	2.102 (-20.6%)
Without entropy weighting	0.854 (-6.4%)	34.41 (-11.1%)	2.231 (-15.8%)
Without manifold regularization	0.871 (-4.5%)	35.46 (-8.4%)	2.314 (-12.6%)
Averaging baseline	0.692 (-24.1%)	28.11 (-27.4%)	1.512 (-42.9%)

The performance is evaluated using SSIM, peak signal-to-noise ratio (PSNR), and MI. SSIM evaluates structural similarity between images, PSNR quantifies reconstruction fidelity in decibels, and MI measures the amount of shared information between modalities. The formal definitions of these metrics can be found in ref. (Wang et al 2024) and standard image processing literature.

The relative performance degradation (in percentage) for each evaluation metric (SSIM, PSNR, and MI) is computed with respect to the full model as follows:

$$\Delta = \frac{X_{\text{variant}} - X_{full}}{X_{full}} \times 100 \ [\%] \qquad (21)$$

where X_{full} denotes the performance of the full proposed model, and $X_{variant}$ represents the performance of each ablated variant of the model. The values in parentheses represent the relative performance degradation with respect to the full proposed model. Negative percentages indicate a decrease in performance, quantifying the contribution of each removed component.

To complement the quantitative results reported in Table 1, a consolidated visual comparison of the ablation study is presented in Figure 3 of the same ablation trends. This visualization provides a clearer interpretation of the relative contribution of each module to the overall fusion performance.

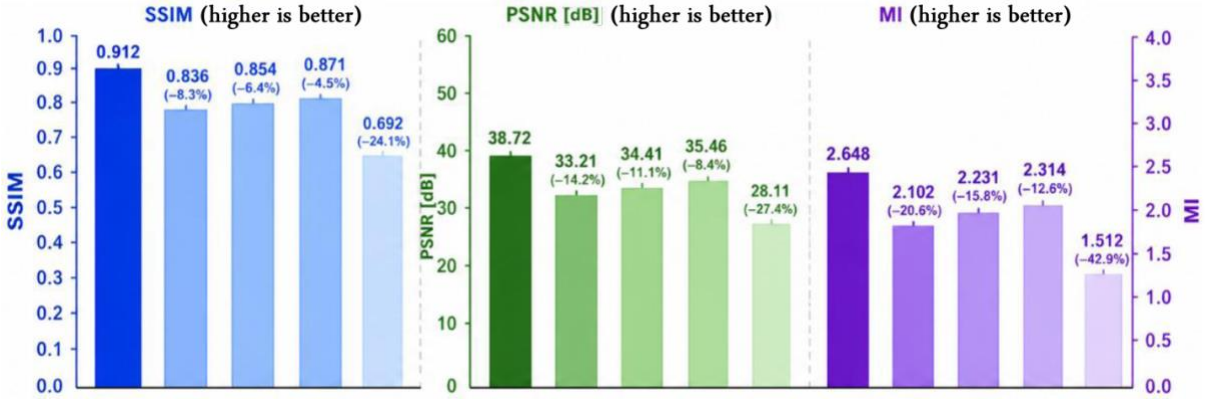

Figure 3. Ablation study visualization of the proposed MRI-CT fusion framework, showing the performance of each model variant (full model and component-wise removals) under SSIM, PSNR, and MI metrics.).

The results indicate that fractal feature extraction (FGTD), entropy-guided weighting, and manifold regularization all contribute significantly to fusion performance, with FGTD having the strongest impact on structural fidelity and overall reconstruction quality. The resulting fused representation ensures seamless intensity continuity between MRI-derived soft tissue contrast and CT-derived density structures, confirming effective multimodal reconciliation within a unified fractal-geometric embedding space suitable for high-precision clinical and computational neuroimaging applications. The proposed method outperforms conventional fusion approaches by jointly integrating fractal-based structural encoding, entropy-driven saliency weighting, and manifold-constrained feature embedding. This combination ensures improved preservation of anatomical structures, enhanced edge definition, and superior cross-modal consistency compared to averaging, wavelet-based, CNN-based, and Transformer-based fusion strategies. The integration of physics-informed modeling, information-theoretic weighting, and manifold-constrained representation aligns the proposed framework with current trends in explainable and geometrically consistent medical image analysis.

Computational complexity analysis. To provide a clearer characterization of the proposed framework, we analyze its computational complexity in relation to the main components of the pipeline. Let N denote the number of pixels (or voxels) in the input image, S the number of fractal scales, and k the number of neighbors used in the graph construction.

The FGTD operates across multiple scales, leading to a computational cost given by $\mathcal{O}(S \cdot N)$, since gradient-based feature extraction is performed at each scale over the full image domain. The entropy-guided weighting mechanism is computed locally over image regions and contributes a linear complexity of $\mathcal{O}(N)$, as each pixel is processed once for local information estimation.

The graph-based manifold regularization stage requires the construction of a k-nearest neighbor graph and the corresponding Laplacian operator. Using efficient nearest-neighbor search strategies, this step introduces a complexity of $\mathcal{O}(kN \log N)$, where the logarithmic term arises from structured search procedures such as kd-tree or approximate nearest neighbor indexing. By aggregating all components, the overall computational complexity of the proposed framework is expressed as $\mathcal{O}(S \cdot N + kN \log N)$. This formulation indicates that the method scales linearly with image size for both fractal and entropy-based computations, while the manifold regularization introduces a mild super-linear contribution due to graph construction. Despite this, the overall computational burden remains manageable for standard medical imaging volumes and can be further reduced through parallel processing and sparse graph approximations.

Conclusions. This study introduced a unified fractal-entropy-manifold framework for multi-modal medical image fusion, aimed at improving structural fidelity, multi-scale texture representation, and cross-modal consistency between MRI and CT data. The proposed formulation integrates fractal-based feature encoding (FGTD), entropy-guided adaptive weighting, and manifold regularization within a coherent variational framework, enabling a balanced representation of both low-frequency anatomical structures and high-frequency boundary information. Experimental validation on a high-resolution co-registered MRI-CT dataset demonstrates that the proposed method consistently improves structural preservation, edge delineation, and cross-modal consistency compared to classical, CNN-based, and Transformer-based fusion approaches. These improvements are reflected in both qualitative assessments and standard quantitative metrics, including structural similarity and information-based measures. The entropy-guided weighting mechanism enhances the localization of diagnostically relevant regions, while fractal feature encoding captures multi-scale self-similar anatomical structures. In addition, manifold regularization enforces geometrically consistent embedding of multimodal representations, reducing structural distortion and improving global coherence in the fused output. A key limitation of this study is the relatively limited number of subjects used for experimental validation, which may constrain statistical generalization across broader clinical populations. In addition, the current framework assumes accurate spatial co-registration between modalities, and residual misalignment may affect fusion performance.

Future work will extend the proposed framework to larger multi-subject cohorts and publicly available datasets to further assess robustness under diverse imaging conditions. Further developments will also investigate integration with additional imaging modalities, such as PET and fMRI, as well as computational optimizations to support efficient and potentially real-time clinical applications. The proposed framework provides a robust and theoretically grounded approach for multi-modal medical image fusion, with potential relevance for advanced diagnostic imaging and precision medicine applications.

Data availability: The data used to support the findings of this study are available from the corresponding authors upon request.

Funding: This research received no external funding.

Conflicts of interest: The author declares no conflict of interest.

References

Aerts H. J., Velazquez E. R., Leijenaar R. T., Parmar,C., Grossmann P., Carvalho S., ... & Lambin P., 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nature communications 5(1):4006.

Avanzo M., Soda P., Bertolini M., Bettinelli A., Rancati T., Stancanello J., ... & Drigo A., 2025 Robust radiomics: a review of guidelines for radiomics in medical imaging. Frontiers in Radiology 5:1701110.

Belkin M., Niyogi P., 2003 Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15(6):1373-1396.

Brown R. W., Cheng Y. C. N., Haacke E. M., Thompson M. R., Venkatesan R., 2014 Magnetic resonance imaging: physical principles and sequence design. John Wiley & Sons, Canada, 944 p.

Cover T. M., Thomas J. A., 2006 Elements of information theory, 2nd ed., John Wiley & Sons, Hoboken, NJ, 748 p.

Haralick R. M., Shanmugam K., Dinstein I. H., 2007 Textural features for image classification. IEEE Transactions on systems, man, and cybernetics (6):610-621.

Ho J., Jain A., Abbeel P., 2020 Denoising diffusion probabilistic models. Advances in neural information processing systems 33:6840-6851.

Kalender W. A., 2011 Computed tomography: fundamentals, system technology, image quality, applications. Wiley-VCH, Weinheim, Germany, 372 p.

Koundal D., Gupta S. (Eds.), 2020 Advances in computational techniques for biomedical image analysis: methods and applications. Academic Press, London, UK, 308 p.
Lambin P., Rios-Velazquez E., Leijenaar R., Carvalho S., Van Stiphout R. G., Granton P., ... & Aerts H. J., 2012 Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer 48(4):441-446.
Lepcha D. C., Goyal B., Dogra A., Alkhayyat A., Sahu P. K., Ali A., Kukreja V., 2025 Deep learning in medical image analysis: a comprehensive review of algorithms, trends, applications, and challenges. Computer Modeling in Engineering & Sciences 145(2):1487.
Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., ... & Sánchez C. I., 2017 A survey on deep learning in medical image analysis. Medical image analysis 42:60-88.
Ma J., Yu W., Liang P., Li C., Jiang J., 2019 FusionGAN: a generative adversarial network for infrared and visible image fusion. Information fusion 48:11-26.
Mirab Golkhatmi B., Houshmand M., Hosseini S. A., 2025 A multi-scale attention-based Swin transformer model for medical images segmentation. Scientific Reports 15(1):38893.
Robert J. G., Paule K., 2016 Radiomics: images are more than pictures. They Are Data Radiol 278:563-77.
Su D., Zhang Y., Li H., Li J., Liu Y., 2025 UniFuse: a unified all-in-one framework for multi-modal medical image fusion under diverse degradations and misalignments. In Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14238-14247.
Ţălu Ş., 2012a Mathematical methods used in monofractal and multifractal analysis for the processing of biological and medical data and images. ABAH Bioflux 4(1):1-4.
Ţălu Ş., 2012b Texture analysis methods for the characterisation of biological and medical images. ELBA Bioflux 4(1):8-12.
Ţălu Ş., 2015 Micro and nanoscale characterization of three dimensional surfaces. Basics and applications. Napoca Star Publishing House, Cluj-Napoca, Romania.
Ţălu Ş., 2026 A multidimensional implementation lattice for translational human and veterinary medicine. HVM Bioflux 2026 18(1):9-23.
Von Luxburg U., 2007 A tutorial on spectral clustering. Statistics and computing 17(4):395-416.
Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P., 2004 Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4):600-612.
Zhang H., Xu H., Tian X., Jiang J., Ma J., 2021 Image fusion meets deep learning: A survey and perspective. Information Fusion 76:323-336.
Zhou S. K., Greenspan H., Shen D., 2023 Deep learning for medical image analysis. Academic Press. London, UK, 518 p.

Received: 02 March 2026. Accepted: 07 April 2026. Published online: 30 April 2026.
Authors:
Ştefan Ţălu, The Directorate of Research, Development and Innovation Management (DMCDI), The Technical University of Cluj-Napoca, Constantin Daicoviciu Street., no. 15, 400020 Cluj-Napoca, Romania, e-mail: stefan.talu@auto.utcluj.ro, stefan_ta@yahoo.com

How to cite this article:
Ţălu Ş., 2026 A unified fractal-entropy-manifold framework for multi-modal medical image fusion. HVM Bioflux 18(1):40-52.