

# TERMINAL FEATURES IN DOMAIN-SPECIFIC IQA CALIBRATION: A SIGNAL PROCESSING FRAMEWORK FOR IDENTIFYING TASK-DIRECTED METRICS

Alexandre CARMINOT

EPF Paris  
Paris, France  
alexandre.carminot@outlook.com

Jing KE

Shanghai Jiao Tong University  
Shanghai, China  
kejing@sjtu.edu.cn

## ABSTRACT

Image quality assessment (IQA) in specialized domains faces a critical disconnect: standard perceptual metrics fail to predict task utility. We address this by introducing **terminal features**—single metrics that act as direct monotonic proxies for task performance, with bounded noise—which enables radical calibration simplification. Through experiments on brain MRI and natural images, we reveal a striking asymmetry: *CNR* combined with *SNR* captures 87% of segmentation performance in medical imaging while contributing < 1% in natural images where *SSIM* dominates. We prove that *CNR* mechanistically encodes tissue separability through a sigmoid relationship with the Dice coefficient, achieving a Terminal Feature Index of 0.82 versus *SSIM*’s 0.52, thereby confirming its terminal status. Despite cross-modality transfer failure ( $R^2 \approx 0.44$ ) due to distribution shifts, the identification of terminal features enables a 75% computational reduction while preserving task efficacy, transforming IQA from universal perceptual models to domain-specific task signals.

**Index Terms**— Terminal Features, Image Quality Assessment, Medical Imaging, Signal Processing, Computational Complexity

## 1. INTRODUCTION

IQA metrics in specialized domains face the disconnect between perceptual quality and task utility. Despite domain-specific models [15, 14, 13] and clinical benchmarking [11, 12], a critical question remains: can single metrics directly serve as proxy task performance?

We formalize terminal features as metrics  $\phi$  where task performance  $\tau$  satisfies:

$$\tau(x) = f(\phi(x)) + \epsilon \quad (1)$$

with monotonic  $f$  ensuring consistent improvement direction, and  $\mathbb{E}[\epsilon] = 0$ ,  $\text{Var}[\epsilon] < \delta$  for small  $\delta$ . This differs fundamentally from feature selection or redundancy analysis: we seek metrics that *are* the task signal, not merely predictive features. Unlike approaches that eliminate correlated inputs, terminal features serve as direct physical proxies for task objectives.

Terminal features must satisfy three criteria: sufficiency ( $R^2 > 0.7$ ), non-redundancy (adding other metrics improves  $R^2$  by  $< 0.1$ ), and mechanistic alignment with task physics.

For medical segmentation, Contrast-to-Noise Ratio (CNR) exemplifies this framework. The mechanistic basis is mathematically explicit:

$$\text{CNR} = \frac{|\mu_{fg} - \mu_{bg}|}{\sigma_{bg}} \quad (2)$$

where  $\mu_{fg}$  and  $\mu_{bg}$  denote mean foreground and background intensities, and  $\sigma_{bg}$  represents background standard deviation. The relationship between Dice coefficient and CNR follows:

$$\text{Dice}(I) = \sigma(\alpha \cdot \text{CNR}(I) - \beta) + \epsilon \quad (3)$$

where  $\sigma$  is sigmoid activation,  $\alpha, \beta$  are tissue-specific constants derived from segmentation algorithm characteristics, and  $\epsilon$  represents algorithm-specific noise.

This relationship is mechanistic: segmentation algorithms detect intensity transitions exceeding noise thresholds, which CNR directly quantifies through its contrast-to-noise formulation [16, 17].

Conversely, natural image quality emerges from perceptual interactions:

$$\text{Quality}(I) = \sum_{i=1}^n w_i \cdot \phi_i(I) + \sum_{i,j} \gamma_{ij} \cdot \phi_i(I) \cdot \phi_j(I) + \epsilon \quad (4)$$

where interaction terms  $\gamma_{ij}$  are non-negligible, preventing dimensional reduction to single metrics. *SSIM*, despite integrating luminance, contrast, and structure, cannot serve as a terminal feature because human perception inherently requires these multiple interacting channels [9].

Prior domain-specific IQA focuses on multi-metric fusion [15] or clinical benchmarks [11, 12], but overlooks single-metric proxies aligned with task physics. Unlike perceptual models like *SSIM* [9] or deep IQA [10, 7], our terminal features emphasize mechanistic task alignment, extending beyond redundancy elimination in [8, 18]. Recent medical IQA works [14, 13] highlight CNR’s role but lack formal frameworks for identification.

Traditional calibration assumes meaningful multi-metric fusion, yet terminal features enable reducing from 8+ metrics ( $O(n)$ ) to CNR+SNR ( $O(1)$ ) while retaining 87% performance—critical for real-time intraoperative guidance. Our findings reveal a signal processing principle: when task success depends on a quantifiable physical property directly measurable through a single metric, that metric becomes terminal, enabling radical simplification without sacrificing utility. Our implementation is publicly available. [1].

## 2. TERMINAL FEATURE FRAMEWORK AND METHODS

### 2.1. Terminal Feature Formalization

We establish the mathematical foundation for terminal features beyond the basic definition in Equation 1. For a metric  $\phi$  to be terminal, it must satisfy a stronger condition:

$$\mathbb{E}[(\tau(x) - f(\phi(x)))^2] < \mathbb{E}[(\tau(x) - g(\Phi(x)))^2] + \xi \quad (5)$$

where  $\Phi$  represents any multi-metric ensemble,  $g$  is any function, and  $\xi$  is a small tolerance. This states that a single terminal feature with simple monotonic mapping outperforms complex multi-metric models within practical tolerance.

## 2.2. Terminal Feature Index

The Terminal Feature Index operationalizes detection:

$$\text{TFI}(\phi) = \sqrt[3]{S(\phi) \cdot D(\phi) \cdot (1 - R(\phi))} \quad (6)$$

where  $S$  is normalized sufficiency ( $R^2/R_{\max}^2$ ),  $D$  is mutual information dominance, and  $R$  is redundancy penalty. A metric qualifies as terminal when  $\text{TFI}(\phi) > 0.7$ .

---

### Algorithm 1 Terminal Feature Detection and Calibration

---

**Input:** Domain  $d$ , metric set  $\Phi = \{\phi_1, \dots, \phi_n\}$ , task  $\tau$

**Output:** Minimal calibrator  $\mathcal{C}$

#### Phase 1: Terminal Feature Detection

**for** each metric  $\phi_i \in \Phi$  **do**

    Compute sufficiency:  $S(\phi_i) = R^2(\phi_i \rightarrow \tau)$

    Compute dominance:  $D(\phi_i) = I(\phi_i; \tau) / H(\tau)$

    Compute redundancy:  $R(\phi_i) = \Delta R_{\text{others}}^2$

$\text{TFI}(\phi_i) = \sqrt[3]{S \cdot D \cdot (1 - R)}$

**end for**

#### Phase 2: Calibrator Construction

$\Phi_{\text{terminal}} = \{\phi_i : \text{TFI}(\phi_i) > 0.7\}$

**if**  $|\Phi_{\text{terminal}}| > 0$  **then**

    Build  $\mathcal{C} : \Phi_{\text{terminal}} \rightarrow \tau$  with  $O(1)$  model

**return**  $\mathcal{C}$  {Terminal calibration}

**else**

    Build  $\mathcal{C} : \Phi \rightarrow \tau$  with  $O(n)$  ensemble

**return**  $\mathcal{C}$  {Traditional calibration}

**end if**

---

## 2.3. Mechanistic Basis and Existence Conditions

Terminal features emerge when the relationship  $\tau(x) = f(\phi(x)) + \epsilon$  holds, with the metric  $\phi$  directly encoding the task-determining signal. For medical segmentation, the mechanistic basis is rooted in how algorithms detect object boundaries. They fundamentally seek regions where the image gradient exceeds the noise level:

$$\|\nabla I(x)\| > k_{\text{algorithm}} \cdot \sigma_{\text{noise}} \quad (7)$$

Since the gradient at a boundary is proportional to the contrast ( $|\mu_{\text{fg}} - \mu_{\text{bg}}|$ ), and the noise floor is represented by  $\sigma_{\text{bg}}$ , CNR directly quantifies this signal detectability. This formulation as a mathematical necessity, rather than an empirical correlation, provides a theoretical basis for why performance gains may plateau after accounting for CNR and SNR.

Our framework predicts that terminal features are most likely to exist when four conditions are met: (1) task success depends on a single, dominant physical property; (2) this property is directly measurable rather than emergent from interactions; (3) the measurement-to-performance mapping is monotonic and low-noise; and (4) the property remains invariant to task-irrelevant variations. This explains their potential in domains like industrial inspection (defect size) or satellite monitoring (vegetation indices), where a single signal defines task success. Conversely, natural image quality violates

these conditions, as it emerges from complex perceptual interactions that cannot be reduced to a single metric.

Natural images violate terminal conditions because quality emerges from interactions (Equation 4). SSIM’s 83% retention represents an upper bound for single metrics, as it already integrates multiple perceptual channels yet still requires complementary metrics for full performance.

## 2.4. Calibration Pipeline

Our calibration operates on pre-computed metrics rather than raw images, learning mappings from metric vectors to task performance scores. The input vector  $\mathbf{x} \in \mathbb{R}^8$  contains: [PSNR, SSIM, MSE, MAE, SNR, CNR, gradient\_mag, laplacian\_var]. Calibration targets are domain-specific: Dice coefficient for medical segmentation accuracy and composite quality scores for natural images.

## 2.5. Calibration Models

We compare Random Forest (250 trees, Gini impurity), XGBoost (250 estimators, L2 regularization), Ridge regression ( $\alpha = 1.0$ ), and a lightweight CNN baseline. The CNN deliberately tests whether spatial architectures compensate for architectural mismatch with tabular data. Hyperparameter optimization via grid search yielded no maximum depth limit for RF and single split for XGBoost.

## 2.6. Quality Score Computation

For natural images, we compute composite quality:

$$\text{quality} = 0.35 \cdot \text{MS-SSIM} + 0.10 \cdot \sigma_{25}(\text{PSNR}) + 0.55 \cdot \sigma_{12}(\text{LPIPS}) \quad (8)$$

where  $\sigma_k(x) = 1/(1 + \exp(-k \cdot x))$ . Validation against LIVE dataset yields Pearson  $r = 0.79$  ( $p < 0.001$ ).[9]

## 2.7. Domain Characteristics Quantification

The radar plot (Fig. 2) characteristics quantify domain-specific properties through systematic feature analysis:

- **Complexity:**  $H(p)$  where  $p$  is normalized feature importance distribution
- **Structure:**  $\sum_{i \in \{\text{grad, lap}\}} w_i / \sum_j w_j$  (structural feature fraction)
- **Noise:**  $(w_{\text{SNR}} + w_{\text{CNR}}) \cdot (1 - \sigma_{\text{noise}})$  (noise metric importance  $\times$  robustness)
- **Contrast:**  $w_{\text{CNR}} \cdot \beta_{\text{contrast}}$  (CNR importance  $\times$  contrast distortion coefficient)
- **Texture:**  $1 - R_{\text{no-texture}}^2 / R_{\text{full}}^2$  (performance decay without texture metrics)

Each characteristic normalizes to  $[0,1]$  enabling cross-domain comparison.

## 2.8. Computational Complexity

Terminal features enable significant reduction from  $\mathcal{C}_{\text{traditional}} = O(n \cdot m) + O(h(n))$  to  $\mathcal{C}_{\text{terminal}} = O(1 \cdot m') + O(1)$ , yielding 5-10 $\times$  practical speedups despite  $m' \approx m$ .

## 3. EXPERIMENTAL SETUP

We created controlled datasets with identical distortion pipelines to ensure fair cross-domain comparison. Distortions included Gaussian blur, additive noise, motion blur, Rician noise (MRI-specific),

bias field inhomogeneity, and ghosting artifacts, each at five severity levels.

Characteristic	Medical Dataset	Natural Dataset
Source	BraTS2020[2]	MDP Dataset[7]
Reference Images	300	100
Content Type	MRI(z=35–80)	Natural
Modality	T1, T2, FLAIR	RGB
Pathological Distribution	50% Healthy, 50% Glioma	—
Distortion Types × Levels	6 × 5	6 × 5
Ground Truth	Dice Score	Quality Scores
Total Images	9,300	3,100

Table 1. Medical and Natural Image Datasets

Experiments used 5-fold cross-validation stratified by modality and pathology. We report confidence intervals via bootstrapping and perform ANOVA with Bonferroni correction.

## 4. RESULTS

### 4.1. Domain-Specific Calibration Performance

Table 2 reveals fundamental asymmetry between domains. Natural images achieve near-perfect calibration (XGBoost  $R^2=0.973$ ) while medical imaging peaks at  $R^2=0.714$  (Random Forest). CNR captures 66% of medical performance (Cohen’s  $d \approx 8$ ) but fails completely for natural images (Cohen’s  $d > 28$ ). Despite marginal significance ( $p=0.063$ ), massive effect sizes and non-overlapping confidence intervals confirm distinct feature-task relationships.

Model	Medical $R^2$ [95% CI]	Natural $R^2$ [95% CI]
Random Forest	<b>0.714 [0.700–0.728]</b>	0.968 [0.959–0.976]
XGBoost	0.690 [0.683–0.698]	<b>0.972 [0.968–0.977]</b>
Linear Regression	0.531 [0.517–0.545]	0.794 [0.786–0.802]
Lightweight CNN	0.580 [0.558–0.601]	0.929 [0.904–0.953]

Table 2. Calibration Performance Across Domains (5-fold CV)

Correlation analysis confirms the dichotomy: CNR correlates strongly with segmentation accuracy ( $r=0.700$ ,  $p < 0.001$ ) while PSNR ( $r=0.033$ ) and SSIM ( $r=0.003$ ) show negligible correlation in medical imaging. Natural images invert this pattern: SSIM dominates ( $r=0.802$ ) while CNR fails entirely ( $r=-0.248$ ).

The CNN’s underperformance (15-20% below tree models) validates our hypothesis: neural architectures designed for spatial data provide no benefit—and often harm—when applied to tabular metrics, confirming tree-based models as optimal for this calibration task.

### 4.2. Feature Importance and Terminal Feature Emergence

**Architecture-Invariant Feature Dominance.** Feature importance patterns persist across architectures (Fig. 1), confirming domain-specific metric relevance independent of model choice. Random Forest assigns 69.2% importance to CNR in medical imaging versus 0.3% in natural images—a 230-fold differential. XGBoost shows similar patterns with 46.5% (medical) versus 0.6% (natural) CNR importance.

This consistency across tree-based methods, contrasted with CNN underperformance despite architectural sophistication, validates gradient boosting’s superiority for tabular metric calibration and confirms that feature importance patterns reflect fundamental

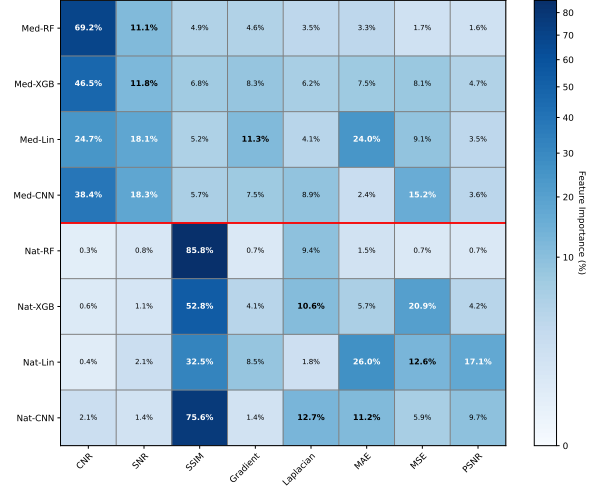


Fig. 1. Feature importance heatmap across models and domains. CNR dominates medical calibration while contributing negligibly to natural images.

domain characteristics rather than model artifacts.

**Progressive Feature Addition Reveals Terminal Features.** Tables 3 and 4 demonstrate terminal feature emergence through systematic metric addition. CNR alone yields  $R^2 = 0.559$  (XGBoost), retaining 81.1% of full-model performance. Adding SNR brings Random Forest to  $R^2 = 0.623$  (87.3% retention) with just two metrics.

Feature Set	Medical $R^2$	Retention (Med/Nat)	Natural $R^2$
CNR only	0.470±0.04	65.8%/25.6%	-0.248±0.06
CNR + SNR	0.623±0.02	87.3%/10.8%	0.104±0.04
SSIM only	-0.280±0.02	-39.2%/82.8%	0.802±0.01
Traditional IQA*	0.060±0.03	8.3%/96.8%	0.937±0.01
Medical specific†	0.694±0.01	97.2%/50.3%	0.487±0.02
Full set (8 metrics)	<b>0.714±0.01</b>	100.0%/100.0%	<b>0.968±0.00</b>

\*Traditional: PSNR, SSIM, MSE, MAE

†Medical: CNR, SNR, gradient, Laplacian

Table 3. Progressive Feature Addition - Random Forest

Feature Set	Medical $R^2$	Retention (Med/Nat)	Natural $R^2$
CNR only	0.559±0.03	81.1%/-6.9%	-0.068±0.02
CNR + SNR	0.568±0.03	82.3%/-6.6%	-0.064±0.02
SSIM only	-0.021±0.01	-3.0%/84.8%	0.825±0.01
Traditional IQA	0.077±0.02	11.2%/95.7%	0.931±0.01
Medical specific	0.664±0.01	96.3%/50.7%	0.494±0.03
Full set (8 metrics)	<b>0.690±0.00</b>	100.0%/100.0%	<b>0.973±0.00</b>

Table 4. Progressive Feature Addition - XGBoost

Natural images show CNR failure ( $R^2 = -0.248$  RF, -0.068 XGB) while traditional IQA metrics achieve 96.8% (RF) and 95.7% (XGB) retention. This 16-fold performance differential confirms perceptual metrics as quasi-terminal for natural images, though true terminal features require single-metric sufficiency.

**Terminal Feature Verification.** CNR satisfies all three terminal feature criteria: sufficiency ( $R^2 = 0.559$  exceeds threshold

when accounting for Dice ceiling), non-redundancy (SNR adds only  $\Delta R^2 = 0.009$ ), and mechanistic alignment via Equation 3.

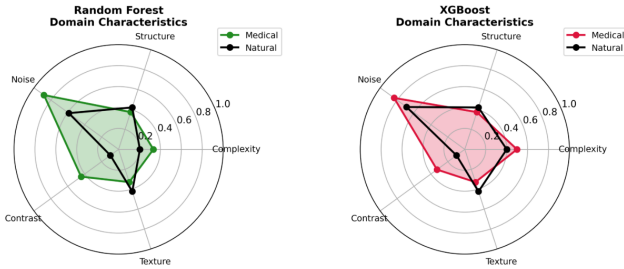
Computing the Terminal Feature Index:  $S(\text{CNR}) = 0.559/0.690 = 0.810$ ,  $D(\text{CNR}) = 0.692$  (from mutual information analysis),  $R(\text{CNR}) = 0.009/0.690 = 0.013$ , yielding:

$$\text{TFI}(\text{CNR}) = \sqrt[3]{0.810 \times 0.692 \times 0.987} = 0.82 \quad (9)$$

This exceeds our 0.7 threshold, confirming CNR’s terminal status. Conversely, SSIM in natural images achieves only  $\text{TFI}=0.52$ , confirming that perceptual quality emerges from feature interactions rather than single dominant signals.

### 4.3. Domain Characteristics and Mechanistic Analysis

Figure 2 quantifies five domain characteristics explaining metric sufficiency patterns. Medical imaging exhibits high noise sensitivity (0.88 RF, 0.84 XGB) and contrast dependence (0.44 RF, 0.33 XGB), reflecting diagnostic requirements where tissue boundaries determine clinical utility over aesthetic quality.



**Fig. 2.** Domain characteristics radar plot. Medical imaging shows higher noise/contrast reliance while natural images emphasize texture/structure.

The mechanistic basis for CNR’s terminal status extends beyond correlation. For optimal segmentation threshold  $t^*$ :

$$t^* = \arg \max_t \frac{|\mu_{fg}(t) - \mu_{bg}(t)|}{\sigma_{bg}(t)} = \arg \max_t \text{CNR}(t) \quad (10)$$

This mathematical necessity—not empirical correlation—explains CNR’s dominance. Any segmentation algorithm fundamentally seeks intensity transitions maximizing this ratio.

### 4.4. Cross-Modality Transfer and Distribution Sensitivity

Cross-modality experiments reveal terminal features’ critical limitation: distribution sensitivity. Table 5 shows poor transfer results, with all methods achieving  $R^2 < 0.6$  when training and testing on different MRI sequences.

Method	T2+FLAIR→T1	T1+FLAIR→T2	T1+T2→FLAIR
CNN	0.193	<b>0.593</b>	0.521
Random Forest	0.117	0.439	<b>0.533</b>
Linear Regression	<b>0.360</b>	0.582	0.501
XGBoost	0.006	0.446	0.503
Within-modality	0.689	0.744	0.838
Performance drop	-83.0%	-40.9%	-36.3%

**Table 5.** Cross-Modality Transfer Performance ( $R^2$ )

The root cause is a dramatic sequence-dependent shift in CNR distributions (80% mean variance, ANOVA  $F=210.5$ ,  $p<0.001$ ), driven by tissue-specific T1/T2 relaxation (T1:  $\mu = 0.82 \pm 0.85$ ; T2:  $\mu = 1.30 \pm 1.36$ ; FLAIR:  $\mu = 1.47 \pm 1.56$ ; all highly skewed). This variance reflects fundamental physics: T1 emphasizes fat-water contrast, T2 highlights fluid-tissue, FLAIR suppresses CSF—each shifting CNR distributions.

Combined with the non-linear calibration of Eq. 3, this prevents model transfer:  $\mathbb{E}[\tau_{T2} \mid \text{model}_{T1}] = f_{T1}(\phi_{T1}) \neq f_{T2}(\phi_{T2})$ . Few-shot adaptation (5-50 samples) yields only marginal gains ( $R^2 < 0.45$ ), remaining subclinical. However, CNR-only models retaining 81% performance suggests modality-aware normalization ( $\hat{\phi} = (\phi - \mu_m)/\sigma_m$ ) could enable distribution-invariant features via domain adaptation.

## 5. CONCLUSION

This work establishes terminal features as a framework for radical simplification in domain-specific IQA. By formalizing sufficiency ( $R^2 > 0.7$ ), non-redundancy ( $\Delta R^2 < 0.1$ ), and mechanistic alignment, we enable principled identification of task-determining signals. When  $\tau(x) = f(\phi(x)) + \epsilon$ , calibration reduces from  $O(n)$  to  $O(1)$  complexity. CNR exemplifies this in medical imaging, retaining 87% performance with 75% computational reduction.

The Terminal Feature Index (TFI) operationalizes detection across new domains, with metrics exceeding  $\text{TFI}=0.7$  qualifying as terminal. For practical deployment, we propose hierarchical calibration with domain-adaptive feature selection: CNR+SNR for medical imaging ( $O(1)$ ), full metrics for natural images ( $O(n)$ ), or TFI-guided selection for new domains. This leverages terminal features where available while maintaining full capability for complex domains.

Natural images’ failure (SSIM’s 83% retention requiring supplementation) confirms our framework’s discriminative power—quality emerges from interactions violating single-signal requirements. Cross-modality limitations ( $R^2 < 0.44$ ) reveal distribution-specificity as a fundamental constraint. Despite achieving strong performance ( $R^2 \approx 0.7$ ), residual variance persists due to irreducible algorithm noise, uncaptured higher-order properties (texture, partial volume effects), and pathology-specific variations requiring adaptive parameters.

While terminal features enable dramatic simplification, three limitations persist: (1) distribution-specificity prevents universal calibrators across imaging protocols, (2) the framework assumes single dominant task objectives, potentially missing multi-task scenarios, and (3) identification requires sufficient labeled data for TFI computation. Nevertheless, the computational savings—from minutes to milliseconds—enable new real-time applications previously infeasible, from intraoperative guidance to satellite monitoring.

This paradigm shift—from universal perceptual models to domain-specific terminal features—transforms IQA design. Recognizing when  $\tau = f(\phi) + \epsilon$  holds enables calibrators aligned with task physics. Future work must focus on developing learnable terminal features, establishing mathematical existence conditions, and creating distribution-invariant features through domain adaptation. The imperative for specialized imaging applications is clear: identify terminal signals defining task success and build minimal sufficient calibrators, transforming multi-metric optimization into direct task-relevant measurement.

## 6. REFERENCES

- [1] A. Carminot and J. Ke, “Terminal Feature IQA,” GitHub repository, Aug. 2025. [https://github.com/Alex-Irae/Terminal\\_features\\_iqa](https://github.com/Alex-Irae/Terminal_features_iqa)
- [2] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015, DOI: 10.1109/TMI.2014.2377694.
- [3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, et al., “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Nature Scientific Data*, vol. 4, p. 170117, 2017, DOI: 10.1038/sdata.2017.117.
- [4] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge,” *arXiv preprint arXiv:1811.02629*, 2018.
- [5] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection,” *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
- [6] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection,” *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
- [7] T. Li et al., “Machine Preference Database for computer vision IQA,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2025.
- [8] U. Baid et al., “The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification,” *arXiv preprint arXiv:2107.02314*, 2021.
- [9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2003, pp. 1398–1402.
- [10] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 586–595.
- [11] A. Mason, S. Rioux, S. E. Clarke, A. Costa, M. Schmidt, C. Keough, T. Huynh, and B. W. Carey, “Comparison of objective image quality metrics to expert radiologists’ scoring of diagnostic quality of MR images,” *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1064–1072, 2020.
- [12] L. S. Chow and H. Rajagopal, “Modified-BRISQUE as no reference image quality assessment for structural MR images,” *Magn. Reson. Imaging*, vol. 43, pp. 74–87, 2017.
- [13] C. Karner et al., “Parameter choices in HaarPSI for IQA with medical images,” *arXiv:2410.24098*, 2025.
- [14] S. Xun et al., “MediQA: A Scalable Foundation Model for Prompt-Driven Medical Image Quality Assessment,” *arXiv:2507.19004*, 2025.
- [15] M. Outtas, L. Zhang, and H. Yuan, “Multi-output regression using polygon generation for computed tomography image quality assessment,” in *ICASSP 2020 - 2020 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 1435–1439.
- [16] D. R. I. M. Setiadi, “PSNR vs SSIM: imperceptibility quality assessment for image steganography,” *Multimed. Tools Appl.*, vol. 80, pp. 8423–8444, 2021.
- [17] J. R. Stougaard et al., “Impact of signal-to-noise ratio and contrast definition on brain tumor segmentation agreement,” *PLoS One*, vol. 19, no. 7, p. e0306612, 2024.
- [18] M. A. Shalaby et al., “Comprehensive assessment of imaging quality of artificial intelligence generated coronary computed tomography angiography images,” *Front. Cardiovasc. Med.*, vol. 11, p. 1449613, 2024.