MIQANet: A Novel Dual-Branch Deep Learning Framework for MRI Image Quality Assessment

Yueran Ma, Huasheng Wang, Yingying Wu, Jean-Yves Tanguy, Richard White, Phillip Wardle, Elizabeth Krupinski, Padraig Corcoran, and Hantao Liu

***Abstract*—Image quality assessment (IQA) algorithms have significantly advanced over the past two decades, primarily focusing on natural images. However, applying these methods directly to medical imaging often yields suboptimal performance due to inherent differences such as the structural complexity of medical images and the limited availability of annotated databases. In this study, we conduct a comprehensive evaluation of state-of-the-art IQA methods, including 29 traditional full-reference (FR), 4 traditional no-reference (NR), and 9 deep learning-based approaches, to assess their effectiveness in the context of medical imaging. Our evaluation is performed on a recently developed MRI image quality assessment benchmark, revealing critical performance gaps in existing methods. Building on these findings, we propose a novel dual-branch deep learning framework specifically designed for medical IQA (MIQANet). The proposed approach effectively combines global contextual information with local structural details, enhancing the model's ability to capture subtle degradations and structural inconsistencies in MRI scans. Experiential results demonstrate the superiority of our approach over existing methods, providing valuable theoretical and practical insights for enhancing quality assessment of medical images.**

***Index Terms*—Image quality assessment, medical image, deep learning, artifacts, MRI.**

I. Introduction

Image quality assessment (IQA) plays a fundamental role in computer vision and image processing, encompassing both subjective and objective quality assessment methods. Subjective IQA involves human observers evaluating the perceived quality of images, typically quantified through the mean opinion score (MOS) [1]. Although subjective assessment is considered the most reliable method due to its alignment with human perception, it is inherently time-consuming, expensive, and impractical for real-time applications. To overcome these limitations, objective IQA methods are developed to automatically predict image quality in a manner that closely approximates human visual perception [2].

Over the past two decades, substantial progress has been made in objective IQA, resulting in the development of methods of three primary categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) IQA [3]. FR methods require full access to the original reference image to evaluate the quality of a distorted image, RR methods rely on partial reference features, and NR methods assess the quality of a distorted image without any reference.

Recently, researchers have recognised the limitations of applying general-purpose IQA models across diverse image domains. To address this, domain-specific datasets and metrics have been developed to better capture the unique degradation patterns and perceptual characteristics of specialised imaging scenarios. Representative examples include benchmark datasets and metrics for underwater image enhancement [4], [5], single image super-resolution [6], HDR tone mapping [7], and underwater IQA with degradation-aware models [8]. In parallel, perception-driven approaches that explicitly model human visual sensitivity have gained attention [9]. These efforts highlight the growing trend of developing domain-specific IQA frameworks, further motivating our investigation into medical image quality assessment.

In medical imaging, an increasing number of domain-specific IQA methods have been developed for individual modalities or quality degradation types, such as CT noise estimation, ultrasound speckle suppression, and endoscopic image clarity grading, as summarised in recent reviews in [10], [11]. These surveys highlight a growing body of research dedicated to medical IQA, ranging from traditional approaches based on handcrafted features or rule-based heuristics to more recent deep learning-based frameworks. Representative examples include CNN-based models for MRI quality classification [12], motion artifact detection [13], and multi-level artifact grading [14], as well as retinal fundus quality assessment using CNN-guided local features [15] and HVS-inspired hybrid networks [16]. While these studies demonstrate the rapid progress in modality-specific IQA, most existing methods focus on classification rather than perceptual quality prediction (e.g., MOS) and remain limited in scalability, generalisability, and benchmark availability. This indicates the need for a unified benchmark and a generalisable no-reference IQA framework specifically tailored for medical imaging – a domain that remains critically important yet comparatively underexplored.

Despite significant advances in IQA for natural images, applying these techniques to medical imaging remains challenging [17], [18]. Medical images differ fundamentally from natural images not only in their visual characteristics and content but also in how their quality is perceived [10], [11]. Natural images typically exhibit rich colours and diverse textures; degradations such as blur, noise, compression artifacts, or colour shifts primarily affect perceived aesthetics

Yueran Ma, Huasheng Wang, Yingying Wu, Padraig Corcoran and Hantao Liu are with the School of Computer Science and Informatics, Cardiff University, Cardiff, United Kingdom.

Jean-Yves Tanguy is with the Department of Radiology, Angers University Hospital, Angers, France.

Richard White is with the Department of Radiology, University Hospital of Wales, Cardiff, United Kingdom.

Phillip Wardle is with the National Imaging Academy Wale, United Kingdom.

Elizabeth Krupinski is with the Department of Radiology and Imaging Sciences, Emory University, Atlanta, United States.

Corresponding author: Huasheng Wang (e-mail: WangHS@cardiff.ac.uk)

and semantic clarity. In contrast, medical images, such as MRI scans, are often monochromatic, characterised by subtle grayscale variations, blurred anatomical boundaries, low-contrast structures, and motion artifacts. These degradations increase structural complexity and reduce visual clarity, but more importantly, they directly affect diagnostic utility [19]. In diagnostic imaging, "quality" extends beyond visual appeal to encompass clinical interpretability. Even subtle degradations can severely compromise diagnostic accuracy, particularly when they alter tissue contrast, boundaries, or small structures. For example, a slightly blurry photograph may remain understandable, whereas a similarly degraded MRI scan could obscure a small lesion or lead to an incorrect diagnosis. Consequently, natural image IQA methods, which rely on detecting prominent features such as edges and colours, may fail to capture the fine-grained texture variations and subtle anomalies critical in medical imaging. Furthermore, medical imaging modalities such as MRI, CT, and ultrasound introduce additional complexities due to variations in acquisition techniques, resolution, and noise characteristics [20]. These differences pose significant challenges for existing IQA methods, which are often tailored to statistical properties of natural images rather than the structural and anatomical features inherent in medical images. The direct application of IQA techniques developed for natural images to medical imaging often results in suboptimal performance. Therefore, there is a need to systematically evaluate existing IQA techniques in the context of medical imaging and develop tailored solutions to address its unique challenges.

In this paper, we focus on MRI image quality assessment. Full-reference (FR) approaches are commonly applied in compressed transmission or reconstruction quality evaluation, such as when assessing the impact of compression, denoising, or super-resolution algorithms on fidelity relative to the original data [21]. However, in most practical scenarios, there is typically no pristine reference image available due to factors such as patient motion, hardware limitations, noise, or other acquisition related artifacts. As a result, quality assessment methods must rely solely on the acquired data, making no-reference (NR) approaches the standard choice in clinical imaging [11]. The contributions of this study are as follows:

1) **Comprehensive Evaluation of IQA Methods for Medical Imaging:** We conduct a thorough evaluation of 29 traditional FR methods, 4 NR methods, and 9 deep learning-based methods, systematically analysing their performance and applicability for medical imaging applications.
2) **Development of a Novel Deep Learning Framework for Medical IQA:** Based on our evaluation results, we propose a novel deep learning framework specifically designed to optimise medical image quality assessment by addressing domain-specific challenges.
3) **Extensive Experimental Validation:** We perform rigorous experiments to validate the effectiveness and reliability of the proposed IQA method against existing methods, and ablation studies to assess the impact of different architectural components in the proposed framework.

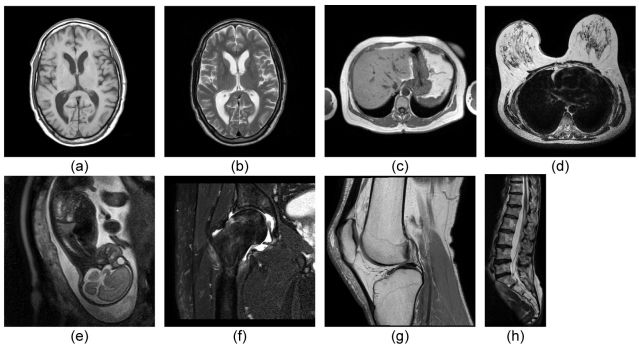


Fig. 1. Reference MRI images used in our RAD-IQMRI database. The images are referred to (a) Brain_T1, (b) Brain_T2, (c) Liver, (d) Breast, (e) Fetus, (f) Hip, (g) Knee, and (h) Spine.

II. Materials and Evaluation Framework

A. Database

The RAD-IQMRI database [22] was developed to tackle the challenges associated with MRI image quality assessment, with a focus on the presence of visual artifacts that can impact diagnostic accuracy. The dataset comprises eight original high-quality MRI images, representing various anatomical regions, including the brain, liver, breast, fetus, hip, knee and spine, acquired using a Philips Achieva 1.5T MRI system. Figure 1 illustrates representative samples of these source images. To generate a diverse set of distorted images, common MRI artifacts were systematically simulated and introduced to the source images, including ghosting and noise, across multiple severity/energy levels. These simulated artifacts closely reflect real-world challenges frequently encountered in clinical MRI scans, ensuring the dataset's relevance for evaluating IQA algorithms. A total of 112 distorted MRI images were generated. A fully controlled perception experiment was conducted to obtain reliable image quality ratings for the MRI images contained in the RAD-IQMRI database. Thirteen radiologists participated in the psychovisual test, rating image quality using a standard simultaneous double-stimulus IQA method [X] as the interface illustrated in Figure 2. The radiologists' ratings were then used to compute the mean opinion score (MOS) for each image, establishing a benchmark for evaluating the performance of IQA models on MRI images.

B. Implementation Details

All experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 3060 GPU, using PyTorch 1.10.0 and CUDA 11.7. For traditional IQA models, publicly available implementations were used directly when available; otherwise, the methods were implemented from scratch. Following conventional practice, the performance is measured by applying an IQA model to the entire database. For deep learning-based IQA models, due to the limited size of the RAD-IQMRI dataset [22], we adopted a k-fold (k=8 in our study) cross-validation approach for critical evaluation. The dataset was divided into eight non-overlapping partitions, where in each iteration, seven partitions were used for training

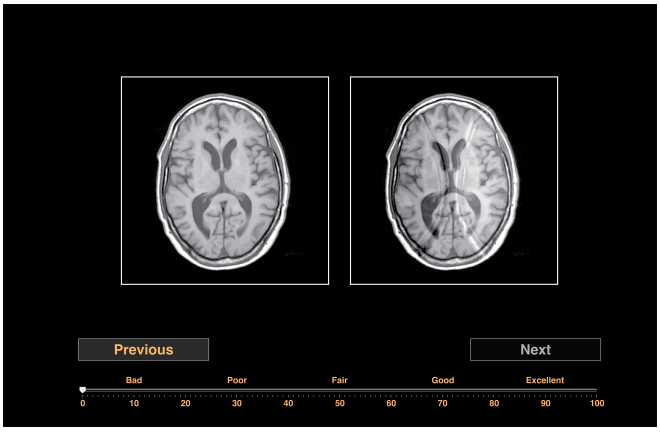


Fig. 2. Illustration of the scoring interface used in the psychovisual experiment. The interface presents two stimuli side-by-side, with the reference image on the left and the test image on the right.

and the remaining partition was used for testing. This process was repeated eight times, ensuring each fold serves as the test set once, and the final performance was averaged across all test results.

For all deep learning-based models, the backbone networks were initialized with ImageNet pre-trained parameters [23]. Hyperparameters such as learning rates, batch sizes, and optimizer settings were adjusted based on preliminary experiments to achieve optimal performance during training. Additionally, we employed learning rate scheduling to dynamically adjust the learning rate during training, which facilitated better model convergence.

C. Evaluation Criteria

The performance of IQA models is evaluated using two metrics, i.e., Pearson Linear Correlation Coefficient (PLCC) and Spearman's Rank-order Correlation Coefficient (SROCC). These metrics are widely used to measure the accuracy and consistency of model's predictive ability in relation to subjective judgments.

Since objective IQA scores often exhibit a non-linear relationship with subjective ratings, a non-linear regression, i.e., a five-parameter modified logistic function [X] is often applied prior to computing PLCC. The PLCC is calculated as follows:

$$\text{PLCC} = \frac{\sum_{i=1}^{N}(s_i - \bar{s})(\hat{s}_i - \bar{\hat{s}})}{\sqrt{\sum_{i=1}^{N}(s_i - \bar{s})^2}\sqrt{\sum_{i=1}^{N}(\hat{s}_i - \bar{\hat{s}})^2}} \tag{1}$$

where s_i and $\hat{s}_i$ denote the subjective and predicted quality scores for the i-th image, and $\bar{s}$ and $\bar{\hat{s}}$ are their corresponding means. N is the total number of images in the evaluation set.

The SROCC metric assesses the monotonic relationship between the rankings of predicted and subjective scores, which is calculated as:

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{2}$$

where d_i represents the difference between the ranks assigned to the i-th image in the predicted and subjective scores.

Both PLCC and SROCC take absolute values in the range [0, 1], where values closer to 1 indicate a stronger correlation with subjective assessments, reflecting more accurate and reliable predictions of an IQA model.

III. Comparative Analysis of IQA Algorithms for Medical Imaging

IQA methods have been extensively developed for natural images; however, their applicability to medical imaging remains under-explored. This study systematically evaluates the effectiveness of established IQA methods from the natural image domain in medical imaging. We assess 29 traditional FR methods, 4 traditional NR methods, and 9 deep learning-based methods, using open-source implementations where available and reproducing others (based on the original publications) as needed. Evaluation is conducted on the RAD-IQMRI database. Table I summarises the IQA methods.

A. IQA Methods

1) Traditional FR IQA: FR IQA based on "Pixel Statistics": The simplest method is to calculate the pixel differences between the reference and distorted images and then fuse these differences to an image quality score. The most common methods are the Mean Square Error (MSE) and Signal-to-Noise Ratio (SNR)/Peak Signal-to-Noise Ratio (PSNR) [X]. These metrics are of low computational complexity; however, they do not consider the properties of the human visual system (HVS) [51], and therefore, they often do not correlate with subjective IQA. Many methods have been proposed to improve the MSE/PSNR metrics, integrating vision models such as the contrast sensitivity function (CSF). The improved models include WSNR [24], NQM [24], PSNR-HVS [25], PSNR-HVS-M [26], PSNR-HA [27], PSNR-HMA [27], IW-PSNR [28]. Some methods are calculated in the discrete wavelet domain, including PSNR-DWT and AD-DWT [29]. *FR IQA based on "Structural Information"*: Based on the observation that the HVS has the ability to extract image structure information, the Structural Similarity (SSIM) index is proposed in [21], which is a modified version of the Universal image Quality Index (UQI). A series of improved models of SSIM have been proposed, including MSSSIM [30], IW-SSIM [28], WSSI [31], SSIM-DWT [29]. In addition, some models are built based on the structural similarity framework, including RFSIM [32], FSIM [33], SR-SIM [34]. In GMSD [35], a gradient similarity method is proposed. In MCSD [36], a multiscale contrast similarity method is proposed. In DSS [37], a DCT subbands similarity method is proposed. *FR IQA based on "Information Theory"*: Information theory-based methods quantify mutual information between the reference and distorted images, including IFC [38], VIF [39], VIFP [39], DWT-VIF [40], and VIF-DWT [29]. *FR IQA based on "Mixed Strategy"*: Some mixed strategy-based methods have been proposed. For example, MAD [41] considers that HVS has a different emphasis under different conditions of image quality and applies different modelling strategies for image of high and low quality.

TABLE I
SUMMARY OF IMAGE QUALITY ASSESSMENT (IQA) CATEGORIES AND METHODS EVALUATED IN THIS STUDY

Category	Method
Traditional FR IQA	**Pixel Statistics** - SNR, PSNR, WSNR [24], NQM [24], PSNR-HVS [25], PSNR-HVS-M [26], PSNR-HA [27], PSNR-HMA [27], IW-PSNR [28], PSNR-DWT [29], AD-DWT [29]
	Structural Information - UQI, SSIM [21], MSSSIM [30], IW-SSIM [28], WSSI [31], SSIM-DWT [29], RFSIM [32], FSIM [33], SR-SIM [34], GMSD [35], MCSD [36], DSS [37]
	Information Theory - IFC [38], VIF [39], VIFP [39], DWT-VIF [40], VIF-DWT [29]
	Mixed Strategy - MAD [41]
Traditional NR IQA	BRISQUE [42], NIQE [43], ILNIQE [44], dipIQ [45]
Deep Learning-Based NR IQA	SimpleCNN [46], DBCNN [47], HyperIQA [48], CLIP-IQA [49], MANIQA [50]

2) Traditional NR IQA: Traditional NR IQA methods evaluate the quality of distorted images by analysing natural scene statistics (NSS) or using hand-crafted features relevant to visual distortions, including BRISQUE [42], NIQE [43], ILNIQE [44]. The Digital Image Pairs Inferred Quality Index (dipIQ) [45] method uses features from FR IQA methods and a learning-to-rank approach to create the quality index.

3) Deep Learning-Based NR IQA: The deep learning-based NR IQA methods exploit the convolutional neural networks (CNNs) to directly extract discriminative features from images, demonstrating superior performance over traditional approaches [46], [52]–[54]. A CNN-based metric (SimpleCNN) is proposed in [46] that integrates feature learning and regression into a unified end-to-end optimisation process. Additionally, the backbone network can be substituted with architectures such as LeNet [55], VGG19 [56], ResNet18 [57], or ResNet34 [57] to develop alternative NR-IQA methods. Other popular deep learning-based approaches include DBCNN [47]: a deep bilinear CNN model designed to capture both local and global image quality features; HyperIQA [48]: a model that adapts to varying perceptual rules by leveraging a self-adaptive hyper network; CLIP-IQA [49]: a model that utilises the visual-language understanding capabilities of CLIP to assess perceived quality of images, and MANIQA [50]: A multi-dimension attention network that combines Vision Transformer features, transposed attention blocks, and scale swin transformer blocks.

B. Comparative Analysis

1) Comparison of Different Model Genres: Table II outlines the performance (measured by PLCC and SROCC) of existing IQA methods as mentioned above, including traditional full-reference (T-FR), traditional no-reference (T-NR), deep learning with no-finetuning (DL-NFT), and deep learning with finetuning (DL-FT) IQA. The results reveal that state-of-the-art (SOTA) T-FR IQA models can achieve strong performance by directly applying to medical imaging, e.g. IWSSIM, MCSD, GMSD, and MAD produce PLCC/SROCC values exceeding 0.85. This can be attributed to their reliance on a reference. However, FR IQA methods are impractical in medical imaging, as reference images are often unavailable in many clinical scenarios.

T-NR models, on the other hand, perform poorly, suggesting that traditional no-reference approaches do not transfer well from natural images to medical imaging. Deep learning-based IQA models show more promising results, particularly when finetuned on medical image data. Notably, most existing IQA models are typically pre-trained on natural images, which differ significantly from medical images in terms of strucure and noise characteristics. Without finetuning, these models fail to capture domain-specific features, such as anatomical structures and modality-specific noise patterns; finetuning on medical image data allows the model to adapt to these unique features, leading to improved performance. In particular, the SOTA deep learning-based IQA model, MANIQA, which is the top-performing model for natural images, achieves the best results on medical imaging after finetuning. Therefore, the following of the paper will focus on the deep learning-based NR IQA.

2) Comparison of Different Transfer Learning Methods: Transfer learning has become increasingly important in medical imaging due to the limited availability of large-scale annotated datasets - a common challenge in this domain. Leveraging knowledge from models pre-trained on large natural image datasets can improve performance on medical imaging tasks where labeled data is scarce. In this study, we adopt a three-step strategy to assess the impact of transfer learning on IQA in medical imaging. **Train_LIVE** *Training on natural IQA dataset without finetuning*: Models are initially trained on the LIVE database [X] (widely used natural IQA database) without finetuning and directly tested on the RAD-IQMRI database for medical imaging. **Train_ImageNet_LIVE** *Pre-training on ImageNet and finetuning on LIVE*: Models are pre-trained on ImageNet, a large-scale natural image classification dataset, and then finetuned on LIVE before being tested RAD-IQMRI. **Finetune_RADIQMRI** *Pre-training on ImageNet, finetuning on LIVE, and further finetuning on RAD-IQMRI*: Models are pre-trained on ImageNet, and finetuned on LIVE, and subsequently further finetuned on RAD-IQMRI to evaluate their performance.

Figure 3 illustrate the performance (measured by PLCC) of different deep learning-based IQA models under each training strategy. The **Train_LIVE** strategy result in the lowest PLCC values across all models. The **Train_ImageNet_LIVE** strategy significantly improves model performance, yielding higher PLCC values in all methods compared to **Train_LIVE**. Notably, the **Finetune_RADIQMRI** strategy lead to further improvements. Particularly for transformer-based approaches i.e., ClipIQA and MANIQA, the models significantly benefit from finetuning on medical data. For CNN-

TABLE II
PERFORMANCE COMPARISON OF IMAGE QUALITY ASSESSMENT (IQA) METHODS

Type	Methods	PLCC	SROCC	Type	Methods	PLCC	SROCC
	UQI	0.1835	0.2247		dipIQ	0.0231	0.0484
	AD_DWT	0.2796	0.1912	**T-NR**	BRISQUE	0.2694	0.0879
	PSNR	0.4999	0.4687		IL-NIQE	0.3021	0.2819
	SNR	0.4999	0.4687		NIQE	0.4305	0.2154
	PSNR_HVSM	0.5052	0.4626				
	PSNR_DWT	0.5119	0.5363				
	PSNR_HVS	0.6307	0.5610		ClipIQA (no finetuning)	0.1068	0.0939
	IW_PSNR	0.6611	0.6665		HyperIQA (no finetuning)	0.2040	0.1890
	IFC	0.7033	0.7060		SimpleCNN (no finetuning)	0.2253	0.3148
	PSNR_HMA	0.7051	0.6819		MANIQA (no finetuning)	0.3163	0.3208
	VIFP	0.7201	0.7264	**DL-NFT**	DBCNN (no finetuning)	0.3556	0.2392
	DWT_VIF	0.7716	0.7764		LeNet5 (no finetuning)	0.4496	0.3478
	RFSIM	0.7743	0.7643		ResNet34 (no finetuning)	0.5224	0.3280
T-FR	VIF	0.7908	0.7857		ResNet18 (no finetuning)	0.5325	0.4088
	PSNR_HA	0.7935	0.7582		VGG19 (no finetuning)	0.5758	0.4544
	SRSIM	0.7951	0.8676				
	VIF_DWT	0.7995	0.7907		SimpleCNN	0.5979	0.5494
	SSIM_DWT	0.8015	0.8346		LeNet5	0.6181	0.6207
	SSIM	0.8089	0.8648		HyperIQA	0.6233	0.6455
	WSSI	0.8122	0.8319		DBCNN	0.6289	0.6609
	FSIM	0.8321	0.8797	**DL-FT**	ClipIQA	0.6501	0.6870
	DSS	0.8361	0.8527		VGG19	0.6909	0.6820
	MSSSIM	0.8465	0.8659		ResNet18	0.6923	0.6513
	IWSSIM	0.8613	0.8896		ResNet34	0.7048	0.7549
	MCSD	0.8739	0.8769		MANIQA	0.9147	0.9012
	GMSD	0.8838	0.8731				
	MAD	0.8863	0.8808				

Note: T-FR represents Traditional Full-Reference methods, T-NR represents Traditional No-Reference methods, DL-NFT represents Deep Learning-based methods without finetuning, and DL-FT represents Deep Learning-based methods with finetuning.

based methods like ResNet34, ResNet18, and VGG19, both **Train_ImageNet_LIVE** and **Finetune_RADIQMRI** produce comparable results, though **Finetune_RADIQMRI** consistently achieves the best overall performance.

These findings suggest that transfer learning, especially finetuning on task-specific medical datasets like RAD-IQMRI, is highly effective in improving IQA performance in medical imaging. This is particularly evident in transformer-based models, where finetuning provides significantly improved generalisation to medical images.

IV. PROPOSED METHOD

In natural image processing, CNNs and Transformers have demonstrated strong capabilities, each with distinct advantages and limitations. CNNs excel in capturing local features through their hierarchical spatial hierarchies, making them particularly effective for analysing fine details in medical images. However, their reliance on local receptive fields can limit their ability to model global image context. In contrast, Transformers leverage self-attention mechanisms to capture long-range dependencies, providing a broader perspective on image content and enhancing contextual understanding.

In natural image IQA, both local and global features contribute to perceived quality, but their influence on overall perception is relatively less content- or task-sensitive [9]. In contrast, medical image (e.g., MRI) quality perception is strongly shaped by diagnostic tasks and clinically relevant content, making the balance between local detail and global anatomical structure particularly critical [58], [59]. For instance, subtle local degradations can obscure fine details such as lesions or vessel boundaries, while global degradations may compromise the overall visibility of anatomical context and organ morphology [60]. Both aspects jointly determine whether an image is clinically useful, as radiologists rely on fine local details for detection tasks and on global coherence for orientation and diagnosis [61]. Our model design is therefore motivated by the need to explicitly capture this dual nature of medical image quality.

To address these MRI-specific challenges, we propose a dual-branch model for medical image IQA (MIQANet), as shown in Figure 4. The design reflects the perceptual needs of radiologists, who rely on both local and global cues when assessing MRI quality. The local branch focuses on extracting fine-grained features that are sensitive to subtle degradations in diagnostically critical regions, while the global branch captures holistic structural information and contextual consistency across the entire image. By combining these complementary representations, MIQANet provides a more comprehensive

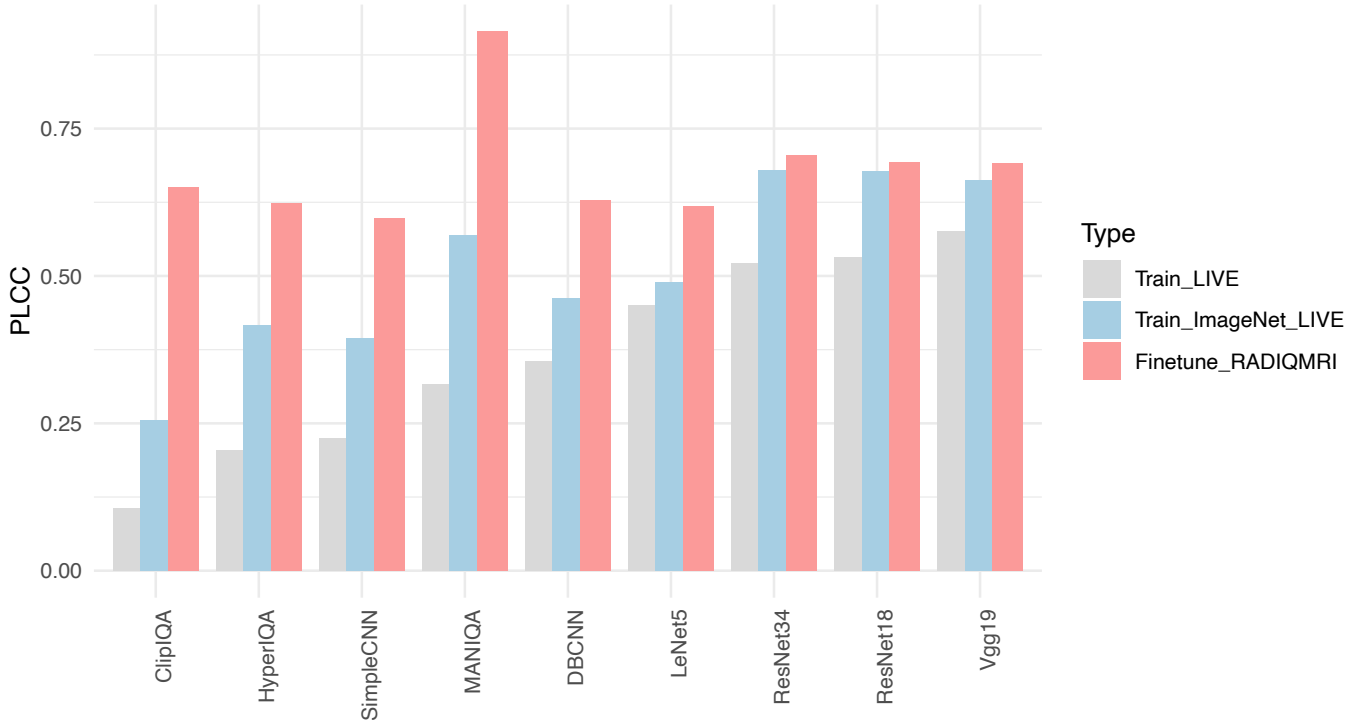


Fig. 3. Performance comparison of deep learning-based IQA models based three different training strategies. Train_LIVE: Trained on LIVE and tested on RAD-IQMRI; Train_ImageNet_LIVE: Pretrained on ImageNet, finetuned on LIVE and tested on RAD-IQMRI; Finetune_RADIQMRI: Pretrained on ImageNet, finetuned on LIVE and RAD-IQMRI, then tested.

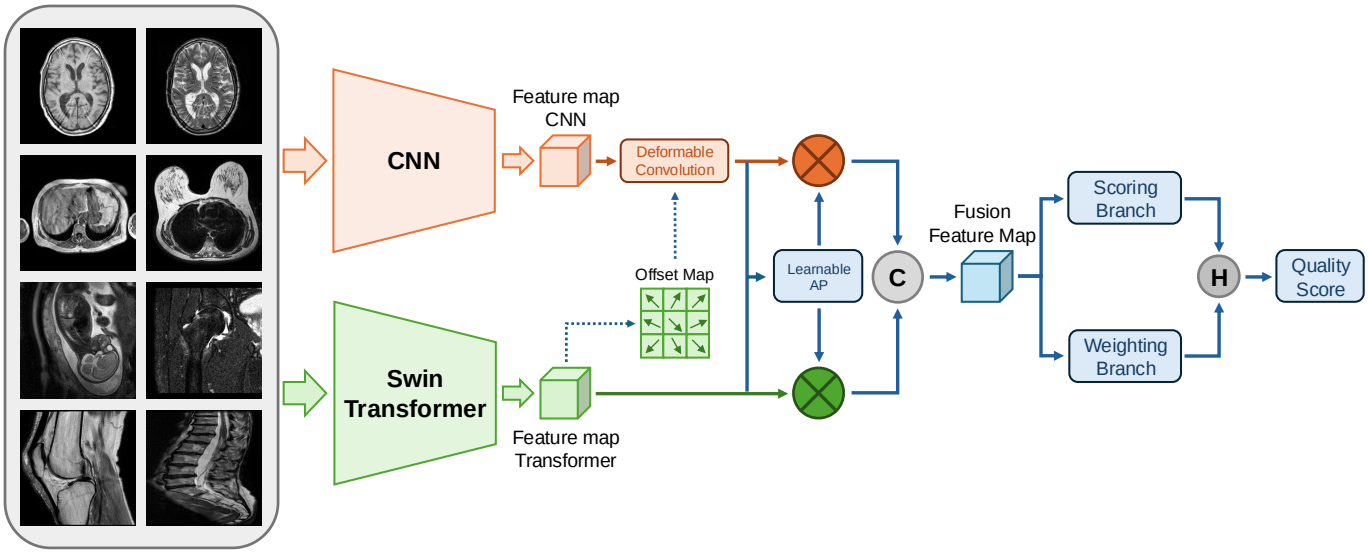


Fig. 4. Architecture of the proposed approach: The model integrates a CNN branch to capture local features and a Swin Transformer branch for global context analysis. A learnable Adaptive Parameter (AP) dynamically adjusts the contributions of each branch to optimise image quality assessment. In the diagram, 'C' represents concatenation operations, while 'H' denotes Hadamard product.

assessment of medical image quality than either branch could achieve independently. This architecture explicitly integrates the fine-grained local analysis capabilities of CNNs with the global context modelling efficiency of Swin Transformers [62]. Notably, the Swin Transformer's hierarchical design facilitates scalable and computationally efficient processing of high-resolution medical images, while its shifted windowing scheme further enhances global information integration without incurring the computational overhead of standard Transformers. Our proposed method is specifically designed for no-reference image quality assessment (NR-IQA), as reference-quality MRI images are generally unavailable in practical clinical environments.

The architecture of combining CNN and Transformer has

been widely explored in radiological image analysis, as thoroughly reviewed in [63]. Some recently proposed dual-branch architectures such as TransUNet [64] and SwinUNet [65] have demonstrated the consistent performance improvements over purely CNN- or Transformer-based baselines. However, most of these studies have focused on classification and segmentation tasks. To the best of our knowledge, no prior work has explicitly applied a CNN-Transformer hybrid architecture to medical IQA. Inspired by this design philosophy, our proposed MIQANet adopts a similar dual-branch strategy, but tailored for the distinct challenges of no-reference medical IQA, including several novel components: (1) adaptive deformable convolution module for aligning local features with the global attention context; (2) learnable weighting parameter to dynamically balance the contribution of CNN and Transformer branches; and (3) dual-branch decoder with patch-wise scoring and spatial attention fusion to provide perceptually consistent quality predictions.

A. CNN-Based Branch

The CNN-based branch of our proposed MIQANet model features a flexible backbone, enabling compatibility with various architectures to enhance adaptability across different IQA tasks. For our analysis, we have utilised models including simpleCNN, VGG, ResNet, and LeNet to demonstrate the branch's versatility and capability. Additionally, this backbone can be easily replaced with alternative CNN architectures, allowing for continuous adaptation and optimisation across a wide range of imaging conditions and assessment requirements.

B. Transformer-Based Branch

The Transformer-based branch of our proposed MIQANet model is structured for IQA using a hierarchical approach that processes images at multiple scales, known as the Swin Transformer [62]. The Swin Transformer operates by dividing the input image into non-overlapping local windows, within which self-attention is computed, enabling efficient and scalable feature extraction.

1) Local Window Attention: The input image is divided into fixed-size patches, forming non-overlapping local windows. Self-attention is then computed independently within each window, enabling the model to capture fine-grained local features. When computing self-attention within a local window, the standard self-attention mechanism can be represented by the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors.

2) Shifted Window Mechanism: The windows are shifted at each subsequent layer. This shift allows for cross-window connections, ensuring that information flows between neighboring patches without the need for global self-attention across the entire image:

$$\text{Shifted Window}(i, j) = W_{i,j} \rightarrow W_{i+\Delta i, j+\Delta j} \tag{4}$$

where $W_{i,j}$ represents the original window position at indices (i, j), and Δi and Δj are the shifts applied to the window in the horizontal and vertical directions, respectively.

3) Hierarchical Representation: The Swin Transformer constructs a hierarchical representation of the image by progressively reducing the resolution at each stage. The reduction is achieved through patch merging operations, which combine patches from the previous stage to form a coarser representation:

$$P_{m+1} = \text{Merge}(P_m) \tag{5}$$

where P_m represents the patches at level m, and $\text{Merge}(\cdot)$ denotes the operation that combines multiple patches into a single one, effectively reducing the resolution and increasing the receptive field.

C. Feature Fusion

The Transformer-based branch is employed to learn an offset map, which is used to apply deformable convolution on the CNN feature maps. This allows the CNN's local features to be adjusted and aligned with the global context provided by the Transformer. Additionally, we introduce an adaptive parameter (AP) that determines the relative contribution of each branch to the overall prediction task, ensuring that the combined output reflects the optimal balance between the two branches.

D. Patch-wise Prediction

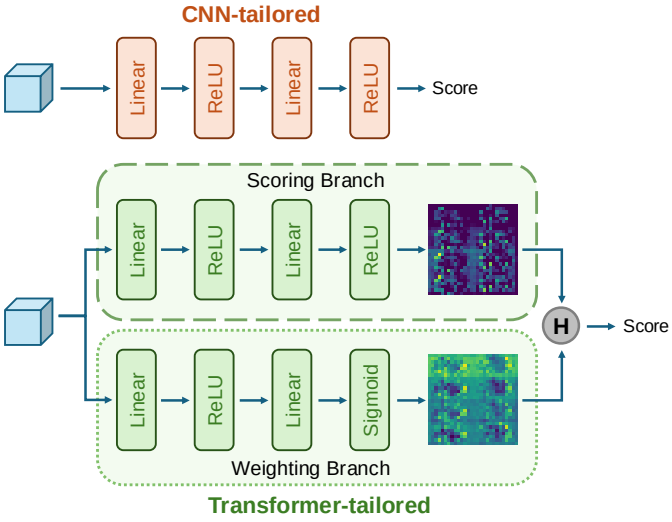


Fig. 5. Dual-branch structure of the patch-wise prediction module with CNN-tailored and Transformer-tailored decoding pathways.

The Patch-wise Prediction module is designed to overcome the limitations of traditional spatial pooling methods, which often result in the loss of critical information by neglecting the relationships between different image patches. Instead of relying on max-pooling or average-pooling to produce a single IQA score, this module utilises a dual-branch structure consisting of a scoring branch and a spatial attention branch.

In this module, each pixel in the deep feature map - corresponding to a specific patch of the input image - is individually

assessed. The scoring branch computes a quality score for each pixel, while the spatial attention branch generates an attention map that reflects the significance of each corresponding score. These two outputs are combined through a weighted summation, where the attention map serves as the weight for the patch scores. This weighted sum operation allows the model to account for the varying importance of different regions, simulating the human visual system's selective focus on areas of interest.

To enhance the adaptability and efficacy of the IQA decoder across different backbone encoder architectures, our MIQANet model features two specialised decoding pathways: one tailored for CNNs and another for Transformers, as illustrated in Figure 5. Each pathway is designed to optimise the interpretation of features according to the specific characteristics of the encoder used, ensuring the patch-wise perception of image quality is maximally accurate. The final predicted IQA score s_f for the entire image is calculated as:

$$s_f = \frac{\mathbf{s} * \mathbf{w}}{\sum \mathbf{w}} \tag{6}$$

where $\mathbf{s}$ represents the score map, $\mathbf{w}$ denotes the corresponding attention map, and $*$ signifies the Hadamard product.

To further refine the predictions, the module generates both weight and score projections from the feature map $\tilde{F}$ through two independent linear transformations. The final IQA score is derived by multiplying the patch scores with their corresponding weights and summing across all patches. This design introduces a balancing mechanism between the scoring and weighting branches, mitigating overfitting and ensuring that the final score reflects a balanced consideration of both patch scores and their relative importance.

E. Performance Evaluation

TABLE III
PERFORMANCE COMPARISON AND STATISTICAL SIGNIFICANCE RESULTS

Method	PLCC	SROCC	Significance (vs MIQANet)
SimpleCNN	0.5979	0.5494	1-1
LeNet5	0.6181	0.6207	1-1
HyperIQA	0.6233	0.6455	1-1
DBCNN	0.6289	0.6609	1-1
ClipIQA	0.6501	0.6870	1-1
VGG19	0.6909	0.6820	1-1
ResNet18	0.6923	0.6513	1-1
ResNet34	0.7048	0.7549	1-1
MANIQA	0.9147	0.9012	1-1
MIQANet (Ours)	**0.9320**	**0.9289**	-

Note: Statistical significance testing consists of two methods: (1) hypothesis testing following ITU-T Rec. P.1401, which performs a pairwise comparison of RMSE values between each SOTA model and our MIQANet model; and (2) hypothesis testing based on residuals between each SOTA model and MIQANet. The test results indicate whether the performance difference is statistically significant, where "1" represents significance and "0" represents non-significance.

1) Overall Performance and Statistical Significance: Table III illustrates the performance of our proposed MIQANet model compared to SOTA learning-based NR IQA models. To ensure a fair comparison, all models have undergone the same training strategy (i.e., the "Finetune_RADIQMRI" method) and k-fold cross-validation as detailed in section III. The results show that that MIQANet outperforms SOTA models, achieving a Pearson PLCC value of 0.9320 and a SROCC value of 0.9289.

To determine whether the observed performance improvement is statistically significant, we conduct hypothesis testing. However, it should be noted that statistical evaluation is inherently challenging, as conclusions depend on assumptions made about data distribution and the choice of statistical methods. To enhance the robustness of our findings, we adopt two well-established significance testing approaches widely used in the IQA literature: *P1401 Hypothesis Testing (HT_P1401)*: Assuming a normal distribution of the compared data, we follow the methodology outlined in ITU-T Rec. P.1401 [66]. This test is performed through pairwise comparison of RMSE values between MIQANet and each SOTA model. *Residual-Based Hypothesis Testing (HT_Res)*: HT_Res [67] evaluates residuals between each SOTA model and MIQANet. We first assess the normality of the residual distributions (i.e., $|MOS - \text{NR IQA}|$ and $|MOS - \text{MIQANet}|$). If the residuals are normally distributed, an independent samples t-test is performed; otherwise, we employ the non-parametric Mann-Whitney U test. These complementary methods help ensure the reliability of our statistical analysis. As shown in Table III, both HT_P1401 and HT_Res confirm that the performance improvements achieved by MIQANet are statistically significant compared to all SOTA models.

TABLE IV
ABLATION STUDY ASSESSING THREE KEY COMPONENTS: THE CNN BRANCH WITH INTERCHANGEABLE BACKBONE NETWORKS, THE SWIN TRANSFORMER BRANCH, AND THE ADAPTIVE PARAMETER (AP). PERFORMANCE OF DIFFERENT MODEL CONFIGURATIONS, WHERE ✓ INDICATES THAT THE COMPONENT WAS USED IN THE MODEL.

CNN						
simpleCNN	VGG19	ResNet34	Swin-T	AP	PLCC	SROCC
✓	–	–	–	–	0.5979	0.5494
–	✓	–	–	–	0.6909	0.6820
–	–	✓	–	–	0.7048	0.7549
–	–	–	✓	–	0.9014	0.8885
✓	–	–	✓	–	0.8794	0.8964
–	✓	–	✓	–	0.9056	0.9178
–	–	✓	✓	–	0.9166	0.9178
✓	–	–	✓	✓	**0.9208**	**0.9294**
–	✓	–	✓	✓	**0.9233**	**0.9206**
–	–	✓	✓	✓	**0.9320**	**0.9289**

2) Ablation Study: To evaluate the effectiveness of the proposed MIQANet model, we conducted an ablation study assessing three key components: the CNN branch with interchangeable backbone networks, the Swin Transformer branch, and the adaptive parameter (AP) that learns the contribution of each branch. Table IV shows the performance of different model configurations measured by the PLCC and SROCC.

First, we evaluate three CNN backbones - SimpleCNN, VGG19, and ResNet34 - individually, without the Swin Transformer branch or the adaptive parameter (AP). As shown in Table IV, SimpleCNN achieves a PLCC of 0.5979, VGG19 improves upon this with a PLCC of 0.6909, and ResNet34 out-

performs both with a PLCC of 0.7048. These results suggest that deeper network architectures like ResNet34 extract more relevant features for IQA tasks compared to shallower/simpler architectures such as SimpleCNN.

Next, integrating the Swin Transformer branch yields substantial performance gains. When combined with ResNet34, the PLCC increases to 0.9166, demonstrating the benefits of using both local features from the CNN branch and global contextual information captured by the Swin Transformer. Similar improvements are observed when combining Swin Transformer with SimpleCNN and VGG19, resulting in PLCC values rising to 0.8794 and 0.9056, respectively.

Finally, incorporating the adaptive parameter (AP) further enhances performance by dynamically learning the optimal weight distribution between the CNN and Swin Transformer branches. The best-performing model, ResNet34 + Swin-T + AP, achieves a PLCC of 0.9320, demonstrating that adaptive weighting allows the model to better leverage the strengths of both branches. Similarly, VGG19 + Swin-T + AP and SimpleCNN + Swin-T + AP achieves PLCC values of 0.9233 and 0.9208, respectively, confirming the overall benefit of incorporating adaptive parameter (AP) in refining the model's predictions.

This ablation study clearly shows the complementary contributions of each component - CNN backbone, Swin Transformer, and adaptive parameter - with the highest IQA performance achieved through their integration.

3) Complexity Analysis and Real-Time Feasibility: We evaluate the training efficiency and parameter budget of MIQANet to determine whether its superior performance results from architectural design rather than increased model complexity. Table V repots the training time per epoch and parameter count of MIQANet compared with MANIQA, the best-performing baseline. The two models exhibit comparable computational costs, indicating that MIQANet achieves higher accuracy without sacrificing efficiency.

To further isolate the effect of architectural design, we construct two single-branch model variants with comparable parameter sizes: a CNN-only model (ResNet152 backbone) and a Transformer-only model (Swin-L backbone). Both variants are trained under identical settings on the RAD-IQMRI dataset. As shown in Table VI, these model variants achieve lower performance than MIQANet, despite having similar parameter budgets. These results demonstrate that the performance gain of MIQANet primarily arises from the proposed dual-branch architecture with adaptive weighting, rather than from higher model capacity (parameter count).

In terms of practical feasibility, MIQANet's inference time is on the order of tens of milliseconds per MRI slice on a modern GPU, which is well within real-time requirements for clinical quality monitoring scenarios. Furthermore, as the method is fully feedforward at inference time, it can be further optimized for deployment using standard acceleration frameworks such as TensorRT or ONNX, ensuring efficient integration into practical clinical workflows.

4) Discussion: The generalisation capability of MIQANet is currently constrained by the limited availability of annotated medical IQA datasets. Most existing studies rely on private data, and to the best of our knowledge, none of the medical IQA databases reported in literature are fully public. Nevertheless, we acknowledge that validation on an independent dataset is essential and plan to evaluate MIQANet on other emerging datasets [10], [11], once they become publicly accessible. To mitigate potential overfitting under limited data conditions, we adopted a multi-stage transfer learning strategy – pre-training on ImageNet, fine-tuning on LIVE, and further fine-tuning on RAD-IQMRI (as detailed in Section III) – combined with k-fold cross-validation to ensure consistent evaluation across disjoint subsets. These procedures enhance model robustness and promote reliable generalisation despite the modest dataset size. Finally, our RAD-IQMRI dataset is planned for public release and will be expanded into a larger-scale resource to support future validation and benchmarking studies.

TABLE V
COMPARISON OF TRAINING EFFICIENCY AND MODEL SIZE.

Model	Time / epoch	Param count
MANIQA	34 mins	135.62M
MIQANet (ours)	35 mins	137.65M

TABLE VI
COMPARISON OF MODEL PERFORMANCE UNDER MATCHED PARAMETER BUDGETS (≈137M PARAMETERS).

Model	PLCC ↑	Time / epoch	Param count
CNN-only (ResNet152)	0.7332	26 mins	137.28M
Transformer-only (Swin-L)	0.9188	38 mins	137.51M
MIQANet (ours)	**0.9320**	35 mins	137.65M

V. CONCLUSION

In this study, we systematically evaluated a broad spectrum of image quality assessment (IQA) techniques within the context of medical imaging, with a particular focus on MRI data. Our analysis revealed the limitations of existing IQA models when applied to medical images, and the need for domain-specific approaches. To bridge this gap, we proposed MIQANet, a novel deep learning-based model tailored for medical IQA, which demonstrated superior performance and potential as a new benchmark in the field.

These findings emphasise the importance of developing specialised IQA techniques that account for the unique characteristics of medical images. Future research involves refining our proposed MIQANet model, extending its application to other medical imaging modalities, and exploiting larger and more diverse datasets to enhance the generalisability and robustness of medical IQA models.

While MIQANet achieves strong performance, several limitations remain. First, the current model is trained and evaluated on simulated MRI artifacts from a single curated dataset; future work should explore its generalisability across real-world clinical data and multiple scanner types. Second, the model currently relies on supervised learning using subjective scores, which may introduce bias. Incorporating semi-supervised or self-supervised strategies could mitigate this limitation. Finally, expanding the framework to handle dynamic sequences

and integrating uncertainty quantification are promising future directions.

REFERENCES

[1] Z. Wang and A. C. Bovik, "Modern image quality assessment," Ph.D. dissertation, Springer, 2006.

[2] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *arXiv preprint arXiv:1406.7799*, 2014.

[3] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 29–40, 2011.

[4] Q. Jiang, Y. Gu, C. Li, R. Cong, and F. Shao, "Underwater image enhancement quality evaluation: Benchmark dataset and objective metric," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5959–5974, 2022.

[5] Q. Jiang, Y. Kang, Z. Wang, W. Ren, and C. Li, "Perception-driven deep underwater image enhancement without paired supervision," *IEEE Transactions on Multimedia*, vol. 26, pp. 4884–4897, 2023.

[6] Q. Jiang, Z. Liu, K. Gu, F. Shao, X. Zhang, H. Liu, and W. Lin, "Single image super-resolution quality assessment: a real-world dataset, subjective studies, and an objective metric," *IEEE Transactions on Image Processing*, vol. 31, pp. 2279–2294, 2022.

[7] Q. Jiang, X. Li, X. Wang, Z. Wang, and G. Zhai, "Dataset and metric for quality assessment of hdr tone mapping: Detail visibility, color naturalness, and overall quality," *IEEE Transactions on Multimedia*, 2025.

[8] Q. Jiang, Y. Gu, Z. Wu, C. Li, H. Xiong, F. Shao, and Z. Wang, "Deep underwater image quality assessment with explicit degradation awareness embedding," *IEEE Transactions on Image Processing*, 2025.

[9] Q. Jiang, Z. Liu, S. Wang, F. Shao, and W. Lin, "Toward top-down just noticeable difference estimation of natural images," *IEEE Transactions on Image Processing*, vol. 31, pp. 3697–3712, 2022.

[10] H. Herath, H. Herath, N. Madusanka, and B.-I. Lee, "A systematic review of medical image quality assessment," *Journal of Imaging*, vol. 11, no. 4, p. 100, 2025.

[11] R. Rodrigues, L. Lévêque, J. Gutiérrez, H. Jebbari, M. Outtas, L. Zhang, A. Chetouani, S. Al-Juboori, M. G. Martini, and A. M. Pinheiro, "Objective quality assessment of medical images and videos: review and challenges," *Multimedia Tools and Applications*, pp. 1–34, 2024.

[12] S. J. Esses, X. Lu, T. Zhao, K. Shanbhogue, B. Dane, M. Bruno, and H. Chandarana, "Automated image quality evaluation of t2-weighted liver mri utilizing deep learning architecture," *Journal of Magnetic Resonance Imaging*, vol. 47, no. 3, pp. 723–728, 2018.

[13] S. J. Sujit, I. Coronado, A. Kamali, P. A. Narayana, and R. E. Gabr, "Automated image quality evaluation of structural brain mri using an ensemble of deep learning networks," *Journal of Magnetic Resonance Imaging*, vol. 50, no. 4, pp. 1260–1267, 2019.

[14] J. J. Ma, U. Nakarmi, C. Y. S. Kin, C. M. Sandino, J. Y. Cheng, A. B. Syed, P. Wei, J. M. Pauly, and S. S. Vasanawala, "Diagnostic image quality assessment and classification in medical imaging: Opportunities and challenges," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 337–340.

[15] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proceedings of the 26th IEEE international symposium on computer-based medical systems*. IEEE, 2013, pp. 95–100.

[16] S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, and D. Qian, "Human visual system-based fundus image quality assessment of portable fundus camera photographs," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1046–1055, 2015.

[17] A. W. Moawad, A. Janas, U. Baid, D. Ramakrishnan, L. Jekel, K. Krantchev, H. Moy, R. Saluja, K. Osenberg, K. Wilms *et al.*, "The brain tumor segmentation (brats-mets) challenge 2023: Brain metastasis segmentation on pre-treatment mri," *ArXiv*, 2023.

[18] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer Jr, R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammography," in *Digital Mammography: Nijmegen, 1998*. Springer, 1998, pp. 457–460.

[19] Y. Wen, L. Chen, Y. Deng, and C. Zhou, "Rethinking pre-training on medical imaging," *Journal of Visual Communication and Image Representation*, vol. 78, p. 103145, 2021.

[20] H. Kasban, M. El-Bendary, and D. Salama, "A comparative study of medical imaging techniques," *International Journal of Information Science and Intelligent System*, vol. 4, no. 2, pp. 37–58, 2015.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[22] Y. Ma, J. Lou, J.-Y. Tanguy, P. Corcoran, and H. Liu, "Rad-iqmri: A benchmark for mri image quality assessment," *Neurocomputing*, p. 128292, 2024.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[24] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE transactions on image processing*, vol. 9, no. 4, pp. 636–650, 2000.

[25] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on hvs," in *Proceedings of the second international workshop on video processing and quality metrics*, vol. 4, 2006, p. 4.

[26] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of dct basis functions," in *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4. Scottsdale USA, 2007.

[27] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting," in *2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*. IEEE, 2011, pp. 305–311.

[28] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on image processing*, vol. 20, no. 5, pp. 1185–1198, 2010.

[29] S. Rezazadeh and S. Coulombe, "A novel discrete wavelet transform framework for full reference image quality assessment," *Signal, image and video processing*, vol. 7, pp. 559–573, 2013.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[31] S. Rezazadeh and S. Coulombe, "A novel approach for computing and pooling structural similarity index in the discrete wavelet domain," in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 2209–2212.

[32] L. Zhang, L. Zhang, and X. Mou, "Rfsim: A feature based image quality assessment metric using riesz transforms," in *2010 IEEE International Conference on Image Processing*. IEEE, 2010, pp. 321–324.

[33] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[34] L. Zhang and H. Li, "Sr-sim: A fast and high performance iqa index based on spectral residual," in *2012 19th IEEE international conference on image processing*. IEEE, 2012, pp. 1473–1476.

[35] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE transactions on image processing*, vol. 23, no. 2, pp. 684–695, 2013.

[36] T. Wang, L. Zhang, H. Jia, B. Li, and H. Shu, "Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment," *Signal Processing: Image Communication*, vol. 45, pp. 1–9, 2016.

[37] A. Balanov, A. Schwartz, Y. Moshe, and N. Peleg, "Image quality assessment based on dct subband similarity," in *2015 IEEE international conference on image processing (ICIP)*. IEEE, 2015, pp. 2105–2109.

[38] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[39] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.

[40] S. Rezazadeh and S. Coulombe, "Low-complexity computation of visual information fidelity in the discrete wavelet domain," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 2438–2441.

[41] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.

[42] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[43] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[44] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, 2015.

[45] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[46] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733–1740.

[47] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.

[48] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3667–3676.

[49] J. Wang, K. C. Chan, and C. C. Loy, "Exploring clip for assessing the look and feel of images," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 2555–2563.

[50] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, "Maniqa: Multi-dimension attention network for no-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1191–1200.

[51] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[52] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, pp. 355–362, 2018.

[53] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on image processing*, vol. 27, no. 1, pp. 206–219, 2017.

[54] K.-Y. Lin and G. Wang, "Hallucinated-iqa: No-reference image quality assessment via adversarial learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 732–741.

[55] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[58] H. L. Kundel and P. S. La Follette Jr, "Visual search patterns and experience with radiological images," *Radiology*, vol. 103, no. 3, pp. 523–528, 1972.

[59] S. Waite, A. Grigorian, R. G. Alexander, S. L. Macknik, M. Carrasco, D. J. Heeger, and S. Martinez-Conde, "Analysis of perceptual expertise in radiology–current knowledge and a new perspective," *Frontiers in human neuroscience*, vol. 13, p. 213, 2019.

[60] R. Bertram, J. Kaakinen, F. Bensch, L. Helle, E. Lantto, P. Niemi, and N. Lundbom, "Eye movements of radiologists reflect expertise in ct study interpretation: a potential tool to measure resident development," *Radiology*, vol. 281, no. 3, pp. 805–815, 2016.

[61] H.-B. Park, L. Azer, S. Ahn, T.-D. Dinh, G. Macias, G. Zhang, B. B. Chen, H. Ma, M. Botejue, E. H. Choi *et al.*, "Contributions of global and local processing on medical image perception," *Journal of Medical Imaging*, vol. 10, no. S1, pp. S11 911–S11 911, 2023.

[62] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[63] J. W. Kim, A. U. Khan, and I. Banerjee, "Systematic review of hybrid vision transformer architectures for radiological image analysis," *Journal of Imaging Informatics in Medicine*, pp. 1–15, 2025.

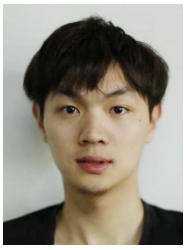
[64] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[65] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.

[66] P. ITU-T, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models (p. 1401)," 2020.

[67] J. Antkowiak, T. J. Baina, F. V. Baroncini, N. Chateau, F. FranceTelecom, A. C. F. Pessoa, F. S. Colonnese, I. L. Contin, J. Caviedes, and F. Philips, "Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000," *Final report from the video quality experts group on the validation of objective models of video quality assessment march*, vol. 10, 2000.

Yueran Ma received the B.Eng. from Beijing Jiaotong University, in 2016 and the M.S. degree from Southern Methodist University in 2018. He is now pursuing his Ph.D. degree at the School of Computer Science and Informatics, Cardiff University, Cardiff, UK. His interests are Image Processing, Biomedical Image Processing, Image Quality Assessment and Saliency Prediction.



Huasheng Wang received the M.S. degree from Dalian University of Technology in 2021 and the Ph.D. degree from Cardiff University in 2024. He is currently an Algorithm Engineer at Taobao, Alibaba. His research interests include Image and Video Quality Assessment and Saliency Prediction.

Yingying Wu received MSc in Data Science and Analytics from Cardiff University, United Kingdom in 2020. She is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics at Cardiff University, United Kingdom. Her research interests include image data analysis, human visual perception, and machine learning.

Jean-Yves Tanguy M.D., is a Neuroradiologist and Head and Neck imaging specialist at the University Hospital Center in Angers, France. He has given lessons on technical aspects of medical imaging to medicine students, and future radiologists, technicians, and engineers in Angers Faculty of Medicine, and ESEO since the beginning of his career.

Richard White received degrees in medical science and medicine from the University of St Andrews, St Andrews, U.K., and The University of Manchester, Manchester, U.K. He underwent radiology training in Dundee, U.K., during which time he was conferred with Fellowship of the Royal College of Radiologists. He took up a consultant post with the University Hospital of Wales, Cardiff, U.K., in 2013 and is currently a Consultant Vascular and Interventional Radiologist with research interests, which include vascular and interventional radiology, innovations, and AI. He is also the radiology R&D lead and a Registered Clinical Radiation Expert with Health Research Authority.

Phillip Wardle received specialist training in Clinical Radiology through the All Wales Specialist Training Scheme and completed Musculoskeletal Radiology fellowships in Melbourne, Australia. He currently serves as Director of the National Imaging Academy Wales and Consultant Musculoskeletal Radiologist at Cwm Taf Morgannwg University Health Board, based at the Royal Glamorgan Hospital.

Elizabeth Krupinski received the Ph.D. degree in Psychology and is currently Professor and Vice Chair for Research in the Department of Radiology and Imaging Sciences at Emory University School of Medicine, Atlanta, GA, USA. Prior to this, she was Professor and Vice Chair of Research in Radiology at the University of Arizona, with joint appointments in Psychology and Public Health. She is a past Chair of the SPIE Medical Imaging Conference, past President of the American Telemedicine Association, and past Chair of the Society for Imaging Informatics in Medicine. Her research interests include medical image perception, telemedicine, and imaging informatics. She is a member of the Cancer Prevention and Control Research Program at Winship Cancer Institute.

Padraig Corcoran is a Reader and the Director of Research in the School of Computer Science and Informatics at Cardiff University, UK. His research interests are in the fields of network science and operations research.

Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands in 2011. He is currently a Professor at the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests sit at the intersection of Image Processing, Machine Learning, Computer Vision, Applied Perception, and Medical Imaging.