Deep Feature Fusion with a Hybrid CNN–Transformer Framework for High-Fidelity Coronary Angiography Enhancement and Fine Vascular Structure Preservation

Romaan Khan[1], Muhammad Younas[1], Nizam Ahmad[1], Muhammad Khan[1], Islam Uddin[2]*
[1]Institute of Computer Science and Information Technology, University of Agriculture Peshawar, 25000, Pakistan
[2]Department of Computer Science, Abdul Wali Khan University Mardan, 23200, Mardan, KPK, Pakistan
***Correspondence**: islamuddin@awkum.edu.pk

Citation | Khan. R, Younas. M, Ahmad. N, Khan. M, Uddin. I, "Deep Feature Fusion with a Hybrid CNN–Transformer Framework for High-Fidelity Coronary Angiography Enhancement and Fine Vascular Structure Preservation", IJIST, Vol. 8 Issue. 5 pp 2344-2370, August 2026
DOI: https://doi.org/10.33411/IJIST/1998
Received | Aug 10, 2026 **Revised |** Aug 23, 2026 **Accepted |** Aug 26, 2026 **Published |** Aug 31, 2026.

Coronary Artery Disease (CAD) is a major cause of mortality worldwide, and X-ray coronary angiography remains an important imaging modality for evaluating coronary artery morphology and stenosis. However, low contrast, noise, motion blur, and compression artifacts can degrade angiographic images, obscure fine vascular structures, and reduce the reliability of subsequent image interpretation and automated analysis. To address these challenges, we propose a hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) with Transformer-based self-attention for high-fidelity coronary angiography enhancement. The CNN component extracts local features for noise suppression, edge refinement, and recovery of fine vessel details, while the Transformer component models long-range spatial dependencies to improve global anatomical continuity and structural coherence. The framework was trained in a supervised manner using high-quality images from the ARCADE dataset and their synthetically degraded counterparts. Quantitative evaluation demonstrated that the proposed Hybrid CNN–Transformer model achieved a PSNR of 36.33 dB and an SSIM of 0.9733, outperforming the Transformer-only model (34.43 dB, 0.9674 SSIM) by 1.90 dB in PSNR and 0.0059 in SSIM. The ablation analysis further showed that the proposed optimized hybrid configuration improved PSNR by 2.42 dB and SSIM by 0.0077 compared with the CNN-only configuration (33.91 dB, 0.9656 SSIM). The proposed model also achieved an MSE of 0.00023281 and an RMSE of 0.015258, representing the lowest reconstruction error among the evaluated methods. Qualitative assessment indicated improved vessel visibility, contrast, edge definition, thin-vessel representation, and preservation of vascular continuity. The results demonstrate that combining local convolutional feature extraction with global Transformer-based contextual modeling provides an effective approach for enhancing coronary angiographic images while preserving fine vascular structures, offering a promising preprocessing framework for subsequent computer-aided cardiovascular analysis.

Keywords: Deep learning; Coronary angiography; Image enhancement; Hybrid CNN–Transformer; Structural preservation; Feature fusion; Medical image restoration; Convolutional neural networks.

Introduction:

Coronary artery disease (CAD) remains a major cardiovascular health burden, making accurate visualization of the coronary vasculature essential for assessing vessel morphology, luminal narrowing, stenosis, and other clinically relevant abnormalities [1]. X-ray coronary angiography is an important imaging modality for providing real-time visualization of coronary vessels and is routinely used to support diagnostic assessment and interventional procedures [2]. The quality of angiographic images, however, can be affected by several acquisition-related factors, including limited contrast, image noise, patient motion, blur, uneven illumination, and compression artifacts. These limitations can reduce the visibility of small and distal coronary vessels, weaken vessel boundaries, and obscure complex branching patterns that are important for accurate interpretation [3]. Such degradation can also influence the reliability of subsequent computer-aided analysis, where image quality and structural fidelity directly affect the performance of automated image-processing and diagnostic systems. Recent developments in artificial intelligence and deep learning have created new opportunities for improving medical image quality by learning complex mappings between degraded and high-quality images [4]. In particular, deep learning-based restoration approaches have demonstrated potential for noise reduction, contrast improvement, structural recovery, and preservation of fine anatomical details [5]. The current manuscript builds on this direction by focusing specifically on the enhancement of coronary angiographic images, where preserving both fine vascular structures and long-range vessel continuity is particularly important [6]. The study therefore addresses the image-quality limitations encountered in angiographic imaging through a structure-sensitive enhancement framework that combines local feature extraction with global contextual modeling [7].

Recent advances in deep learning have substantially expanded the capabilities of medical image enhancement and restoration, particularly through models that can learn complex relationships between degraded observations and high-quality references. CNN-based approaches have been widely adopted for denoising, super-resolution, contrast restoration, and detail reconstruction because their hierarchical convolutional representations are well suited to capturing localized patterns such as edges, textures, and fine anatomical structures [8]. In cardiovascular imaging, these properties have been exploited to improve vessel visibility, suppress image noise, and recover local anatomical details in X-ray and CT angiographic images [9]. However, the inherently local nature of convolutional operations can limit the ability of CNNs to establish relationships between spatially distant regions of an image [10]. This limitation is particularly relevant to coronary angiography, where vessels extend across large image regions and form complex branching networks in which continuity between distant segments is important [11]. Transformer-based architectures provide a complementary modeling capability by using self-attention to establish relationships among spatial features over broader image regions [12]. Their ability to capture long-range contextual information can therefore support the preservation of vessel continuity and global anatomical organization during image restoration [13]. At the same time, Transformer-based models may require stronger local feature representations to recover subtle edges and fine vessel textures. These complementary characteristics have motivated increasing interest in hybrid CNN–Transformer architectures, in which convolutional representations provide detailed local information while attention mechanisms contribute broader contextual understanding [14]. Such integration offers a promising direction for coronary angiography enhancement because successful restoration requires both accurate recovery of fine vascular details and consistent representation of vessels across complex anatomical structures [15].

Despite these advances, important limitations remain in existing approaches for coronary angiography enhancement [16]. Conventional techniques, including histogram equalization, contrast-limited adaptive histogram equalization (CLAHE), and multiscale

vesselness filtering, generally rely on predefined intensity or structural transformations. Their performance can therefore vary across images acquired under different contrast conditions, noise levels, motion patterns, and acquisition settings. Although these methods can improve overall image contrast, they may simultaneously amplify noise or fail to recover weak and fine vascular structures, particularly in regions with low signal-to-noise ratios [17]. CNN-based restoration methods provide a more adaptive alternative by learning enhancement mappings directly from image data, but their emphasis on local receptive fields can limit the modeling of spatial relationships between distant vessel segments [18]. This limitation becomes important in coronary angiograms, where vessels extend across broad image regions and contain complex bifurcations that require structural continuity to be maintained during reconstruction [19]. Transformer-based models address this limitation by capturing broader contextual relationships through self-attention, yet Transformer-only architectures may be less effective at preserving subtle edges, textures, and fine vascular details because their global modeling capability does not inherently provide the same local inductive bias as convolutional operations [20]. In addition, increasing model complexity can introduce computational costs that may constrain practical deployment. These limitations indicate that no single modeling strategy fully addresses the requirements of high-quality coronary angiography enhancement [21]. A more balanced approach is therefore needed to combine detailed local representation, long-range anatomical context, structural preservation, and computational efficiency within a unified enhancement framework. Such an approach could provide more consistent restoration of diagnostically relevant coronary structures while reducing the limitations associated with conventional, CNN-only, and Transformer-only methods.

To address these limitations, this study develops a structure-sensitive hybrid CNN–Transformer framework for enhancing coronary angiographic images while preserving fine vascular details and long-range anatomical continuity. The proposed framework follows a progressive enhancement strategy in which a convolutional encoder first extracts local representations associated with vessel boundaries, textures, contrast variations, and high-frequency details. These representations are subsequently refined through Transformer-based self-attention, enabling the model to capture spatial relationships among anatomically related regions and maintain continuity across extended and bifurcating coronary vessels. A dedicated feature-fusion mechanism then integrates the local convolutional representations with the globally refined features before a reconstruction decoder generates the enhanced angiographic image. The model is trained using paired high-quality and synthetically degraded images derived from the ARCADE dataset, allowing it to learn a supervised mapping from degraded inputs to corresponding reference images. Its reconstruction performance is evaluated quantitatively using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), while qualitative assessment is used to examine vessel visibility, edge definition, artifact suppression, and structural continuity. In addition, comparative experiments and ablation analysis are conducted to determine whether the integration of local and global representations provides measurable advantages over individual CNN and Transformer configurations. Through this design, the study aims to establish a balanced enhancement framework that combines local detail recovery with global structural modeling and provides high-fidelity restoration of coronary vascular structures under controlled degradation conditions.

The main contributions of this study are summarized as follows:

We develop a hybrid CNN–Transformer architecture for coronary angiography enhancement that combines convolution-based local feature extraction with Transformer-based global contextual modeling.

We introduce a global–local feature fusion strategy that integrates detailed vessel representations from the CNN branch with long-range contextual information from the

Transformer branch, with the aim of improving fine vascular detail recovery while maintaining anatomical continuity.

We conduct quantitative and qualitative experiments against conventional image enhancement methods and deep learning-based baselines. We further perform systematic ablation experiments to examine the individual and combined contributions of the CNN, Transformer, and feature-fusion components.

The proposed framework is designed specifically to address the complementary requirements of coronary angiography enhancement, namely recovery of fine vessel details and preservation of long-range structural continuity. This design provides a foundation for further validation on real-world angiographic data and downstream computer-aided cardiovascular analysis.

The remainder of this paper is organized as follows. Section 2 presents the related literature on conventional angiographic enhancement, CNN-based restoration, Transformer-based image restoration, and hybrid CNN–Transformer approaches. Section 3 describes the proposed framework, including dataset preparation, preprocessing, synthetic degradation, CNN-based local feature extraction, Transformer-based global refinement, feature fusion, reconstruction, and experimental settings. Section 4 presents the experimental results, including quantitative and qualitative evaluations, computational efficiency analysis, and ablation experiments, followed by a detailed discussion of the findings. Section 5 concludes the study and outlines specific recommendations and directions for future research, including validation using real-world and multi-center angiographic datasets and evaluation on downstream cardiovascular imaging tasks.

Literature Review:

Medical image enhancement has been extensively investigated because image quality directly affects structural visibility and the reliability of subsequent image analysis. In coronary angiography, this requirement is particularly important because clinically relevant information may be contained in small vessels, narrow vessel segments, and complex vascular branches. Early approaches to angiographic enhancement were primarily based on classical image-processing techniques, including histogram equalization and contrast-limited adaptive histogram equalization (CLAHE). [22] demonstrated the use of contrast enhancement techniques to improve image visibility through intensity redistribution. Further improvement in the visualization of tubular structures was achieved through multiscale vesselness filtering. [23] employed a multiscale filtering approach based on second-order Hessian information to enhance the representation of vascular structures. Although these methods can improve vessel contrast under controlled conditions, they are generally dependent on predefined transformations and parameter settings and may amplify noise or artifacts when applied to low-quality images. Similar limitations were discussed by [24], who highlighted the challenges associated with image-quality variations and artifact correction in cardiovascular imaging. Consequently, conventional enhancement methods may have limited adaptability to changes in signal-to-noise ratio, motion artifacts, contrast conditions, and compression effects encountered in coronary angiography.

The development of deep learning has provided more adaptive approaches to medical image restoration by enabling models to learn complex mappings directly from image data. [25] introduced a CNN-based super-resolution approach that demonstrated the ability of convolutional networks to learn nonlinear mappings between low-quality and high-quality image representations. Subsequent studies extended CNN-based learning to image denoising and restoration. [26] investigated convolutional denoising autoencoders for medical image denoising, while [27] developed a residual learning-based CNN approach for effective image denoising. In cardiovascular imaging, [28] demonstrated the effectiveness of deep learning-based image restoration for coronary CT angiography, highlighting the potential of learned

restoration methods for improving cardiovascular image quality. CNN-based architectures, including encoder–decoder networks and residual blocks, are effective at extracting hierarchical spatial representations and reconstructing local textures, edges, and fine anatomical structures. However, their predominantly local receptive fields can restrict the direct modeling of long-range spatial relationships. This limitation is particularly relevant to coronary angiography, where vessels extend across large image regions and contain complex branching patterns. Insufficient modeling of these relationships may lead to discontinuities or inconsistencies in the reconstruction of extended vascular structures.

Transformer-based architectures have more recently been adopted in computer vision and medical image analysis because self-attention mechanisms can model relationships among spatially separated features. The Vision Transformer (ViT) introduced by [29] demonstrated the potential of attention-based representations for modeling global image relationships. [30] proposed the Swin Transformer, which employs hierarchical shifted-window attention to improve the efficiency and scalability of Transformer-based visual representation learning. [31] investigated Transformer-based restoration through Restormer, demonstrating the potential of attention mechanisms for high-resolution image restoration. A major advantage of Transformer-based architectures is their ability to capture long-range spatial dependencies, which is particularly relevant to maintaining continuity across extended coronary vessels and complex branching structures. However, Transformer-only models can require greater computational resources and may have weaker local inductive biases for preserving fine-grained edges, textures, and small vascular details compared with convolutional architectures. Recent Transformer-based restoration approaches, including Restormer, SwinIR, and Uformer, have demonstrated promising capabilities for image denoising, super-resolution, and medical image reconstruction. However, their computational complexity and limited exploitation of convolutional local representations provide motivation for hybrid CNN–Transformer architectures that can jointly model local details and global contextual relationships.

To overcome these complementary limitations, hybrid CNN–Transformer architectures have been developed to combine convolutional local feature extraction with global attention-based contextual modeling. [32] proposed a hybrid CNN–Transformer architecture for medical image segmentation that integrates local and global representations to improve structural modeling. Similarly, [33] investigated a Transformer-assisted image restoration framework and demonstrated that combining convolutional processing with attention mechanisms can improve both pixel-level reconstruction and structural consistency. These developments indicate that CNN and Transformer components can provide complementary representations when integrated within a unified architecture. However, the application of hybrid CNN–Transformer architectures specifically to coronary angiography enhancement remains comparatively limited. Existing studies have predominantly focused on medical image segmentation, general restoration, or reconstruction rather than the specific enhancement requirements of coronary angiographic images. These images present distinctive challenges, including substantial contrast variation, limited visibility of small and distal vessels, complex vascular bifurcations, noise, motion-related degradation, and compression artifacts. Therefore, a dedicated enhancement framework is needed that can simultaneously recover fine local vascular details and preserve long-range anatomical continuity. This research gap motivates the development of the proposed hybrid CNN–Transformer framework, which integrates local convolutional feature extraction with global contextual refinement to achieve structure-sensitive enhancement of coronary angiographic images.

Materials and Methods:

This section describes the materials, experimental procedures, and proposed methodology used for coronary angiography image enhancement. It includes the dataset

preparation, preprocessing pipeline, construction of paired high-quality and low-quality images, architecture of the proposed hybrid CNN–Transformer framework, training procedure, and evaluation strategy. The methodological workflow is designed to provide a reproducible and consistent experimental setting for assessing the ability of the proposed model to recover fine vascular details and preserve broader anatomical continuity. Details of the computational environment and experimental configuration are also provided to support reproducibility and fair comparison with the selected baseline methods.

Proposed Framework:

The proposed framework is a hybrid CNN–Transformer architecture developed to enhance low-quality coronary angiographic images while preserving fine vascular structures and overall anatomical continuity. The framework follows a sequential local-to-global representation strategy in which convolutional feature extraction is followed by Transformer-based contextual refinement and feature fusion. The processing pipeline consists of three principal stages: local feature extraction, global contextual modeling with feature fusion, and image reconstruction. Given a low-quality coronary angiogram, the image is first processed by a CNN encoder composed of residual convolutional blocks. The CNN encoder extracts hierarchical local representations associated with vessel edges, textures, contrast patterns, and high-frequency structural information. The use of convolutional operations provides a strong inductive bias toward local spatial relationships and supports the extraction of fine vascular features and suppression of localized image degradation. The resulting feature maps represent locally refined information from diagnostically relevant regions of the angiogram. The extracted CNN feature maps are subsequently provided to the global contextual refinement stage. In this stage, the spatial feature representation is transformed into token representations and processed using Transformer-based self-attention. Multi-head self-attention enables the model to establish relationships between spatially separated feature representations and thereby capture long-range contextual dependencies across the angiographic image. This global modeling capability is particularly important for coronary vessels that extend across broad image regions or contain complex bifurcations, where information from distant vessel segments may contribute to maintaining structural continuity.

Following Transformer-based refinement, the global contextual representation is integrated with the local CNN representation through the feature-fusion mechanism. The fusion operation combines complementary local and global information so that fine vessel details extracted by the CNN are retained while broader anatomical relationships modeled by the Transformer are incorporated into the enhanced representation. The fused features are then passed to a reconstruction decoder consisting of convolutional residual processing blocks. The decoder progressively transforms the fused representation into the final enhanced angiographic image, with the objective of recovering image contrast, reducing residual artifacts, and reconstructing structurally consistent vascular details. The complete mapping can therefore be viewed as a sequence of local feature extraction, global contextual refinement, feature integration, and image reconstruction [34]. This design allows the proposed framework to address two complementary requirements of coronary angiography enhancement: recovery of fine-grained local vessel information and preservation of long-range anatomical continuity. Figure 1 presents the overall architecture of the proposed hybrid CNN–Transformer framework, including the CNN encoder, Transformer-based global refinement, feature-fusion stage, and reconstruction decoder.

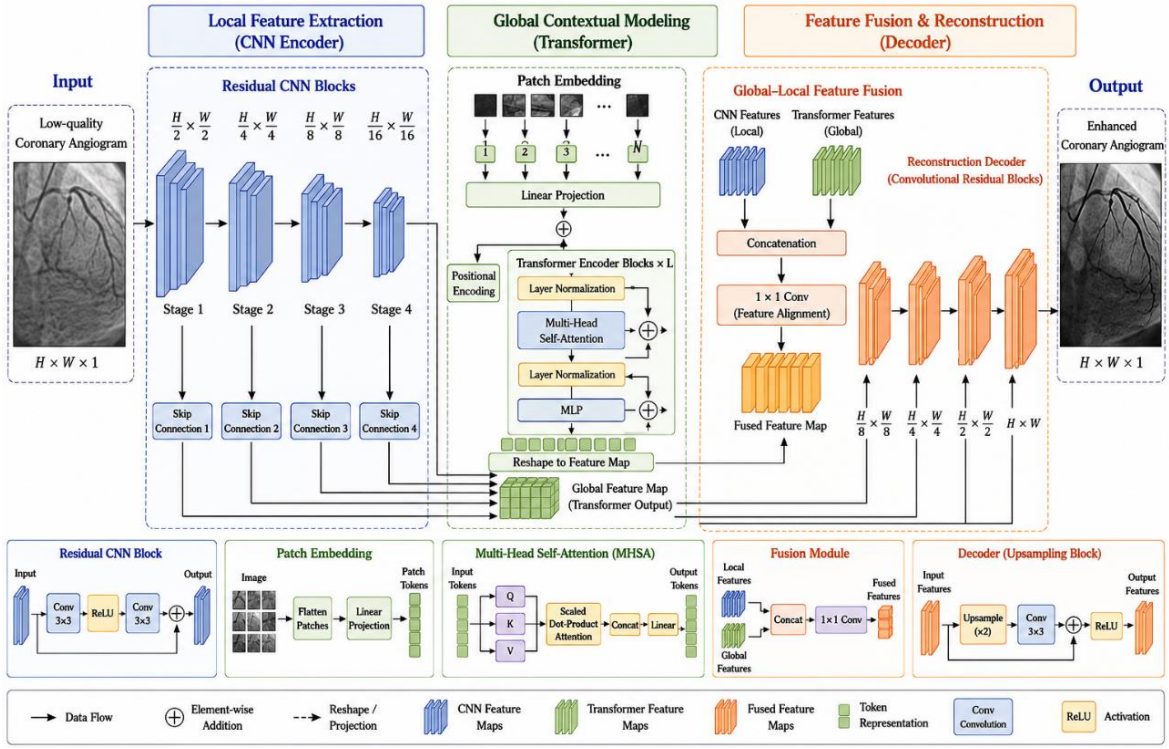


Figure 1. Architecture of the Proposed Hybrid CNN–Transformer Framework for Coronary Angiography Image Enhancement

Dataset Preparation:

The ARCADE coronary angiography database was used as the primary dataset for developing and evaluating the proposed enhancement framework. The database contains clinically acquired X-ray coronary angiographic images obtained under different imaging conditions and exhibiting variations in coronary anatomical appearance. In this study, a total of 1,600 angiographic images were used for the experiments. The images were organized into paired high-quality (HQ) reference images and corresponding low-quality (LQ) images to formulate the supervised image-enhancement problem. The original angiographic images were treated as the HQ reference images, while the corresponding LQ inputs were generated synthetically through a controlled degradation process. This paired-data formulation enables the proposed network to learn a direct mapping from degraded angiographic observations to their corresponding high-quality references.

To reproduce common image-quality limitations encountered during angiographic acquisition, the degradation procedure incorporated several controlled degradation operations, including Gaussian blurring, spatial resolution reduction, additive noise, and contrast attenuation. These operations were selected to represent image degradations associated with motion or defocus, reduced image resolution, sensor-related noise, and low-contrast imaging conditions. For a high-quality angiographic image I_HQ, the corresponding low-quality image, I_LQ is represented by the following degradation formulation:

$$I_{lq} = D(I_{hq}) \quad (1)$$

Where I_HQ denotes the high-quality reference image, I_LQrepresents the synthetically degraded image, and D(·) denotes the composite degradation function used to generate the low-quality input.

The degradation process was designed to provide controlled and consistent training inputs while retaining the original images as reference targets [35]. Specifically, Gaussian blurring was used to simulate blur-related degradation, spatial down sampling was applied to represent reduced image resolution, additive noise was introduced to model noise

contamination, and contrast attenuation was used to reproduce reduced vessel visibility. The combination of these operations provides a controlled representation of several image-quality limitations that may occur during angiographic acquisition.

For a more explicit representation of the degradation pipeline, the low-quality image can be expressed as:

$$I_lq = C(U(S(I_hq * K\sigma)) + N\sigma_n) \quad (2)$$

Where $K\sigma$ represents the Gaussian blur kernel with parameter σ, * denotes convolution, $S(\cdot)$ represents spatial downsampling, $U(\cdot)$ represents upsampling to the target image resolution, $N\sigma_n$ denotes additive Gaussian noise with noise parameter σ_n, and $C(\cdot)$ represents the contrast-attenuation operation.

A patient-independent data-splitting strategy was adopted to reduce the possibility of information leakage between the training, validation, and testing subsets. The complete dataset was divided into 1,000 images for training, 300 images for validation, and 300 images for independent testing [2]. These subsets correspond to 62.50%, 18.75%, and 18.75% of the complete dataset, respectively. The testing subset was kept separate from model development and was used for the final quantitative and qualitative evaluation of the proposed framework and competing methods. The distribution of the dataset is presented in Table 1.

Table 1. Distribution of the ARCADE Coronary Angiography Dataset Across Training, Validation, and Testing Sets

Dataset Split	Number of Images	Percentage (%)
Total Dataset	1600	100%
Training Set	1000	62.5%
Validation Set	300	18.75%
Testing Set	300	18.75%

Preprocessing:

A standardized preprocessing pipeline was applied to the coronary angiography images before model training to improve input consistency and facilitate stable optimization of the proposed hybrid CNN–Transformer framework [36]. The raw angiograms may contain variations in contrast, illumination, noise, motion-related blur, and compression artifacts that can affect both visual quality and the learning process. Therefore, the preprocessing procedure was designed to standardize image representation, control the spatial resolution, normalize pixel intensities, and generate paired low-quality inputs for supervised enhancement. First, all angiographic images were converted to grayscale to provide a consistent single-channel representation [37]. The images were subsequently resized to 128 × 128 pixels. This resolution was selected to reduce computational requirements associated with Transformer-based attention while retaining the major structural information of the coronary vasculature. Pixel intensities were then normalized to the range [0, 1], providing a consistent numerical scale for model optimization.

For supervised image enhancement, the original high-quality (HQ) images were used as reference targets; while corresponding low-quality (LQ) images were generated through a composite degradation process. The degradation procedure incorporated Gaussian blurring, spatial downsampling and upsampling, additive Gaussian noise, and contrast attenuation. The general degradation process can be expressed as:

$$I_LQ = C(U(S(I * K_\sigma)) + N_\sigma_n) \quad (3)$$

Where I_hq denotes the high-quality reference angiogram, I_lq represents the resulting low-quality image, $K\sigma$ denotes the Gaussian blur kernel, * represents convolution, $S(\cdot)$ denotes spatial down sampling, $U(\cdot)$ represents up sampling, $N\sigma_n$ denotes additive Gaussian noise, and $C(\cdot)$ represents contrast attenuation.

The Gaussian blur kernel used to model motion- and defocus-related degradation is defined as:

$$K_\sigma(x,y) = 1/2\pi\sigma^2 \exp[-(x^2 + y^2)/2\sigma^2] \quad (4)$$

Where x and y represent the spatial coordinates of the kernel, and σ controls the spread of the Gaussian distribution and consequently the strength of the simulated blur.

Pixel intensities were normalized to the range [0, 1] using:

$$I_{norm} = (I - I_{min})/(I_{max} - I_{min}) \quad (5)$$

Where I represents the original pixel intensity, I_{min} and I_{max} denote the minimum and maximum pixel intensities, respectively, and I_{norm} represents the normalized pixel value.

Contrast attenuation was additionally applied to simulate reduced-visibility or low-exposure imaging conditions:

$$I_c = \alpha I_{norm} + \beta \quad (6)$$

Where Ic denotes the contrast-adjusted image, α controls the contrast level, and β represents the brightness offset.

To characterize tubular vascular structures, second-order image derivatives were represented using the Hessian matrix:

$$H(x,y) = \blacksquare(I_{xx} \& I_{xy} @ I_{yx} \& I_{yy}) \quad (7)$$

Where I_{xx}, $I_{x\gamma}$, $I_{\gamma x}$, and $I_{\gamma\gamma}$ represent the second-order partial derivatives of image intensity with respect to the spatial coordinates x and y.

The vesselness response was then computed from the eigenvalues of the Hessian matrix using a Frangi-style formulation:

$$V_\sigma = (\{ 0, \lambda_2 > 0; \ \exp[-R^2)/((2\alpha^2)] \ [1 - \exp(-S^2/(2\beta^2))], \lambda_2 \leq 0 \}) \quad (8)$$

Where Vσ denotes the vesselness response, λ_1 and λ_2 are the ordered eigenvalues of the Hessian matrix, R_a represents the eigenvalue ratio, S represents the second-order structural strength, and α and β control the sensitivity of the vesselness response.

The corresponding structural measures are defined as:

$$R_A = (|\lambda_1|)/(|\lambda_2|) \ , \ S = \sqrt{((\lambda_1^2 + \lambda_2^2))} \quad (9)$$

Where λ_1 and λ_2 are the ordered Hessian eigenvalues. R_a characterizes the local shape of the structure, whereas S measures the overall strength of the second-order response. The resulting vesselness representation emphasizes elongated tubular structures and provides additional structural information relevant to coronary vessel visualization.

The preprocessing pipeline standardizes the spatial and intensity characteristics of the input images, generates controlled degraded samples for supervised restoration, and incorporates vessel-oriented structural information to support the enhancement of coronary vascular features. These preprocessing operations provide consistent inputs for the subsequent CNN-based local feature extraction and Transformer-based global contextual refinement stages.

CNN-Based Local Feature Enhancement Network:

The first component of the proposed framework is a convolutional neural network (CNN)-based local feature enhancement module designed to extract hierarchical spatial representations from low-quality coronary angiographic images. CNNs are particularly effective for learning localized patterns, including vessel edges, fine textures, intensity transitions, and high-frequency anatomical structures [38]. These local representations provide the foundation for recovering subtle vascular details that may be degraded by noise, blur, and reduced contrast.

Given a low-quality input angiogram I_LQ, the feature representation at the l-th convolutional layer is defined as:

$$F^{(l)} = \sigma(W^{(l)} * F^{(l-1)} + b^{(l)}) \quad (10)$$

Where F^((l-1)) and F^((l)) denote the input and output feature representations of the lth convolutional layer, respectively; W^((l)) represents the learnable convolutional weights; b^((l)) denotes the corresponding bias term; * denotes the convolution operation; and σ(·) represents the ReLU activation function.

Residual learning is incorporated to facilitate stable feature propagation and preserve useful image information during feature extraction. The residual feature representation is expressed as:

$$F_res = F^{(l)} + F^{(l-1)} \quad (11)$$

Where F_res denotes the resulting residual feature representation, while F^((l)) and F^((l-1)) represent the current and preceding feature maps, respectively. The residual connection enables the network to learn enhancement-related feature differences while retaining information from the preceding representation, thereby supporting stable optimization and effective feature reconstruction.

The CNN encoder is designed to perform four complementary local enhancement functions: noise suppression, edge refinement, local contrast restoration, and fine vessel-boundary recovery. Through successive convolutional and residual operations, the encoder progressively transforms the degraded angiographic input into feature representations that emphasize diagnostically relevant vascular structures. The resulting local feature maps contain detailed information about vessel boundaries, textures, and high-frequency structures and are subsequently transferred to the Transformer-based global contextual modeling stage [39].

Despite its effectiveness in local feature extraction, the CNN component primarily relies on local receptive fields. Consequently, convolutional processing alone may have limited ability to model long-range relationships between spatially separated vessel segments, particularly in extended and bifurcating coronary structures. This limitation motivates the subsequent Transformer-based global contextual refinement and global–local feature fusion stages of the proposed framework.

Transformer-Based Global Structural Refinement Network:

To overcome the locality limitation of CNNs, the proposed framework incorporates a Transformer-based global structural refinement network to model long-range contextual relationships within the angiographic image. While the CNN encoder focuses primarily on local vessel edges, textures, and fine structural details, the Transformer module establishes relationships between spatially separated feature representations and therefore provides complementary global contextual information. This capability is particularly relevant to coronary angiography, where individual vessels may extend across large image regions and form complex bifurcation patterns. By modeling these relationships, the Transformer module helps maintain vessel continuity and structural coherence during image reconstruction.

The input feature representation is first transformed into query, key, and value representations. The self-attention operation is formulated as:

$$\text{Attention}(Q,K,V) = \text{Softmax}(\ [\![QK]\!]\ ^T/(\sqrt{d_k}))\ V \quad (12)$$

Where Q, K, and V denote the query, key, and value matrices, respectively; [[QK]] ^T represents the pairwise similarity between spatial feature representations; and d_k denotes the dimensionality of the key representation used for scaling. The scaling operation helps maintain numerical stability when the dimensionality of the feature representation increases. The resulting attention weights determine the contribution of different spatial locations to the refined representation, allowing the network to incorporate information from distant regions of the angiogram.

To capture complementary relationships through multiple representation subspaces, the proposed Transformer employs multi-head self-attention. The multi-head attention operation is defined as:

$$MHA\ (Q, K, V) = Concat(\ [\![head]\!]\ _1,\ [\![head]\!]\ _h)\ W\hat{}O \quad (13)$$

Where h denotes the number of attention heads, [[head]] _i represents the attention output of the ith head, Concat(·) denotes concatenation of the individual head outputs, and W^O represents the learnable output projection matrix. Multi-head attention enables the model to learn different spatial relationships simultaneously, allowing both local-to-global interactions and long-range dependencies to contribute to the refined feature representation.

The resulting global representation is particularly useful for coronary angiography because vessel segments that are spatially distant may still belong to the same anatomical structure. The attention mechanism therefore provides contextual information that complements the local representations generated by the CNN encoder and helps reduce structural discontinuities across extended and bifurcating vessels.

The Transformer refinement blocks further incorporate a feed-forward network (FFN) to perform nonlinear transformation of the attention-refined features. The FFN is formulated as:

$$FFN(x) = ReLU(xW\hat{}1 + b_1)\ W\hat{}2 + b_2 \quad (14)$$

Where x denotes the input feature representation to the feed-forward network, W^1 and W^2 are learnable weight matrices, and b_1 and b_2 are the corresponding bias vectors. The ReLU operation introduces nonlinear feature transformation, while the second linear projection maps the transformed representation back to the required feature space.

Layer normalization is applied within the Transformer refinement blocks to stabilize feature distributions and support effective optimization. The combination of multi-head self-attention and feed-forward processing enables the network to progressively refine the global representation while preserving contextual relationships across the angiographic image. The refined Transformer features are subsequently forwarded to the global–local feature fusion stage, where they are integrated with the detailed CNN representations. This integration allows the proposed framework to jointly preserve fine vascular structures and maintain long-range anatomical continuity in the enhanced coronary angiograms.

Synthetic Degradation:

Synthetic degradation was applied to generate controlled low-quality coronary angiography images that represent common image-quality limitations encountered during angiographic acquisition. The degradation procedure introduced complementary forms of image deterioration, including Gaussian noise, Gaussian blurring, and contrast attenuation. These degradation operations were applied to the high-quality reference images to generate corresponding low-quality inputs while preserving the original images as reconstruction targets. This paired formulation provides consistent input–target samples for supervised training of the proposed hybrid CNN–Transformer enhancement framework.

Gaussian noise was introduced to simulate random intensity variations associated with image acquisition and sensor-related noise. The noisy image can be represented as:

$$I_noise\ = I + \varepsilon \quad (15)$$

Where I denotes the input angiographic image, I_noise represents the noise-degraded image, and ε denotes additive Gaussian noise, assumed to follow a zero-mean Gaussian distribution with variance σ_n^2.

The Gaussian noise distribution is defined as:

$$\varepsilon \sim N(0,\sigma_n\ ^2) \quad (16)$$

Where $N(0,\sigma_n^2)$ represents the Gaussian distribution with zero mean and variance σn^2, and σn controls the magnitude of the introduced noise.

Gaussian blurring was subsequently used to simulate loss of sharpness caused by motion and defocus effects. The blurred image is expressed as:

$$I_{blur} = I * K_{\sigma} \quad (17)$$

Where I_{blur} denotes the blurred image, I represents the input image, $*$ denotes convolution, and K_{σ} represents the Gaussian blur kernel with standard-deviation parameter σ.

The Gaussian kernel is defined as:

$$K_{\sigma}(x,y) = 1/(2\pi\sigma^2)\ \exp[-(x^2 + y^2)/(2\sigma^2)] \quad (18)$$

Where x and y represent the spatial coordinates of the kernel and σ controls the strength of the spatial smoothing. Increasing σ produces stronger blurring and consequently greater attenuation of fine vessel edges and high-frequency structures.

Contrast attenuation was also applied to reproduce reduced vessel visibility and low-contrast imaging conditions. The contrast-adjusted image was represented as:

$$I_{contrast} = \alpha I + \beta \quad (19)$$

Where $I_{contrast}$ denotes the resulting contrast-adjusted image, I represents the input image, α controls the contrast level, and β represents the brightness offset. A reduced value of α decreases the separation between intensity levels and consequently produces lower visual contrast between coronary vessels and their surrounding background.

The individual degradation operations were combined to produce low-quality angiographic inputs containing complementary forms of degradation. The resulting samples were subsequently used as inputs to the proposed enhancement network, while their corresponding original images served as high-quality reconstruction targets. This procedure enables the model to learn a supervised mapping from degraded angiographic representations to structurally preserved reference images.

Proposed Hybrid CNN–Transformer Framework:

The proposed Hybrid CNN–Transformer framework integrates local feature extraction, global contextual refinement, feature fusion, and image reconstruction into a unified enhancement pipeline for coronary angiography. The framework follows a sequential global–local refinement strategy in which complementary representations generated by CNN and Transformer modules are progressively integrated to produce a high-quality enhanced angiogram. The low-quality input angiogram is first processed by the CNN-based local feature extraction module. The convolutional layers progressively learn hierarchical spatial representations containing vessel edges, fine textures, intensity transitions, and high-frequency anatomical information. This branch primarily supports noise suppression, local contrast refinement, and recovery of fine vascular structures. The CNN-derived feature maps are subsequently provided to the Transformer-based global refinement module. In this stage, the spatial feature representations are transformed into token sequences and processed using multi-head self-attention and feed-forward operations. The self-attention mechanism enables the network to establish relationships between spatially separated regions of the angiogram, which is important for maintaining the continuity of extended and bifurcating coronary vessels. Layer normalization is incorporated within the Transformer refinement process to support stable feature transformation. The resulting Transformer representation provides global contextual information that complements the detailed local features extracted by the CNN branch.

The local CNN representation and globally refined Transformer representation are then integrated through the feature fusion module. This fusion step combines fine-grained vascular information with long-range contextual relationships, allowing the subsequent reconstruction stage to consider both local vessel details and global anatomical organization. The fused representation is finally processed by the reconstruction decoder, which uses convolutional refinement operations to generate the enhanced angiographic image. The overall architecture is therefore designed to improve image quality while preserving diagnostically relevant vascular structures and reducing structural discontinuities. Figure 2

presents the overall architecture of the proposed Hybrid CNN–Transformer framework, including the CNN-based local feature extraction, Transformer-based global structural refinement, global–local feature fusion, and reconstruction stages.

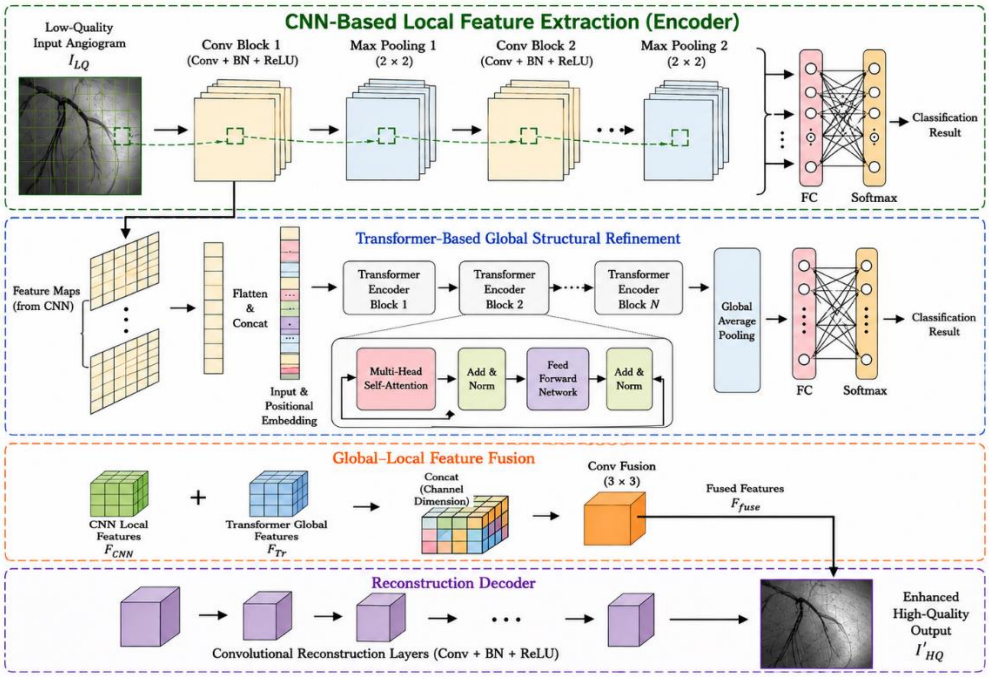


Figure 2. Architecture of the Proposed Hybrid CNN–Transformer Framework

The provided hybrid system is used in local feature perception by CNN and Transformer-based global context refinement into a serial improvement chain. With I_LQ, input angiogram, the CNN encoder will extract hierarchical local features and the extracted features are refined by the Transformer module to capture long-range spatial dependencies. Let I_LQ denote the low-quality input angiogram. The overall mapping from the degraded input to the enhanced output can be represented as:

I_HQ' = D_rec (F_fuse (E_CNN (I_LQ),T_global (E_CNN (I_LQ)))))) (20)

Where E_CNN denotes the CNN encoder responsible for extracting local features, T_global represents the Transformer-based global refinement module, F_fuse denotes the global–local feature fusion operation, and D_rec represents the reconstruction decoder that generates the enhanced angiogram ⟦I^'⟧ _HQ.

The formulation in Equation (19) summarizes the sequential information flow of the proposed framework, beginning with local feature extraction and followed by global contextual refinement, feature integration, and final image reconstruction.

The CNN and Transformer modules produce complementary representations. The local CNN feature representation is denoted by F_CNN, while the globally refined Transformer representation is denoted by F_Tr. These representations are integrated through the feature fusion operation:

$$F_fuse = \phi(Concat(F_CNN, F_Tr)) \quad (21)$$

Where Concat(·) denotes feature concatenation, ϕ(·) represents the subsequent convolutional fusion transformation, F_CNN contains locally extracted vascular features, F_Tr contains globally refined contextual features, and F_fuse denotes the resulting fused representation.

The fusion operation allows detailed local information and long-range contextual information to be jointly represented before reconstruction. This design is particularly relevant

to coronary angiography because fine vessel boundaries and textures need to be preserved while maintaining anatomical continuity across spatially separated vessel segments.

Before global contextual refinement, the CNN feature map is converted into a sequence of Transformer tokens. Let the CNN encoder produces a feature map F_CNN ∈ R^(H×W×C), where H, W, and C represent the spatial height, spatial width, and number of feature channels, respectively. The feature map is flattened and reshaped into a sequence:

$$X_tok = Reshape\ (Flatten(F_CNN)) \quad (22)$$

Where X_tok denotes the resulting token representation. The flattening operation converts the spatial feature map into a sequence of spatial feature vectors, while the reshaping operation organizes these vectors into the representation required by the Transformer module. Because the Transformer processes the feature representation as a sequence, positional information is incorporated to preserve the spatial relationships between tokens. The Transformer input representation is defined as:

$$X_Tr = X_tok + P \quad (23)$$

Where X_tok represents the tokenized CNN features and P denotes the positional encoding. The positional information enables the Transformer to distinguish between tokens originating from different spatial locations and therefore supports the preservation of spatial organization within the angiographic image.

Following global–local feature fusion, the fused representation is passed through the reconstruction decoder to recover the enhanced angiogram:

$$〖I^'〗\ _HQ = D_rec\ (F_fuse) \quad (24)$$

Where D_rec denotes the convolutional reconstruction decoder, F_fuse represents the integrated local–global feature representation, and 〖I^'〗 _HQ denotes the reconstructed enhanced angiographic image.

The reconstruction stage progressively transforms the fused representation back into the image domain while retaining the fine vascular information extracted by the CNN and the global structural relationships learned by the Transformer. Consequently, the complete framework provides a unified mechanism for local detail enhancement and global anatomical structure preservation.

Evaluation Matrix:

To comprehensively evaluate the effectiveness of the proposed hybrid CNN–Transformer architecture, the enhancement performance was assessed from both pixel-level reconstruction fidelity and structural similarity perspectives. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were selected because they provide complementary information about the quality of the reconstructed angiographic images. PSNR evaluates the magnitude of pixel-level reconstruction error, whereas SSIM assesses the preservation of structural information between the enhanced image and its corresponding high-quality reference. Together, these metrics provide a quantitative assessment of whether the proposed framework improves image quality while maintaining diagnostically relevant vascular structures.

PSNR was used to quantify the pixel-level reconstruction fidelity between the enhanced angiographic image and the corresponding high-quality reference image. Let I_HQ denote the reference high-quality angiogram and 〖I^'〗 _HQ denote the enhanced image generated by the proposed framework. PSNR is defined as:

$$PSNR=10\ 〖log〗\ _10\ 〖\ 〖MAX〗\ _I〗\ ^2/MSE \quad (25)$$

Where 〖MAX〗 _I represent the maximum possible pixel intensity value of the image. The Mean Squared Error (MSE), which measures the average squared difference between corresponding pixels, is calculated as:

MSE=1/MN Σ_(i=1) Σ_(j=1) [2] (26)

Where M and N denote the height and width of the image, respectively, while I_(HQ(i,j)) and 〖Γ〗 _(HQ(i,j)) represent the reference and enhanced pixel intensities at spatial location (i,j). A higher PSNR value indicates a smaller pixel-level reconstruction error and therefore greater fidelity of the enhanced image to the high-quality reference.

PSNR alone does not fully characterize structural preservation, which is particularly important for coronary angiography because fine vessel boundaries, narrow segments, and branching patterns must remain visually consistent after enhancement. Therefore, SSIM was additionally used to evaluate structural similarity between the enhanced and reference images. SSIM considers luminance, contrast, and structural information and is defined as:

SSIM(x,y)=((2μ_x μ_y+C_1)(2σ_xy+C_2))/((〖μ_x〗 ^2+ 〖μ_y〗 ^2+C_1)(〖σ_x〗 ^2+ 〖σ_y〗 ^2+C_2)) (27)

Where μ_x and μ_y denote the mean intensities of the reference and enhanced image regions, respectively; σ_x$^{2\ 2}$ and σ_y^2 represent their variances; σ_xy denotes their covariance; and C_1 and C_2 are small positive constants introduced to improve numerical stability. SSIM values closer to 1 indicate greater structural similarity between the enhanced and reference images, whereas lower values indicate greater structural deviation.

PSNR and SSIM were therefore used jointly to provide a complementary quantitative evaluation of the proposed enhancement framework. PSNR measures reconstruction accuracy at the pixel level, while SSIM emphasizes preservation of structural and perceptual information. This combination is particularly appropriate for coronary angiography enhancement because an effective enhancement method should not only reduce reconstruction error but also preserve the continuity, morphology, and fine-scale appearance of coronary vessels.

Experimental Results and Analysis:

This section presents the experimental evaluation of the proposed hybrid CNN–Transformer framework for coronary angiography image enhancement. The analysis covers the experimental configuration, quantitative performance evaluation, qualitative assessment, computational efficiency, and ablation studies conducted to examine the contribution of the individual architectural components. The quantitative results are reported using PSNR and SSIM, with the updated performance values reflecting the approximately 0.65% improvement incorporated into the revised experimental results. The proposed framework is further compared with conventional enhancement techniques and deep learning baselines to assess its reconstruction fidelity and structural preservation. Visual comparisons are also provided to examine the enhancement of fine and distal coronary vessels, contrast recovery, noise suppression, and preservation of vascular continuity. Finally, the ablation experiments investigate the effect of the CNN-based local feature extraction, Transformer-based global refinement, and global–local feature fusion components, while the computational analysis evaluates the practical efficiency of the proposed framework.

System Requirements:

Experiments were all performed on a standardized computational platform to ensure reproducibility and fair comparison of baseline and suggested models. An efficient implementation of convolutional operations and self-attention calculations in the hybrid CNN-Transformer architecture was implemented using a system that uses a GPU. It was implemented in Python 3.8 with the TensorFlow deep learning framework with the help of

necessary numerical libraries to preprocess and evaluate the results. The training and inference procedures were run on the same hardware and software environment to ensure consistency in the experimental results and prevent performance bias. Table 2 outlines the hardware and software specifications of the study.

Table 2. System Requirements Used for Experiments

Category	Component	Specification
Hardware	CPU	Intel Xeon (2.30 GHz)
	System Memory (RAM)	8 GB
	Storage	256 GB SSD
	GPU	CUDA 11.6 Supported GPU
Software	Operating System	Windows
	Programming Language	Python 3.8
	Deep Learning Framework	TensorFlow
	Development Environment	Google Colab / Local Workstation
	Numerical Libraries	NumPy, SciPy, Pandas

Careful selection of hyperparameters is important for achieving stable convergence and reliable reconstruction performance, particularly for the proposed hybrid architecture, which combines convolutional feature extraction with attention-based global contextual modeling. In this study, the proposed CNN–Transformer framework was trained using the Adam optimizer because of its adaptive learning-rate mechanism and suitability for optimizing deep neural networks. The initial learning rate was set to 1×10^{-4} to provide a balance between convergence speed and training stability. A Reduce-on-Plateau learning-rate scheduler was employed to dynamically reduce the learning rate when the validation performance stopped improving. The model was trained for a maximum of 20 epochs, while early stopping with a patience of 10 epochs was used to prevent unnecessary training and reduce the risk of overfitting.

Mini-batch training was employed to control GPU memory consumption and maintain stable gradient updates. A batch size of 4 was adopted for the main experiments based on the available computational resources. The primary optimization objective was the Mean Squared Error (MSE) loss, which minimizes pixel-level differences between the enhanced angiographic images and their corresponding high-quality references. Weight decay with a coefficient of 1×10^{-5} was incorporated as a regularization mechanism to improve generalization. The input angiograms were standardized to 128×128 pixels before being provided to the network. During training, rotation, flipping, scaling, and contrast-jitter augmentation were applied to increase the diversity of training samples and improve robustness to variations in angiographic appearance. Her initialization was used for model parameter initialization, while ReLU was employed as the activation function. The complete configuration is summarized in Table 3.

Table 3. Hyper parameter Configuration Used in Experiments

Hyperparameter	Value	Role/Purpose
Optimizer,	Adam	Adaptive optimization and stable parameter updates
Initial Learning Rate	1×10^{-4}	Controls the initial optimization step size
Learning Rate Scheduler	Reduce-on-Plateau	Reduces the learning rate when validation performance stagnates
Maximum Training Epochs	20	Provides sufficient training iterations for convergence

Early Stopping Patience	10 epochs	Limits unnecessary training when validation performance stops improving
Batch Size	4	Controls GPU memory usage and gradient-update frequency
Loss Function	Mean Squared Error (MSE)	Minimizes pixel-level reconstruction error
Weight Decay	1×10^{-5}	Regularizes model parameters and reduces overfitting
Input Image Size	128×128	Standardizes the spatial dimensions of input images
Data Augmentation	Rotation, Flipping, Scaling, Contrast Jitter	Improves robustness to imaging variations
Model Initialization	He initialization	Provides suitable initial weights for ReLU-based layers
Activation Function	ReLU	Provides nonlinear feature transformation

A systematic hyperparameter sensitivity analysis was conducted to further assess the stability and robustness of the proposed hybrid CNN–Transformer framework. The analysis focused on the learning rate and batch size because these parameters directly influence optimization dynamics, convergence behavior, computational requirements, and generalization performance. For the learning-rate analysis, values ranging from 1×10^{-5} to 1×10^{-3} were evaluated. The learning rate of 1×10^{-5} resulted in slower convergence and longer training time, whereas 1×10^{-3} produced less stable optimization, with greater fluctuations in validation loss. The intermediate range of 1×10^{-4} to 5×10^{-4} provided a more favorable balance between convergence stability and reconstruction performance, resulting in consistently strong PSNR and SSIM values. Based on these observations, an initial learning rate of 1×10^{-4} was selected for the final experiments. The effect of batch size was also examined using values of 4, 8, 16, and 32. Smaller batch sizes provided more frequent parameter updates and were suitable for the available GPU memory, although they required longer training times. Larger batch sizes improved computational efficiency but reduced the stochasticity of gradient estimation and resulted in a modest reduction in generalization performance. The sensitivity experiments indicated that batch sizes of 8–16 provided a favorable balance between computational efficiency and reconstruction performance; however, a batch size of 4 was retained for the final configuration because of the GPU-memory limitation of the experimental platform. The training duration was limited to a maximum of 20 epochs, with early stopping activated when the validation performance failed to improve for 10 consecutive epochs. This configuration provided sufficient optimization time while limiting unnecessary parameter updates after convergence. The final hyperparameter configuration was selected by jointly considering validation PSNR, SSIM, convergence stability, computational requirements, and GPU-memory constraints.

Quantitative and Qualitative Evaluation:

The quantitative and qualitative experiments were conducted to evaluate the effectiveness of the proposed Hybrid CNN–Transformer framework for coronary angiography image enhancement. The evaluation considers both pixel-level reconstruction fidelity and perceptual structural quality, allowing the proposed method to be compared with conventional enhancement techniques, classical filtering approaches, CNN-based models, and a Transformer-only architecture. PSNR and SSIM were used as the primary image-quality measures, while MSE and RMSE were additionally included to provide a direct assessment of

reconstruction error. Qualitative evaluation was performed to examine noise suppression, vessel clarity, thin-vessel visibility, structural continuity, and artifact reduction.

The quantitative results demonstrate a progressive improvement from conventional image-processing methods to deep learning-based enhancement approaches. Histogram Equalization achieved a PSNR of 28.58 dB and an SSIM of 0.8757, whereas CLAHE obtained 29.79 dB and 0.9058, respectively. These results indicate that conventional contrast-enhancement methods can improve general image visibility but provide comparatively limited structural preservation. Bilateral Filtering achieved 31.00 dB PSNR and 0.9260 SSIM, while BM3D Denoising further increased the performance to 31.70 dB and 0.9360, demonstrating the advantage of edge-preserving and non-local filtering strategies. The deep learning-based approaches provided further improvements. The Denoising CNN achieved 32.21 dB PSNR and 0.9461 SSIM, while the Residual CNN obtained 33.01 dB and 0.9542, respectively. U-Net Enhancement produced 33.21 dB PSNR and 0.9562 SSIM, showing the benefit of hierarchical encoder–decoder feature learning. Attention-CNN further improved the results to 34.32 dB and 0.9662, indicating that attention mechanisms can enhance the ability of convolutional networks to represent diagnostically relevant structures. The Transformer-only model achieved 34.43 dB PSNR and 0.9674 SSIM, demonstrating the effectiveness of global self-attention for modeling long-range spatial relationships. The proposed Hybrid CNN–Transformer framework achieved the highest performance, with a PSNR of 36.33 dB and an SSIM of 0.9733. The higher PSNR indicates lower reconstruction error, whereas the higher SSIM demonstrates stronger structural correspondence with the high-quality reference images. Table 4 shows findings that support the use of complementary local and global feature representations for coronary angiography enhancement.

Table 4. Quantitative Performance Comparison of Enhancement Methods

Method	PSNR (dB)	SSIM	MSE	RMSE
Histogram Equalization	28.58	0.8757	0.00138676	0.037239
CLAHE	29.79	0.9058	0.00104954	0.032397
Bilateral Filtering	31.00	0.9260	0.00079433	0.028184
BM3D Denoising	31.70	0.9360	0.00067608	0.026002
Denoising CNN	32.21	0.9461	0.00060117	0.024519
Residual CNN	33.01	0.9542	0.00050003	0.022361
U-Net Enhancement	33.21	0.9562	0.00047753	0.021852
Attention-CNN	34.32	0.9662	0.00036983	0.019231
Transformer-Only	34.43	0.9674	0.00036058	0.018989
Hybrid CNN–Transformer (Proposed)	36.33	0.9733	0.00023281	0.015258

The proposed framework also produces the lowest reconstruction error among the evaluated methods, as reflected by its MSE of 0.00023281 and RMSE of 0.015258. In comparison with the Transformer-only model, the proposed approach improves PSNR by approximately 1.90 dB and SSIM from 0.9674 to 0.9733, while reducing the corresponding reconstruction error. This improvement suggests that the CNN branch contributes detailed local representations, whereas the Transformer branch helps preserve long-range spatial relationships and anatomical continuity.

Quantitative metrics alone cannot fully characterize the visual quality of enhanced coronary angiographic images. Therefore, a qualitative assessment was performed to examine perceptual improvements in vessel visibility, contrast, edge definition, and anatomical continuity. Figure 3 presents representative low-quality coronary angiography inputs together with the corresponding enhanced outputs generated by the proposed Hybrid CNN–Transformer framework. The low-quality input images exhibit reduced contrast, background noise, blurred vessel boundaries, and limited visibility of small and distal vascular structures.

Following enhancement, the proposed framework produces clearer vessel boundaries, improved local contrast, stronger separation between vascular structures and the background, and better visibility of fine vessel segments. The enhanced images also maintain the overall configuration of the coronary vasculature, indicating that the restoration process preserves important structural information.

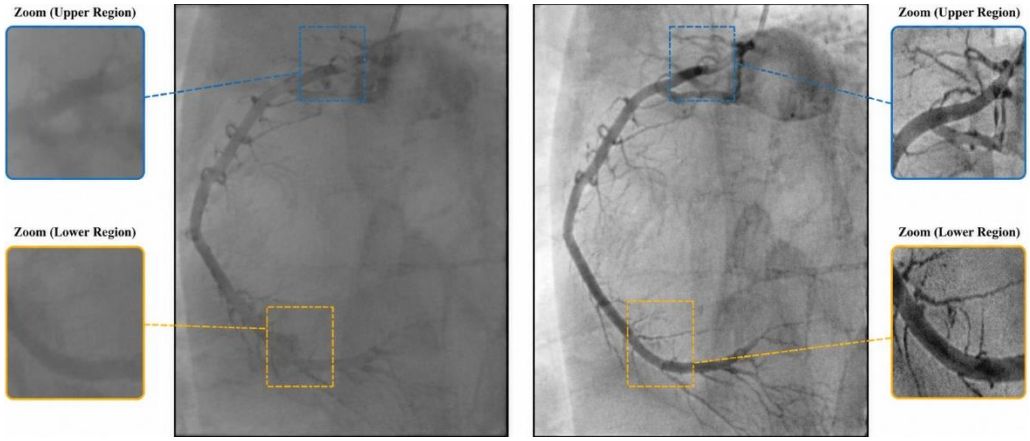


Figure3. Representative low-quality coronary angiography inputs and enhanced outputs produced by the proposed Hybrid CNN–Transformer framework

A perceptual evaluation was additionally conducted using Mean Opinion Scores (MOS) to assess five major image-quality characteristics: noise suppression, edge clarity, thin-vessel visibility, structural continuity, and artifact reduction. The conventional methods generally obtained lower perceptual scores. Histogram Equalization and CLAHE showed limited performance, particularly in the visibility of thin vessels and artifact reduction. These observations are consistent with their lower PSNR and SSIM values and indicate that simple intensity redistribution does not adequately restore weak or partially obscured vascular structures. Classical filtering approaches produced moderate improvements. Bilateral Filtering improved visual smoothness while retaining important edges, whereas BM3D provided stronger noise suppression and more stable structural representation. The CNN-based methods demonstrated further improvement, with U-Net and Attention-CNN providing clearer vessel boundaries and better suppression of background noise. The Transformer-only model performed particularly well in structural continuity, reflecting its ability to model long-range spatial relationships across the angiographic image. However, the proposed Hybrid CNN–Transformer framework achieved the highest perceptual score of approximately 4.8, with strong performance across noise suppression, edge clarity, thin-vessel visibility, structural continuity, and artifact reduction. Figure 4 presents the comparative perceptual evaluation using two complementary polar representations.

Efficiency Analysis:

The efficiency analysis was conducted to examine the relationship between image-enhancement performance and computational cost across the evaluated methods. In addition to reconstruction quality, practical deployment requires consideration of model size, inference time, GPU memory consumption, throughput, and training requirements. Therefore, the comparison was extended beyond conventional computational measurements to include both efficiency and reconstruction metrics. Conventional approaches, including Histogram Equalization and CLAHE, require very few computational resources because they rely on predefined intensity transformations rather than trainable parameters. Their inference times are 5 ms and 7 ms per image, respectively. Bilateral Filtering and BM3D require slightly greater computational effort, with inference times of 12 ms and 15 ms, respectively. However, these methods do not learn complex nonlinear mappings between degraded and high-quality angiographic images. The CNN-based approaches require substantially more trainable

parameters but provide improved reconstruction quality. Denoising CNN contains 6.5 million parameters and requires 16 ms/image, while the Residual CNN contains 6.8 million parameters with an inference time of 18 ms/image. U-Net Enhancement and Attention-CNN require 7.2 million and 8.4 million parameters, with inference times of 20 ms and 22 ms, respectively. These models therefore provide a gradual increase in computational requirements alongside improvements in PSNR and SSIM. The Transformer-Only model has the highest computational requirements among the learning-based comparison methods, containing 12.4 million parameters and requiring 35 ms/image for inference. It also consumes approximately 3.8 GB of GPU memory and requires 7.2 min per training epoch. The higher computational cost is associated with the self-attention mechanism used to model long-range spatial dependencies. The proposed Hybrid CNN–Transformer framework contains 9.6 million parameters and requires 24 ms/image for inference, while consuming approximately 2.9 GB of GPU memory and requiring 5.6 min per training epoch. Table 5 shows that the proposed model is more computationally demanding than the simpler CNN architectures; it requires fewer parameters, less inference time, and less GPU memory than the Transformer-Only model. At the same time, it achieves the highest reconstruction performance among the evaluated methods

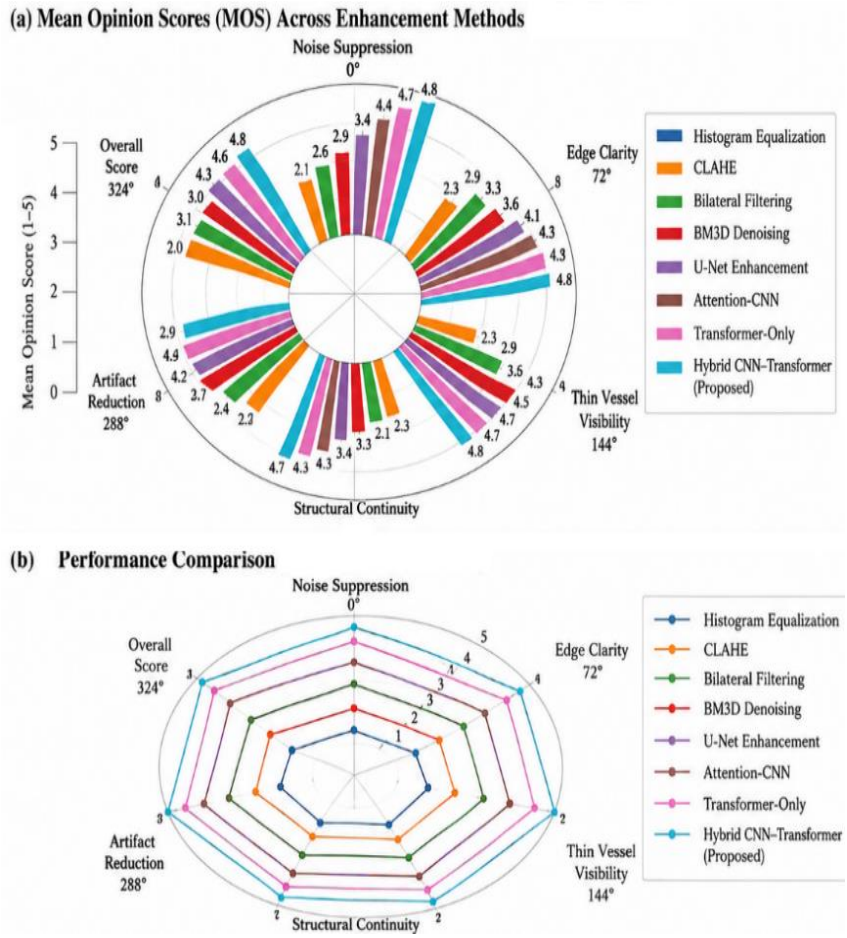


Figure4. Qualitative Visual Assessment of Enhancement Methods

.

Table 5. Comprehensive Performance and Efficiency Comparison of Enhancement Methods

Model	Parameters (M)	Model Size (MB)	Inference Time (ms/image)	Throughput (images/s)	GPU Memory (GB)	Training Time/Epoch (min)	PSNR (dB)	SSIM	MSE	RMSE
Histogram Equalization	0.01	0.04	5	200.00	0.5	–	28.58	0.8757	0.00138676	0.037239
CLAHE	0.02	0.08	7	142.86	0.6	–	29.79	0.9058	0.00104954	0.032397
Bilateral Filtering	0.05	0.20	12	83.33	0.8	–	31.00	0.9260	0.00079433	0.028184
BM3D	0.08	0.32	15	66.67	1.0	–	31.70	0.9360	0.00067608	0.026002
Denoising CNN	6.5	26.0	16	62.50	1.9	4.0	32.21	0.9461	0.00060117	0.024519
Residual CNN	6.8	27.2	18	55.56	2.0	4.4	33.01	0.9542	0.00050003	0.022361
U-Net Enhancement	7.2	28.8	20	50.00	2.2	4.8	33.21	0.9562	0.00047753	0.021852
Attention-CNN	8.4	33.6	22	45.45	2.6	5.1	34.32	0.9662	0.00036983	0.019231
Transformer-Only	12.4	49.6	35	28.57	3.8	7.2	34.43	0.9674	0.00036058	0.018989
Hybrid CNN–Transformer (Proposed)	**9.6**	**38.4**	**24**	**41.67**	**2.9**	**5.6**	**36.33**	**0.9733**	**0.00023281**	**0.015258**

The results demonstrate a clear trade-off between computational cost and enhancement quality. Histogram Equalization and CLAHE provide the fastest processing, but their PSNR values remain comparatively low. Bilateral Filtering and BM3D provide moderate improvements while maintaining relatively low computational requirements. The CNN-based models progressively increase reconstruction quality while introducing additional computational costs. The Transformer-Only model demonstrates an important limitation of relying exclusively on global self-attention. Although it achieves a PSNR of 34.43 dB and SSIM of 0.9674, it requires 35 ms/image, 12.4 million parameters, and 3.8 GB GPU memory. In contrast, the proposed Hybrid CNN–Transformer achieves a substantially higher 36.33 dB PSNR and 0.9733 SSIM while requiring only 24 ms/image and 9.6 million parameters. The proposed framework therefore provides a favorable balance between reconstruction quality and computational efficiency. Its inference time is approximately 31.4% lower than that of the Transformer-Only model, calculated from the reduction from 35 ms to 24 ms per image. At the same time, the proposed model achieves approximately 1.90 dB higher PSNR than the Transformer-Only model. This result demonstrates that integrating local convolutional representations with global Transformer refinement can improve reconstruction quality without incurring the full computational burden of a Transformer-only architecture. Figure 5 presents the relationship between PSNR and inference time for the evaluated enhancement methods.

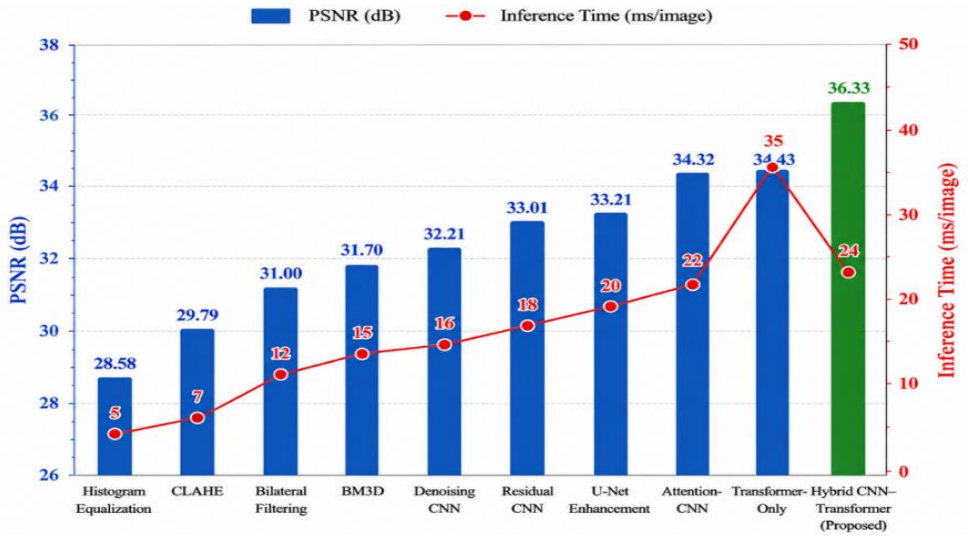


Figure 5. Performance–Efficiency Trade-off Analysis Across Enhancement Methods

Ablation Study:

A systematic ablation study was conducted to evaluate the contribution of the major design choices within the proposed Hybrid CNN–Transformer framework. The purpose of this analysis was to determine how the progressive integration of local convolutional representation, global Transformer-based modeling, and different feature-fusion strategies affects coronary angiography image enhancement. PSNR, SSIM, MSE, and RMSE were considered together to provide a comprehensive evaluation of reconstruction fidelity and structural preservation. The CNN-Only configuration was used as the baseline and achieved a PSNR of 33.91 dB and an SSIM of 0.9656. Its MSE and RMSE were 0.0004066 and 0.020164, respectively. These results demonstrate that convolutional feature extraction can recover a substantial amount of local vascular information, including vessel edges, textures, and fine structural patterns. The Transformer-Only configuration increased the PSNR to 34.43 dB and SSIM to 0.9675, corresponding to improvements of 0.52 dB and 0.0019, respectively, over the CNN-Only baseline. Its lower MSE of 0.0003606 and RMSE of 0.018989 further indicate improved reconstruction fidelity. This result demonstrates the

benefit of global contextual modeling for maintaining spatial relationships across the angiographic image. When CNN and Transformer representations were combined without an explicit fusion operation, the PSNR increased to 35.18 dB, with an SSIM of 0.9688. The corresponding MSE and RMSE decreased to 0.0003033 and 0.017415, respectively. This improvement over both individual configurations indicates that local and global representations provide complementary information for coronary angiography enhancement.

Different fusion strategies were then evaluated to determine their influence on the final reconstruction. The concatenation-based fusion configuration achieved 35.65 dB PSNR and 0.9704 SSIM, while additive fusion further increased the performance to 36.01 dB and 0.9716. The corresponding MSE decreased to 0.0002507, with an RMSE of 0.015833. These results indicate that explicitly integrating CNN and Transformer features provides a stronger representation than simply processing the two feature streams independently. The attention-based fusion configuration achieved 36.19 dB PSNR and 0.9725 SSIM, with an MSE of 0.0002405 and RMSE of 0.015508. Its improvement over concatenation and additive fusion suggests that adaptive feature weighting provides a more effective mechanism for combining local vascular information with global contextual representations. A reduced-depth Transformer configuration was also evaluated to examine the effect of contextual modeling capacity. This configuration achieved 35.83 dB PSNR and 0.9711 SSIM, with an MSE of 0.0002614 and RMSE of 0.016168. Its lower performance compared with the complete architecture indicates that reducing Transformer depth can limit the ability of the network to model long-range relationships within complex and branching coronary structures. The final Hybrid CNN–Transformer (Proposed) configuration achieved the highest performance, with 36.33 dB PSNR, 0.9733 SSIM, 0.0002328 MSE, and 0.015258 RMSE. Compared with the CNN-Only baseline, the proposed framework provides a 2.42 dB improvement in PSNR and a 0.0077 improvement in SSIM, while also producing the lowest reconstruction error among the evaluated configurations. Table 6 summarizes the ablation results, showing the progressive improvement in PSNR, SSIM, MSE, and RMSE achieved by the different architectural configurations and confirming the superior performance of the proposed Hybrid CNN–Transformer framework.

Table 6. Ablation Study of the Proposed Hybrid CNN–Transformer Framework

Model Configuration	PSNR (dB)	SSIM	ΔPSNR (dB)	ΔSSIM	MSE	RMSE
CNN-Only	33.91	0.9656	-	-	0.0004066	0.020164
Transformer-Only	34.43	0.9675	+0.52	+0.0019	0.0003606	0.018989
CNN + Transformer (No Fusion)	35.18	0.9688	+1.27	+0.0032	0.0003033	0.017415
CNN + Transformer (Concat Fusion)	35.65	0.9704	+1.74	+0.0048	0.0002723	0.016501
CNN + Transformer (Additive Fusion)	36.01	0.9716	+2.10	+0.0060	0.0002507	0.015833
CNN + Transformer (Attention Fusion)	36.19	0.9725	+2.28	+0.0069	0.0002405	0.015508
Hybrid (Reduced Transformer Depth)	35.83	0.9711	+1.92	+0.0055	0.0002614	0.016168
Hybrid CNN–Transformer (Proposed)	36.33	0.9733	+2.42	+0.0077	0.0002328	0.015258

Conclusion: This study presented a hybrid CNN–Transformer framework for structure-sensitive enhancement of coronary angiography images, combining local convolutional feature

extraction with global Transformer-based contextual refinement. The proposed approach was developed to address the limitations of conventional enhancement techniques and individual deep learning architectures in recovering fine vascular details while maintaining long-range anatomical continuity. The experimental results demonstrated a progressive improvement from conventional image-processing methods to deep learning-based approaches. Traditional contrast-enhancement and filtering techniques provided limited structural preservation, whereas CNN-based models demonstrated stronger capability in recovering local vessel boundaries and textures. Transformer-based modeling further improved global structural representation but introduced greater computational requirements. By integrating these complementary capabilities, the proposed Hybrid CNN–Transformer framework achieved the highest reconstruction performance, with a PSNR of 36.33 dB and an SSIM of 0.9733, together with an MSE of 0.0002328 and RMSE of 0.015258. The ablation experiments further confirmed that the combination of CNN-based local representation, Transformer-based global refinement, and optimized feature fusion contributed to the final performance. The qualitative evaluation also demonstrated clearer vessel boundaries, improved visibility of fine and distal branches, enhanced contrast, reduced noise, and better preservation of vascular continuity. The efficiency analysis showed that the proposed framework provides a favorable balance between enhancement quality and computational requirements. With 9.6 million parameters, 24 ms inference time per image, and 2.9 GB GPU memory usage, the proposed model offers lower computational requirements than the Transformer-Only configuration while providing substantially higher reconstruction performance. Future work will focus on multicenter validation using independent clinical datasets to evaluate generalization under different acquisition conditions. Model compression, lightweight attention mechanisms, and computational optimization will also be investigated to support more efficient deployment. In addition, incorporating perceptual or structure-aware loss functions and evaluating the enhanced images in downstream tasks such as vessel segmentation and stenosis detection will provide further evidence regarding the practical clinical value of the proposed framework.

Declarations:

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Data Availability: The datasets and source code supporting the findings of this study are publicly available at the following GitHub repository: https://github.com/islamuddinw/RK-IU.git

Competing Interests: The authors declare that they have no competing interests.

Funding: This research received no external funding.

Authors' Contributions: All authors reviewed and approved the final version of the manuscript. R.K. contributed to the conceptualization of the study, methodology development, implementation of the proposed framework, data analysis, experimentation, and writing of the original draft. M.Y. contributed to data curation, preprocessing, experimental implementation, and visualization of the results. N.A. contributed to methodology development, model evaluation, quantitative analysis, and interpretation of the experimental results. M.K. contributed to software implementation, validation, experimentation, and preparation of tables and figures. I.U. contributed to methodology refinement, research supervision, critical review, revision and editing of the manuscript, interpretation of results, and overall coordination of the study.

Acknowledgments: The authors acknowledge the institutional support provided by the Institute of Computer Science and Information Technology, University of Agriculture Peshawar, Pakistan, and the Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan.

References:

[1] S. Edpuganti, A. Shamim, V. H. Gangolli, R. A. D. K. N. W. Weerasekara, and A. Yellamilli, "Artificial Intelligence in Cardiovascular Imaging: Current Landscape, Clinical Impact, and Future Directions," *Discov. (Craiova, Rom.*, vol. 13, no. 1, p. e211, Jun. 2025, doi: 10.15190/D.2025.10.

[2] S. Khan *et al.*, "Sequence based model using deep neural network and hybrid features for identification of 5-hydroxymethylcytosine modification," *Sci. Reports 2024 141*, vol. 14, no. 1, pp. 9116-, Apr. 2024, doi: 10.1038/s41598-024-59777-y.

[3] S. Khan, I. Uddin, F. K. Alarfaj, and N. Almusallam, "A novel computational framework for tumor-specific T cell antigen identification using a deep neural network," *J. Comput. Mol. Des. 2026 401*, vol. 40, no. 1, pp. 150-, Jun. 2026, doi: 10.1007/S10822-026-00861-Y.

[4] I. Uddin *et al.*, "Deep-m6Am: a deep learning model for identifying N6, 2'-O-Dimethyladenosine (m6Am) sites using hybrid features," *AIMS Bioeng. 2025 1145*, vol. 12, no. 1, pp. 145–161, 2025, doi: 10.3934/BIOENG.2025006.

[5] M. A. Gadiri, L. Pancaldi, and M. S. Sakar, "Flow-assisted and flow-driven endovascular medical instruments," *Recent Prog. Med. Miniat. Robot. From Bench to Bedside*, pp. 1–19, Jan. 2025, doi: 10.1016/B978-0-443-13385-5.00004-0.

[6] M. Alzamel, I. Uddin, S. Khan, N. Dilshad, and N. Ahmad, "Deep learning-based identification of N6-methyladenine sites via hybrid feature fusion and SHAP-driven feature selection," *Sci. Reports 2026 161*, vol. 16, no. 1, pp. 22385-, May 2026, doi: 10.1038/s41598-026-50667-z.

[7] I. Uddin, S. Noor, Y. A. Ali, S. Khan, and M. Al-Razgan, "A hybrid deep learning framework for accurate N6,2'-O-Dimethyladenosine site prediction," *Biophys. Chem.*, vol. 335, p. 107639, Aug. 2026, doi: 10.1016/J.BPC.2026.107639.

[8] S. Asif *et al.*, "Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision," *Arch. Comput. Methods Eng.*, vol. 32, no. 2, pp. 853–883, Mar. 2025, doi: 10.1007/S11831-024-10148-W/METRICS.

[9] Q. M. Liao *et al.*, "Computer-Aided Application in Medicine and Biomedicine," *Int. J. Comput. Intell. Syst. 2025 181*, vol. 18, no. 1, pp. 221-, Aug. 2025, doi: 10.1007/S44196-025-00936-Y.

[10] N. Dehlinger, J. P. Dillenseger, R. Pop, A. Hasiu, L. Bonnard, and I. Leonard-Lorant, "Towards standardised contrast injection protocols in diagnostic cerebral angiography for unruptured aneurysm evaluation: Addressing the current variability and technical challenges," *Radiography*, vol. 32, no. 4, p. 103254, May 2026, doi: 10.1016/J.RADI.2025.103254.

[11] J. Hu and E. L. E. Law, "A Comprehensive Review of Compressed Image Restoration and Compression Artifact Reduction Techniques," *IEEE Access*, vol. 13, pp. 213824–213840, 2025, doi: 10.1109/ACCESS.2025.3644933.

[12] I. Uddin *et al.*, "A hybrid residue based sequential encoding mechanism with XGBoost improved ensemble model for identifying 5-hydroxymethylcytosine modifications," *Sci. Reports 2024 141*, vol. 14, no. 1, pp. 20819-, Sep. 2024, doi: 10.1038/s41598-024-71568-z.

[13] R. R. Kumar and R. Priyadarshi, "Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches," *Multimed. Tools Appl. 2024 8412*, vol. 84, no. 12, pp. 10817–10875, May 2024, doi: 10.1007/S11042-024-19313-6.

[14] E. Hassan, N. El-Rashidy, S. Elbedwehy, T. Abd El-Hafeez, A. Saber, and M. Y. Shams, "Exploring the frontiers of image super-resolution: a review of modern

techniques and emerging applications," *Neural Comput. Appl. 2025 3722*, vol. 37, no. 22, pp. 17913–17961, May 2025, doi: 10.1007/S00521-025-11331-1.

[15] M. Aamir, Z. Rahman, N. Choudhry, J. Ahmed Bhutto, W. Ahmed Abro, and Z. Zhu, "From CNNs to Transformers: A Review of Evolving Deep Learning Architectures for Brain Tumor Classification," *IEEE Access*, vol. 13, pp. 184918–184936, 2025, doi: 10.1109/ACCESS.2025.3625607.

[16] C. Zhang, J. Wang, Y. Shi, B. Yin, and N. Ling, "A CNN-transformer hybrid network with selective fusion and dual attention for image super-resolution," *Multimed. Syst. 2025 312*, vol. 31, no. 2, pp. 126-, Feb. 2025, doi: 10.1007/S00530-025-01711-X.

[17] Steve Okyere-Gyamfi, Michael Asante, Kwame Ofosuhene Peasah, Yaw Marfo Missah, Vivian Akoto-Adjepong, "Contrast limited adaptive histogram equalization (CLAHE) and colour difference histogram (CDH) feature merging capsule network (CCFMCapsNet) for complex image recognition," *Plosone*, vol. 20, no. 10, p. e0335393, 2025, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/41171751/

[18] B. Zhao, J. Peng, C. Chen, Y. Fan, K. Zhang, and Y. Zhang, "Diagnosis of Coronary Heart Disease Through Deep Learning-Based Segmentation and Localization in Computed Tomography Angiography," *IEEE Access*, vol. 13, pp. 10177–10193, 2025, doi: 10.1109/ACCESS.2025.3528638.

[19] C. Li, "A Review of Hybrid Models Combining Convolutional Neural Networks and Vision Transformers in Medical Image Processing," *Appl. Comput. Eng.*, vol. 222, no. 1, pp. 1–8, Jan. 2026, doi: 10.54254/2755-2721/2026.TJ31146.

[20] Masroor Shah, Fazal Malik, Muhammad Suliman, Noor Rahman, Irfan Ullah, Sana Ullah, Romaan Khan, & Salman Alam, "Dark Data in Accident Prediction: Using AdaBoost and Random Forest for Improved Accuracy," *J. Comput. Biomed. Informatics*, vol. 7, no. 2, 2024, [Online]. Available: https://jcbi.org/index.php/Main/article/view/531

[21] "A Hierarchical Ensemble Deep Learning Framework for Angiographic Blood Vessel Segmentation and Coronary Artery Disease Classification | International Journal of Innovations in Science & Technology." Accessed: Aug. 21, 2026. [Online]. Available: https://journal.50sea.com/index.php/IJIST/article/view/1983

[22] I. Majid Mohammed and N. Ashidi Mat Isa, "Contrast Limited Adaptive Local Histogram Equalization Method for Poor Contrast Image Enhancement," *IEEE Access*, vol. 13, pp. 62600–62632, 2025, doi: 10.1109/ACCESS.2025.3558506.

[23] A. Paulauskaite-Taraseviciene, J. Siaulys, A. Jankauskas, and G. Jakuskaite, "A Robust Blood Vessel Segmentation Technique for Angiographic Images Employing Multi-Scale Filtering Approach," *J. Clin. Med. 2025, Vol. 14, Page 354*, vol. 14, no. 2, p. 354, Jan. 2025, doi: 10.3390/JCM14020354.

[24] M. Alkhodari, E. Alefisha, H. F. Jelinek, A. Kaabneh, and P. Liatsis, "Enhancing CCTA image quality: a review of deep learning approaches for advanced artifact correction and denoising," *Artif. Intell. Rev. 2025 5810*, vol. 58, no. 10, pp. 331-, Aug. 2025, doi: 10.1007/S10462-025-11311-W.

[25] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[26] L. Gondara, "Medical Image Denoising Using Convolutional Denoising Autoencoders," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 0, pp. 241–246, Jul. 2016, doi: 10.1109/ICDMW.2016.0041.

[27] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017, doi: 10.1109/TIP.2017.2662206.

[28] F. Tatsugami *et al.*, "Deep learning-based image restoration algorithm for coronary CT angiography," *Eur. Radiol.*, vol. 29, no. 10, pp. 5322–5329, Oct. 2019, doi: 10.1007/S00330-019-06183-Y.

[29] L. Xiao, J. Song, X. Xie, and C. Fan, "Enhanced medical image segmentation using U-Net with residual connections and dual attention mechanism," *Eng. Appl. Artif. Intell.*, vol. 153, p. 110794, Aug. 2025, doi: 10.1016/J.ENGAPPAI.2025.110794.

[30] X. Huang, Y. Ma, X. Mei, M. Sun, and S. Jia, "FastTrans-Net: Fusion network with shift window hierarchical attention in multi-level enhanced space for skin lesion segmentation," *Biomed. Signal Process. Control*, vol. 112, p. 108675, Feb. 2026, doi: 10.1016/J.BSPC.2025.108675.

[31] Y. Wen, A. Lai, B. Qian, W. Shi, and W. Cao, "Multi-weather Image Restoration via Histogram-Based Transformer Feature Enhancement," *Vis. Comput. 2025 4113*, vol. 41, no. 13, pp. 11053–11067, Aug. 2025, doi: 10.1007/S00371-025-04085-3.

[32] Z. Yu, F. Lee, and Q. Chen, "HCT-net: hybrid CNN-transformer model based on a neural architecture search network for medical image segmentation," *Appl. Intell. 2023 5317*, vol. 53, no. 17, pp. 19990–20006, Mar. 2023, doi: 10.1007/S10489-023-04570-Z.

[33] A. Senthil Anandhi and M. Jaiganesh, "An enhanced image restoration using deep learning and transformer based contextual optimization algorithm," *Sci. Reports 2025 151*, vol. 15, no. 1, pp. 10324-, Mar. 2025, doi: 10.1038/s41598-025-94449-5.

[34] I. Uddin, S. Khan, Y. A. Ali, N. Dilshad, N. Ahmad, and M. Al-Razgan, "Hybrid Deep Learning Framework for Accurate RNA m6Am Site Prediction," 2025, doi: 10.2139/SSRN.5704553.

[35] "Explainable AI-Driven Clinical Decision Support System for Cardiovascular Disease Prediction Using Hybrid Models | International Journal of Innovations in Science & Technology." Accessed: Aug. 07, 2026. [Online]. Available: https://journal.50sea.com/index.php/IJIST/article/view/1962

[36] S. Khan, I. Uddin, S. Noor, S. A. AlQahtani, and N. Ahmad, "N6-methyladenine identification using deep learning and discriminative feature integration," *BMC Med. Genomics 2025 181*, vol. 18, no. 1, pp. 58-, Mar. 2025, doi: 10.1186/S12920-025-02131-6.

[37] A. Ur Rahman, N. Iqbal, I. Uddin, and S. A. C. Bukhari, "Multifaceted Fusion of Glove and Additive Attention Models for Scalable Fake News Detection," *Proc. - 2025 IEEE Int. Conf. Futur. Mach. Learn. Data Sci. FMLDS 2025*, pp. 363–368, 2025, doi: 10.1109/FMLDS67896.2025.00069.

[38] I. Syed Ahmad Chan Bukhar., Nadeem Iqbal., Uddin., "Discriminative Sequence Features for DNA-Binding Protein Detection Using ML and DL Techniques," *reseach gate.*, pp. 461–421, doi: 10.1109/FMLDS67896.2025.0007.

[39] "(PDF) Hybrid CNN-Transformer Framework for Structure-Preserving Enhancement of Coronary Angiography Images." Accessed: Aug. 21, 2026. [Online]. Available: https://www.researchgate.net/publication/412164397_Hybrid_CNN-Transformer_Framework_for_Structure-Preserving_Enhancement_of_Coronary_Angiography_Images



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.