

Efficient Medical Image Compression using a CNN and Discrete Wavelet Transform

Dalila CHIKHAOUI¹, Mohammed BELADGHAM¹, and Mohamed BEN AISSA ²

⁽¹⁾ LTIT Laboratory, Tahri Mohammed University, Bechar, Algeria

⁽²⁾ LTIT Laboratory, Abou Bekr Belkaid University, Tlemcen, Algeria

Email of the corresponding author: chikhaoui.dalila@univ-bechar.dz

Received: 25 March 2026

Accepted for publication: 22 September 2026

Published: 12 October 2026

Abstract—This study presents a computationally efficient framework for medical image compression that integrates a low-complexity convolutional neural network (CNN) with the discrete wavelet transform (DWT). Unlike conventional CNN–DWT hybrid approaches that embed wavelet processing within large end-to-end architectures, the proposed method maps the input image to a compact three-channel latent representation that is encoded using the CDF 9/7 wavelet, as employed in JPEG 2000. This design avoids custom coding of high-dimensional feature tensors while facilitating compatibility with established image-coding workflows. The CNN reduces spatial redundancy while preserving essential image information, whereas the DWT exploits frequency-domain sparsity and multiresolution characteristics to improve compression efficiency. A two-stage training strategy is adopted: the encoder–decoder is first pretrained without the codec and subsequently fine-tuned with real non-differentiable quantization in the compression loop, enabling optimization against the actual coding process. Experiments on a publicly available low-grade glioma (LGG), magnetic resonance imaging (MRI) dataset demonstrate compression ratios of up to 40:1, with a peak signal-to-noise ratio (PSNR) of 38.6 dB and a multiscale structural similarity index measure (MS-SSIM) of 98.84%. The proposed CNN contains approximately 0.26 million trainable parameters and can be trained on a single GPU, demonstrating its potential for resource-constrained medical image storage and transmission applications.

Keywords— Image compression, convolutional neural network (CNN), discrete wavelet transform (DWT), medical image

I. INTRODUCTION

Medical imaging is integral to modern healthcare and supports disease detection, diagnosis, treatment planning, and longitudinal follow-up across virtually every clinical specialty. Modalities such as X-ray, computed tomography (CT), single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) generate enormous volumes of image data that are routinely archived in Picture Archiving and Communication Systems (PACS) and Hospital Information Systems (HIS) for later consultation and research[1]. These systems require diverse strategies and mechanisms for efficient data processing and transmission, posing challenges to the limited storage and bandwidth resources. Several compression techniques and standards have been proposed to tackle these issues [2]. Image compression methods are generally divided into lossless and lossy

categories. Presently, the technology for compressing medical images is mainly split into lossless and lossy compression techniques [3]. In the realm of lossless compression for medical images, certain methods facilitate the efficient storage and transmission of visualization data [4]. Techniques have been applied to compress volumetric data, such as the work [5], which leveraged the interslice correlation between voxels and used modern adaptive and context-based reversible methods to achieve efficient compression and storage. Although images reconstructed using lossless compression technology exhibit no distortion and have a good reconstruction effect, the compression rate is relatively low, and this technology currently cannot meet the demands of compressing the vast amounts of medical images [3].

Currently, lossy compression in medical imaging is mainly divided into two broad categories: traditional

algorithm-based techniques and deep learning-based methods. Traditional compression algorithms, like JPEG, JPEG2000, and SPIHT, use fixed mathematical transformations—such as the Discrete Cosine Transform (DCT) or Wavelet Transform—to reduce spatial and frequency redundancies in image data [3], [6]. For example, a region of interest (ROI)-based medical image compression framework was proposed for effective image transmission without compromising diagnostic quality [7]. The ROI is compressed using a combination of Block-based Burrows-Wheeler Transform (BWT) and Move-to-Front (MTF) transformation, ensuring minimal loss of critical information. The non-ROI (NROI) region is encoded using a hybrid fractal compression technique, which surpasses traditional fractal methods in terms of efficiency and quality. In another study, a compression technique employing wavelet transform encoding was introduced [8]. The approach compressed medical images using Discrete Cosine Transform combined with Set Partitioning in Hierarchical Trees (SPIHT), and the NROI is handled using Discrete Wavelet Transform (DWT) followed by merging-based Huffman encoding. Additionally, a combined method for simultaneous compression and encryption of medical images was developed to ensure both data efficiency and security [9].

Ballé et al. [10] introduced the first end-to-end optimized CNN-based image codec, combining a generalized divisive-normalization nonlinear operator with a uniform scalar quantizer and an entropy model. This work demonstrated that learned transforms can outperform JPEG and JPEG2000 at comparable bitrates on natural images. Subsequent work [11], [12] scaled the network capacity, replaced pixel-space losses with perceptual and adversarial objectives, and improved entropy modeling, establishing the modern template of transform-quantize-entropy pipelines. In the medical imaging domain, several CNN-based codecs have been proposed. The work in [13] combined lifted wavelets with a CNN to compress MRI and CT images, achieving performance comparable to JPEG-LS and JPEG2000 in both lossless and lossy modes. Reference [14] proposed a fully convolutional autoencoder tailored to medical images, showing faster convergence and improved reconstruction quality over JPEG-based baselines. A variational approach [3] balanced rate and distortion via a learned latent prior and reported lower distortion than both JPEG2000 and earlier learned codecs on several medical modalities.

A major advance in learned compression was the introduction of hyperprior networks by Ballé et al. [11], which model the marginal distribution of the encoder latent using a secondary “hyper” network. The hyper-network emits a per-element Gaussian scale parameter transmitted as side information, enabling far more accurate entropy coding than uniform or factorized priors. Minnen et al. [15] further combined hyperpriors with a learned autoregressive context model, and this framework remained the strongest published baseline for natural images throughout the early 2020s. More recent extensions [16], [17] add discretized Gaussian mixtures, channel-wise contexts, and stochastic gating to scale this template to higher resolutions. In the medical imaging domain, hyperprior-based codecs—including adapted versions of Ballé’s model [3]—have demonstrated substantial improvements over earlier CNN codecs.

Hybrid frameworks attempt to retain the interpretability and standardization of wavelet codecs while adding the

representational power of CNNs. Early efforts used CNNs as post-processors that clean artifacts produced by JPEG2000 [18]. More tightly integrated designs, such as [13], embed the DWT as a structural prior within an otherwise learned end-to-end model. One hybrid approach [18], termed SDWTCNN, combines Singular Value Decomposition (SVD) for ROI feature extraction, DWT for ROI encoding, and CNNs for non-ROI regions to enable scalable and efficient compression. ROI-aware hybrids, such as the ECNN plus optimum zero-tree wavelet framework in [19], use learned encoders for non-critical regions and wavelet-based encoders for the ROI, with metaheuristics such as Sunflower Optimization used to tune thresholds and network parameters jointly.

Traditional codecs are fast, well standardized, and easy to deploy in resource-constrained clinical environments. Region-of-interest (ROI)-based frameworks have further improved their diagnostic fidelity by preserving clinically critical structures at higher quality. However, their hand-crafted basis functions and rigid quantization schemes limit their ability to adapt to the heterogeneous statistics of medical images, especially at high compression ratios where subtle diagnostically relevant features may be lost. Deep learning-based methods address this limitation by learning hierarchical, domain-specific features directly from data and have been shown to outperform traditional codecs on several natural and medical image benchmarks. Reported gains, however, often come at the cost of higher computational complexity, dependence on large training sets, and particularly in medical imaging limited interpretability of the learned latent representations. The two families thus present complementary strengths: traditional codecs are robust, interpretable, and standards-compliant, while deep codecs are adaptive and expressive in sparse coding regimes. This complementarity has driven growing interest in hybrid strategies that combine the two. Despite their variety, prior CNN-DWT hybrids share a common set of limitations that this paper addresses directly. First, the DWT codec is rarely co-designed with the surrounding CNN; most approaches treat the CNN as an add-on whose role is to refine an already quantized bitstream, so the encoder CNN does not have the opportunity to align its representation with the codec’s quantizer. Second, the DWT codec’s hard quantization step is non-differentiable, so prior works either train components in isolation or fall back on straight-through estimators, both of which limit achievable rate-distortion performance.

This paper addresses existing limitations by introducing an effective hybrid approach for compressing medical images. This method utilizes a low-complexity convolutional neural network combined with discrete wavelet transform for lossy compression, achieving a compression ratio of up to 40 while maintaining high image quality. Hybrid models that incorporate deep learning are adaptable and offer cutting-edge solutions in areas such as segmentation [20], feature extraction, and classification [21]. By merging the advantages of traditional techniques with the adaptive learning capabilities of deep models, these hybrids deliver robust, efficient, and interpretable results, making them valuable for a wide range of imaging applications. In the context of compression, deep learning models can enhance performance by utilizing their learning abilities to surpass traditional coding methods. Initially, to boost compression efficiency, a CNN on the encoding side creates an optimal compact representation that retains essential structural information from the original image. On the decoding side, another CNN

aids in the precise reconstruction of the output image. Additionally, a DWT-based image codec positioned between the encoding and decoding CNNs can further compress the compact representation. Lastly, we propose a training strategy for our hybrid model that addresses the issue of training disruption caused by non-differentiable quantization in the DWT codec.

The novelty of this work is threefold and spans architecture, algorithm, and training strategy:

- **Architectural co-design:** As opposed to previous hybrids that treat the CNN as an add-on codec, we treat the encoder CNN, the DWT-based sub-band codec, and the decoder CNN as a single end-to-end optimized pipeline. The encoder CNN produces a compact feature representation that is then compressed via a standardized wavelet codec, and the decoder CNN reconstructs the image directly from the codec output. This co-design preserves compatibility with wavelet-based standards while still benefiting from learned representations.
- **Algorithmic handling of quantization:** We introduce a training strategy that explicitly resolves the non-differentiability of the DWT codec's quantization step, enabling stable end-to-end optimization without resorting to biased straight-through estimators. The encoder and decoder CNNs are jointly tuned to the statistical behavior of the quantized wavelet coefficients, which is not possible in earlier decoupled hybrids.
- **Compact, deployment-friendly footprint:** The CNNs in the proposed framework use a deliberately small parameter count and inference latency profile, matching the computational envelopes of JPEG2000-class codecs while still exploiting learned representations. This makes the model suitable for installation on PACS edge hardware, a deployment scenario not directly targeted by the deeper transformer and foundation-model.

The proposed framework is designed to slot into existing clinical imaging pipelines with minimal disruption. Because the intermediate representation is wavelet-coded, compressed bitstreams can be transmitted, stored, partially decoded, and rendered using the same wavelet-aware tools that already serve JPEG2000 traffic in hospital networks. At the same time, the surrounding CNNs operate on a compact feature representation rather than full-resolution volumetric data, which keeps per-image compute costs compatible with edge or near-edge deployment. Anticipated downstream effects include a tangible reduction in archive storage growth, shorter transmission times for teleradiology sessions, and the ability to preserve diagnostic fidelity in commonly used modalities at high compression ratios.

The remainder of this article is organized as follows. Section II outlines the architecture and functional components of the proposed compression framework, highlighting the role of each module and the data flow within the system. Section III offers a comprehensive analysis of the experimental results, including quantitative comparisons with benchmark techniques and a discussion of the observed trends and implications. Finally, Section IV concludes the paper with a summary of key findings.

II. METHODOLOGY

This section outlines the proposed framework for compressing medical images, which integrates CNNs with a codec based on Discrete Wavelet Transform to achieve effective, low-complexity lossy compression while maintaining diagnostically important features. The architecture is crafted to enhance both the transformation and reconstruction stages of compression through a combined deep learning and signal processing approach.

A. Overview of the Proposed Framework

The framework is composed of three primary stages: an encoding CNN, a wavelet-based compression codec, and a decoding CNN. Fig. 1 depicts the entire pipeline of the proposed model. Initially, the input medical image is processed by a lightweight CNN-based encoder, which converts the image into a compact feature representation. This encoded data is then handled by a DWT-based codec to further minimize frequency redundancy. Finally, the compressed information is decoded by a symmetric CNN, which reconstructs the image with high accuracy. This modular setup allows the CNNs to learn optimal feature transformations, while the DWT codec manages efficient spatial-frequency domain compression, ensuring it aligns with existing image coding standards.

The proposed method treats the CDF 9/7 codec as a fixed, standards-compliant module that is kept outside the network and applied to the CNN output exactly as in the deployed pipeline. The key design novelty is threefold: (a) the learned encoder compresses the input into a compact three-channel feature map, enabling each feature channel to be coded by the DWT codec identically to a standard RGB image channel without requiring custom high-dimensional tensor coding; (b) the architecture is deliberately lightweight (≈ 0.26 M trainable parameters, see Table I, making it suitable for deployment in resource-constrained environments; and (c) because the DWT codec is non-differentiable, training avoids surrogate approximations and instead employs a two-stage optimization strategy that optimizes the model directly against the true quantization process..

B. CNN Architecture for Encoding and Decoding

The CNN modules for both encoding and decoding are specifically crafted to be computationally efficient, making them ideal for use in resource-limited medical environments. The encoder network comprises a series of convolutional layers that gradually reduce spatial dimensions and extract hierarchical features. In contrast, the decoder uses transpose convolutions or up sampling layers to restore the original image resolution. The encoder produces a three-channel output whose values are then passed to the DWT codec, and the decoder takes the reconstructed three-channel coefficients as input. For images normalized to the $[0, 1]$ range, the final decoder layer applies a sigmoid activation to keep the reconstruction within the valid pixel range. All intermediate layers use ReLU, and all convolutional/transposed-convolutional kernels are initialized with normal initialization. The detailed configurations of the CNN encoder and decoder, including the number of layers, kernel sizes, activation functions, output shapes, and the corresponding number of trainable parameters, are summarized in Table I. As shown in Table I, the complete encoder-decoder network contains 262,022 trainable parameters (131,011 in the encoder and 131,011 in the decoder), which keeps the model well below

0.3 M parameters. This blend of learned representations through CNNs and a mathematically grounded DWT codec

allows the model to achieve a strong balance between compression efficiency and reconstruction quality.

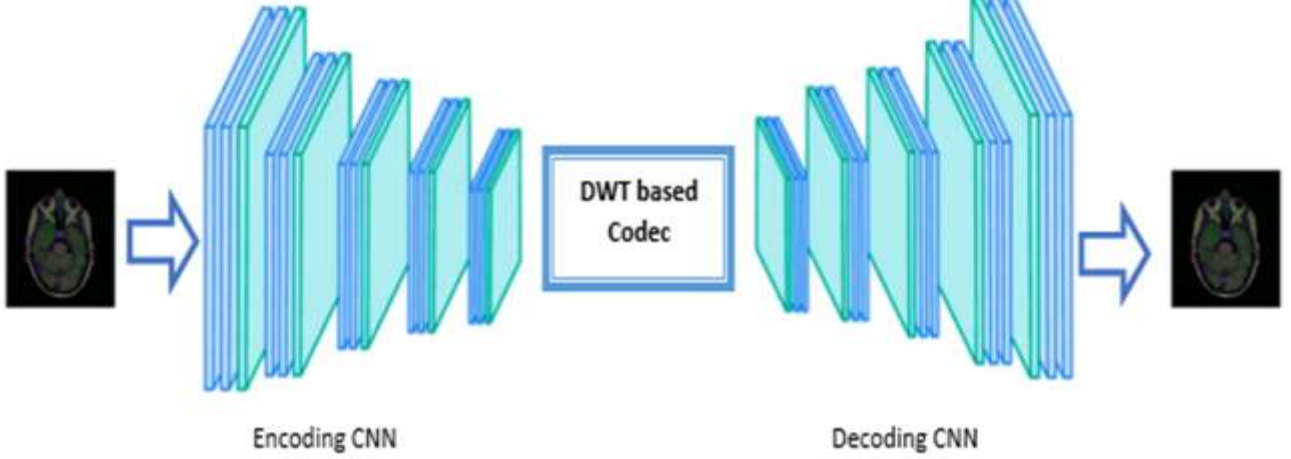


Fig. 1. Overall framework of our proposed approach

TABLE I. ARCHITECTURE DETAILS OF ENCODING AND DECODING CNNs

Encoder			
Layers	Kernel, Filter channels, Stride, Activation	Output size (128×128 input)	Trainable params
	3×3, 32 Conv, strides 1, ReLU	128×128×32	896
	3×3, 64 Conv, strides 2, ReLU	64×64×64	18,496
	3×3, 128 Conv, strides 1, ReLU	64×64×128	73,856
	3×3, 32 Conv, strides 1, ReLU	64×64×32	36,896
	3×3, 3 Conv, strides 1, ReLU	64×64×3	867
Encoder Total			131,011
Decoder			
Layers	3×3, 32 Deconv, strides 1, ReLU	64×64×32	896
	3×3, 128 Deconv, strides 1, ReLU	128×128×64	18,496
	3×3, 64 Deconv, strides 2, ReLU	128×128×128	73,856
	3×3, 32 Deconv, strides 1, ReLU	128×128×32	36,896
	3×3, 3 Deconv, strides 2, ReLU	128×128×3	867
	Decoder Total		131,011

C. Basic DWT Based Codec

A DWT-based codec compresses images by transforming them into a multi-resolution representation where most image energy is concentrated in a few significant coefficients. The CDF 9/7 wavelet, used in the JPEG2000 standard, is particularly effective for lossy compression due to its smooth basis functions and high compression efficiency [22]. It Works by passing the input image through the following steps:

1) *Wavelet Transform (Encoder)*: The image undergoes decomposition into multiple subbands utilizing the CDF 9/7 filter. This process encompasses both low-frequency (approximation) and high-frequency (detail) components in the horizontal and vertical directions. The transform is applied to each of the three encoder output channels independently using implementation of the CDF 9/7 filter, which matches the irreversible 9/7 transform standardized in JPEG2000. The decomposition is performed over 3 dyadic levels, producing one low-pass approximation subband (LL_3) and three sets of detail subbands ($LH_k, HL_k, HH_k, k = 1, 2, 3$) per channel, where the LL subband concentrates most of the image energy and the detail subbands contain progressively sparser high-frequency information.

2) *Quantization*: The transformed coefficients are subjected to quantization, thereby reducing the number of bits required to represent the image while managing distortion. Following the JPEG2000 convention, a uniform scalar dead-zone quantizer is applied, where Δ is the quantization step size $\Delta = 10$, which also serves as the rate-control mechanism. Subbands with lower perceptual importance (finest detail subbands) can be quantized more coarsely.

3) *Entropy Coding*: The quantized coefficients are further compressed through entropy coding techniques to eliminate statistical redundancy. In this work, arithmetic coding is used: the quantized indices of each subband are scanned and entropy-coded with an adaptive arithmetic coder, which achieves near-entropy bit rates by assigning shorter code words to more probable coefficient values. The coded subband bit streams are concatenated with the subband

header information (dimensions, quantization step) to form the final compressed bitstream..

4) *Decoding Process*: During reconstruction, the quantized indices are entropy-decoded, dequantized and the inverse DWT is applied to the decoded coefficients to reconstruct the image. The use of CDF 9/7 ensures smooth image reconstruction with minimal artifacts..

D. Training Strategy

To tackle the challenges posed by the quantization step in the DWT codec and to ensure consistent convergence, the training process is structured into two separate phases. Both phases utilize the Mean Squared Error (MSE) loss function, which quantifies the pixel-level reconstruction error between the original and reconstructed images.

The quantization operation in the DWT codec is non-differentiable: its gradient is zero almost everywhere, so gradients cannot be back-propagated through the codec when the encoder and decoder are trained jointly in a fully end-to-end manner. Simply replacing the quantizer with a smooth relaxation (e.g., additive uniform noise) would make training tractable but would optimize the network against a different distortion process than the one actually deployed. The two-stage strategy resolves this tension: the first phase provides a well-conditioned initialization of the encoder-decoder mapping without any quantizer in the loop, and the second phase then adapts the decoder to the real quantization noise while keeping the encoder's output statistics stationary, so the network is optimized with respect to the true compression pipeline.

Pretraining of CNN Encoder and Decoder in the initial phase, the CNN-based encoder and decoder networks are trained together in an end-to-end manner, without incorporating the DWT codec. The goal at this stage is for the encoder to learn an optimal compact representation of the input image and for the decoder to accurately reconstruct it. This pretraining phase enables the networks to develop a strong mapping between the input and output domains, free from the non-differentiable quantization operations introduced by the wavelet codec. The MSE loss is defined as:

$$L_1(\theta_i, \theta_j) = \frac{1}{N} \sum_{k=1}^N \|R(\theta_j, E(\theta_i, x_k)) - x_k\|^2 \quad (1)$$

Where x_k is the input images, E and R are the encoding and decoding CNNs, θ_i and θ_j are their respective variables, with N representing the number of images in the batch.

Fine-Tuning with DWT Integration in the subsequent phase, the trained encoder is combined with the DWT codec, followed by the CNN-based decoder. The quantization step within the DWT is maintained to accurately represent the compression pipeline. During this phase, only the decoder undergoes fine-tuning, while the encoder weights are either frozen ($\hat{\theta}_i$) depending on the specific training configuration. This fine-tuning process allows the decoder to adjust to the distortions caused by quantization, thereby enhancing reconstruction quality in the complete compression setting. With the integration of the wavelet codec referred to by D , the loss function becomes:

$$L_2(\theta_j) = \frac{1}{N} \sum_{k=1}^N \|R(\theta_j, D(E(\hat{\theta}_i, x_k))) - x_k\|^2 \quad (2)$$

This two-phase training strategy effectively reduces the negative impact of quantization on network optimization and improves the performance of the proposed compression model in real-world applications.

III. RESULTS AND DISCUSSION

A. Experimental Setup

The proposed model for compressing medical images was implemented using the NVIDIA Tesla K80 GPU available on Google Colab. The network's architecture was crafted using Keras with TensorFlow as the backend. To reduce reconstruction error during training, the Adam optimizer was utilized with an initial learning rate of 1×10^{-3} . A learning rate scheduler was employed to halve the learning rate whenever the validation loss plateaued, with a minimum threshold set at 1×10^{-6} . A batch size of 32 was used throughout training. All convolutional layers were initialized with normal initialization with zero-initialized biases. The training was carried out in two stages:

- an initial phase of 400 epochs to learn the compact feature representation.
- a fine-tuning phase of 100 epochs after integrating with the DWT codec to enhance reconstruction quality under realistic compression conditions.

B. Dataset and Preprocessing

The model's evaluation was conducted using the FLAIR MRI Low Grade Glioma (LGG) dataset from Kaggle, as detailed by [23]. The selected LGG MRI dataset comprises brain FLAIR MRI images from 110 patients, with corresponding manually annotated FLAIR abnormality segmentation masks. The dataset contains approximately 3,929 image-mask pairs. At the slice level, approximately 27.7% of the images contain a visible tumor region, whereas 72.3% are tumor-free, indicating a substantial class imbalance between tumor-containing and tumor-free slices. The dataset does not provide tumor-grade labels as the segmentation target; therefore, the class distribution was reported based on the presence or absence of a segmented tumor region rather than tumor grade. This dataset consists of 2D FLAIR MRI images, each with a resolution of 256×256 pixels and 3 color channels. A total of 1000 images containing tumors were selected for the study. To optimize memory usage and improve training efficiency, each image was divided into non-overlapping patches of 128×128 pixels. The dataset was split into three subsets: 70% for training, and the remaining 30% equally divided between validation and testing.

The dataset is publicly available on Kaggle and was used in accordance with the platform's terms of use and the dataset license [23]. The released data were de-identified by the dataset providers — patient identifiers were removed before publication, and no protected health information is involved.

C. Performance Evaluation

The proposed compression model's performance was assessed using standard metrics for medical image compression, including PSNR, MS-SSIM, and Compression Ratio (CR). The evaluation process incorporated both quantitative metrics and qualitative visual assessments. Furthermore, a comparative analysis was performed against traditional state-of-the-art techniques such as JPEG and JPEG2000, as well as deep learning-based methods and ROI-

based compression techniques, to highlight the advantages of our approach.

To assess the trade-off between image quality and compression efficiency, the PSNR, MS-SSIM, and CR results are jointly analyzed across four compression techniques: JPEG, JPEG2000, a deep learning-based model, and the proposed model. This integrated comparison enables a clearer understanding of how each method balances fidelity and compression, which is critical in medical imaging applications where both visual accuracy and data reduction are essential. JPEG and JPEG2000 were selected as the only conventional baselines because they are the de facto standards deployed in clinical imaging practice — and JPEG2000, built on the same CDF 9/7 wavelet used in our codec, making them the most meaningful state-of-the-art conventional references [10]. Modern learned codecs were intentionally excluded from the main comparison for two reasons: (i) they typically require one to two orders of magnitude more parameters and substantially more computation -parameter model proposed here, which contradicts the low-complexity objective of this study; and (ii) the comparison is deliberately scoped to baseline models and the state-of-the-art JPEG/JPEG2000 standards, since these define the operating conditions of real PACS systems.

From the figures (Fig. 2, Fig. 3, and Fig. 4), a clear trend emerges: traditional codecs such as JPEG and JPEG2000 operate at a compression ratio of 30:1, but their image quality varies significantly. JPEG, with the lowest PSNR 28.46 dB and lowest MS-SSIM 87.3%, demonstrates that high compression without perceptual optimization leads to notable quality degradation. JPEG2000 improves on this with 34.38 dB PSNR and 96.7% MS-SSIM, but it still lags behind deep learning-based approaches. The 5.9 dB advantage of JPEG2000 over JPEG at the same 30:1 CR reflects the superior energy compaction of the wavelet transform compared with the block-based DCT, confirming that the wavelet domain is the appropriate substrate for medical-image compression. In contrast, the deep learning model [10] achieves both higher PSNR 35.4 dB and MS-SSIM 96.9%, while pushing the CR up to 38:1, indicating that learning-based methods are more capable of capturing complex image features while compressing aggressively. The proposed model outperforms all others with the highest PSNR of 38.6 dB and MS-SSIM of 98.84%, even at a CR of 40:1. Interpretation rather than restatement: the 3.2 dB gain over the DL baseline [10], achieved while the CR is simultaneously raised from 38:1 to 40:1, shows that the proposed method does not merely exchange bits for quality — it moves the operating point outward on the rate-distortion plane. The learned encoder followed by the CDF 9/7 codec therefore yields a latent representation whose coefficients are more compact (lower entropy) and more robust to quantization, so less distortion is incurred per retained bit. The 98.84% MS-SSIM additionally indicates that structural and textural information is preserved almost entirely the property most relevant to diagnostic readability. This illustrates that our approach is not merely increasing compression but is also improving the perceptual and objective image quality simultaneously.

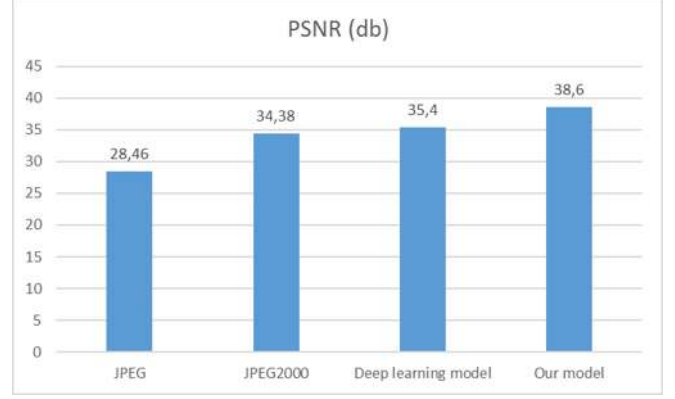


Fig. 2. Graphical illustration of PSNR comparison

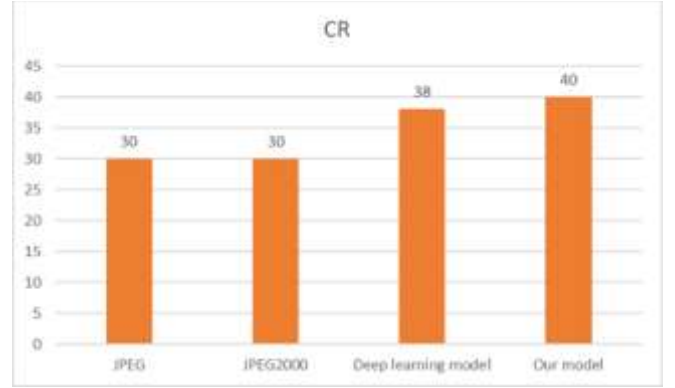


Fig. 3. Graphical illustration of CR comparison

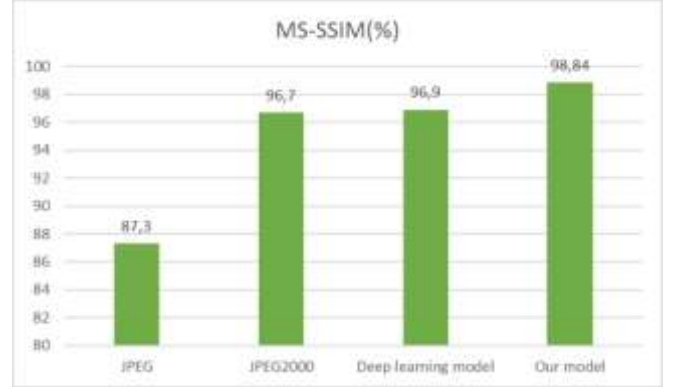


Fig. 4. Graphical illustration of MS-SSIM comparison

To evaluate the performance of the proposed hybrid compression model, a comparative study was performed against several existing medical image compression techniques that focus on hybrid approaches. The reference works selected for this analysis include the methods outlined in method [24], method [7], and method [19]. These particular methods were chosen because they are specifically applied to medical imaging, making them pertinent for benchmarking the current research. Each of these previous approaches employs a multi-stage compression strategy that emphasizes maintaining the integrity of diagnostically important areas while allowing for greater compression of the background or less informative sections. Table II summarizes the PSNR, MSE, CR.

TABLE II. COMPARAISON AGAINST PUBLISHED WORKS

Methods	Metrics		
	PSNR (dB)	MSE	CR
method [24]	33.51	28.96	24.61
method [7]	34.69	22.08	4.63
method [19]	42.77	3.43	14.91
Our method	38.6	8.96	40

As shown in Table II, the proposed method excels by achieving the highest compression ratio (CR) of 40 and the lowest MSE of 8.96, while also maintaining PSNR of 38.6 db. In comparison, the next best CR was achieved by method [24], who reported a CR of 24.61, but their PSNR was only 33.51 dB, which is 4.4 dB lower than our method, indicating a noticeable decline in reconstruction quality. Other methods exhibited much lower compression efficiency, with CRs under 5, despite having moderate PSNR values. For example, method [7] achieved a PSNR of 34.69 dB with a CR of just 4.63. Notably, the method by method [19] achieved the highest PSNR of 42.77 dB, indicating excellent visual quality, but this was at the expense of a lower CR of 14.91, making it less efficient for storage or transmission in bandwidth-limited environments compared to our model. Why [19] attains a higher PSNR at a lower CR: a substantially higher PSNR at roughly $2.7\times$ lower compression is the expected signature of a near-lossless or bit-preserving scheme — retaining more bits naturally lowers distortion. In bandwidth-constrained telemedicine and archiving, however, the extra 4.2 dB beyond our 38.6 dB has limited practical value, because beyond $\approx 35 - 38$ dB residual differences are barely perceptible and are not necessarily correlated with diagnostic utility, whereas the $2.7 \times$ larger CR directly determines storage cost and transmission latency. Our model therefore occupies the practically preferable region of the rate-distortion plane for clinical deployment. This trade-off underscores the strength of our proposed technique, which achieves an optimal balance between reconstruction quality and compression efficiency, a critical requirement for practical use in medical imaging systems.

To further assess the visual performance of the proposed method, Fig. 5 presents a qualitative comparison at the same compression ratios evaluated in Fig. 3, showing (a) the original images, (b) the reconstructed images obtained using the proposed method, and (c) those obtained using JPEG. Overall, the proposed method provides reconstructed images with better visual quality and greater similarity to the original images, while JPEG exhibits more noticeable compression artifacts and overall visual degradation. These results demonstrate the improved visual quality achieved by the proposed method at the considered compression ratios, complementing the quantitative results presented earlier.

The experimental findings and comparative evaluations strongly support the effectiveness of the proposed deep learning-assisted hybrid compression framework. In terms of objective quality, our model achieved the highest PSNR value, indicating that the pixel-wise reconstruction error is of its original size without perceptible loss, directly reducing the storage and bandwidth budgets of telemedicine and cloud-based PACS. The MS-SSIM score of 98.84% further confirms that the perceptual quality of the reconstructed images remains very high. Compared to JPEG and even the deep learning

baseline, our model preserves structural information more effectively. High MS-SSIM values suggest that critical diagnostic features, such as spatial relationships and organ morphology, are well maintained even under aggressive compression. Our model achieved a higher compression ratio of 40 with less perceptible loss in visual quality. The ability to compress more aggressively without sacrificing image fidelity is crucial for modern telemedicine and cloud-based medical record systems, where large volumes of imaging data must be stored and transmitted efficiently.

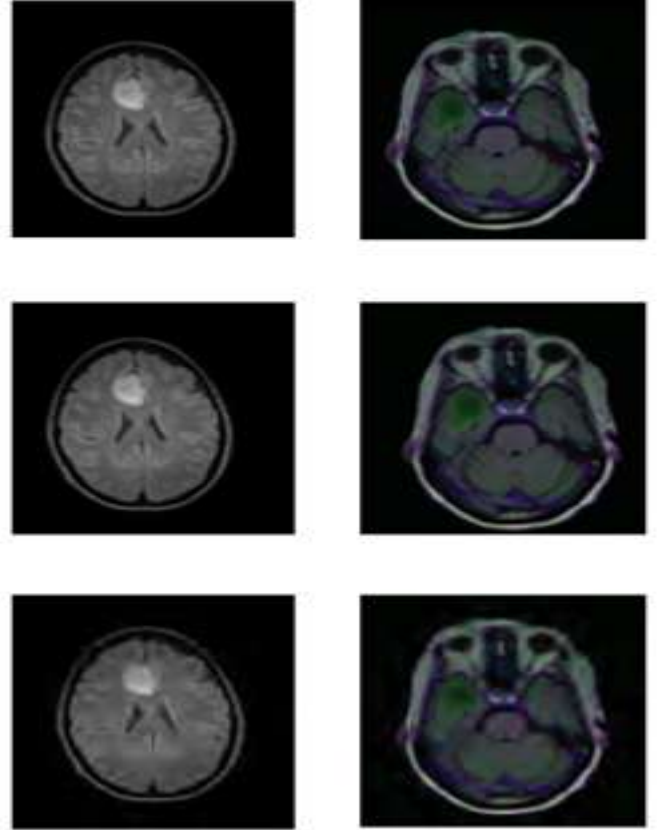


Fig. 5. Visual quality comparison: (a) original images; (b) reconstructed images obtained using the proposed method; and (c) reconstructed images obtained using JPEG.

Compared to ROI-based hybrid compression methods like those in method [24], method [7], and method [19], which involve complex segmentation and multiple encoding schemes, our model simplifies the architecture while delivering better performance. Our method bypasses the need for explicit ROI segmentation by learning to preserve essential features directly through end-to-end training. The CNN learns a diagnostic-saliency-preserving transform from data, while the standards-compliant CDF 9/7 codec handles the coding, keeping the framework compatible with existing JPEG2000-based infrastructure rather than requiring a bespoke decoder. Because the approach avoids the fragile ROI-segmentation stage of [7], [19], [24] whose accuracy degrades on atypical lesions, it is more robust and substantially simpler to deploy, and its ≈ 0.26 M-parameter footprint makes retraining feasible on a single Colab GPU. Broader impact: the same CNN + DWT template generalizes to any modality representable as a 3-channel tensor (CT, X-ray, ultrasound, mammography), and the two-stage training procedure — pretraining without the codec, then fine-tuning the decoder only — is directly reusable for other non-differentiable codecs

(e.g., SPIHT, EBCOT), offering a low-cost recipe for coupling learned representations with standard coding tools.

The principal limitation of this study is that evaluation was carried out on a single dataset and a single modality (FLAIR MRI). This choice was made to keep the first validation controlled (Section B); however, it limits the generalizability of the claims, because the optimal feature transforms learned for MRI may differ for CT, X-ray, ultrasound, or mammography. The following extensions are therefore planned: (i) evaluation on multiple public datasets and across imaging modalities, with modality-appropriate retraining; (ii) comparison with modern learned codecs, once the low-complexity constraint is relaxed; (iii) diagnostic quality assessment by medical experts, reported in the dedicated quality assessment section; (iv) extension of the statistical analysis with additional cross-validation folds and larger test sets; and (v) investigation of learnable quantization to enable fully joint optimization of the encoder and the codec. Addressing these points will further substantiate the clinical applicability of the proposed framework.

IV. CONCLUSION

This study introduced a hybrid method for compressing medical images by integrating a CNN with a DWT-based codec, minimizing redundancy in both spatial and frequency domains while preserving essential diagnostic details. Experimental results demonstrated a compression ratio of 40:1 with a PSNR of 38.6 dB and an MS-SSIM of 98.84%, reflecting high perceptual and structural image quality that supports efficient, high-quality data compression for healthcare and research. Reducing the size of large medical datasets lowers storage and transmission costs, enabling remote professionals to access high-resolution images for diagnosis and treatment. Scientifically, the framework decouples the learned representation from the coding stage: the CNN learns what to preserve, diagnostically salient structures, while the standards-compliant CDF 9/7 codec performs the coding, keeping the method compatible with JPEG2000-based infrastructure and providing a two-stage training recipe transferable to other non-differentiable codecs. Its 0.26 M-parameter footprint is trainable on a single commodity GPU, lowering the entry barrier for resource-limited institutions. Despite these strengths, the study has limitations: evaluation was limited to a single dataset and modality (FLAIR MRI), the comparison included only JPEG, JPEG2000, and a small set of hybrid methods, the reported metrics correspond to a single operating point without rate-distortion analysis or statistical significance testing, and no assessment by medical experts was performed. Future work will therefore validate the framework on multiple datasets and modalities (CT, X-ray, ultrasound, mammography), benchmark against modern hyperprior-based and transformer-based codecs, conduct rate-distortion and statistical analyses, and incorporate expert diagnostic quality assessment. We also plan to explore alternative architectures, integrate the method with other classical and deep learning-based compression techniques in a modular framework, and enhance computational efficiency for real-time encoding and decoding in telemedicine and medical archive systems.

REFERENCES

- [1] P. K. Sran, S. Gupta, and S. Singh, "Segmentation based image compression of brain magnetic resonance images using visual saliency," *Biomedical Signal Processing and Control*, vol. 62, Art. N°. 102089, Sep. 2020, <https://doi.org/10.1016/j.bspc.2020.102089>
- [2] Z. Li, A. Ramos, Z. Li, M. L. Osborn, W. Zaid, X. Li, Y. Li, J. Xu, "Nearly-lossless-to-lossy medical image compression by the optimized JPEGXT and JPEG algorithms through the anatomical regions of interest," *Biomedical Signal Processing and Control*, vol. 83, Art. Numb. 104711, May 2023, <https://doi.org/10.1016/j.bspc.2023.104711>
- [3] X. Liu., L. Zhang, Z. Guo, T. Han, M. Ju, B. Xu, H. Liu "Medical Image Compression Based on Variational Autoencoder," *Mathematical Problems in Engineering*, vol. 2022, Art. Numb. 7088137. Dec. 2022, <https://doi.org/10.1155/2022/7088137>
- [4] S. Aumann, S. Donner, J. Fischer, and F. Müller, "Optical Coherence Tomography (OCT): Principle and Technical Realization," in: J. F. Bille, "High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics", Ed., Cham: Springer International Publishing, 2019, pp. 59–85. https://doi.org/10.1007/978-3-030-16638-0_3
- [5] E. Aldemir, G. Tohumoglu, and M. A. and Selver, "Binary medical image compression using the volumetric run-length approach," *The Imaging Science Journal*, vol. 67, no. 3, Apr. 2019, pp. 123–135, <https://doi.org/10.1080/13682199.2019.1565695>
- [6] M. K. Abdmouleh, H. Amri, A. Khalfallah, and M. S. Bouhlel, "A Fast JPEG2000 Based Crypto-Compression Algorithm: Application to the Security for Transmission of Medical Images," In: Balas, V., Jain, L., Balas, M. (eds) *Soft Computing Applications*. SOFA 2016. *Advances in Intelligent Systems and Computing*, vol. 633. Springer, Cham pp. 164–175. https://doi.org/10.1007/978-3-319-62521-8_14
- [7] C. P. Devadoss and B. Sankaragomathi, "Near lossless medical image compression using block BWT-MTF and hybrid fractal compression techniques," *Cluster Comput*, vol. 22, no. 5, pp. 12929–12937, Sep. 2019, <https://doi.org/10.1007/s10586-018-1801-3>
- [8] P. Sreenivasulu and S. Varadarajan, "An Efficient Lossless ROI Image Compression Using Wavelet-Based Modified Region Growing Algorithm," *Journal of Intelligent Systems*, vol. 29, no. 1, Jan. 2020, pp. 1063–1078, <https://doi.org/10.1515/jisys-2018-0180>
- [9] N. H. Hussein and M. A. Ali, "Medical Image Compression and Encryption Using Adaptive Arithmetic: Coding, Quantization Technique and RSA in DWT Domain," *Iraqi Journal of Science*, Vol. 63, N°. 5 pp. 2279–2296, May 2022, <https://doi.org/10.24996/ijis.2022.63.5.38>
- [10] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end Optimized Image Compression," in *5th International Conference on Learning Representations (ICLR)*, Toulon, France, Apr. 2017. arXiv:1611.01704, <https://doi.org/10.48550/arXiv.1611.01704>
- [11] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," in *6th International Conference on Learning Representations (ICLR)*, Vancouver, Canada, Apr. 2018. arXiv:1802.01436. <https://doi.org/10.48550/arXiv.1802.01436>
- [12] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. V. Gool, "Generative Adversarial Networks for Extreme Learned Image Compression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 02 November 2019, Seoul, Korea (South), pp. 221–231. <https://doi.org/10.1109/ICCV.2019.00031>
- [13] S. Shukla and A. Srivastava, "Medical images Compression using convolutional neural network with LWT," *IJMCTR*, vol. 6, no. 6, Art Numb. 265086, Jun. 2018, Accessed: Jan. 24, 2025. [Online]. Available: <https://www.neliti.com/publications/265086/>
- [14] A. Kar, S. Phani Krishna Karri, N. Ghosh, R. Sethuraman, and D. Sheet, "Fully Convolutional Model for Variable Bit Length and Lossy High Density Compression of Mammograms," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 2591–2594. Accessed: Jan. 24, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018_workshops/w50/html/Kar_Fully_Convolutional_Model_CVPR_2018_paper.html
- [15] D. Minnen, J. Ballé, and G. D. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," in book "Advances in Neural Information Processing Systems", editors: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett. Vol. 31, Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper_files/paper/2018/file/53edebc543333dfbf7c5933af792c9c4-Paper.pdf
- [16] C. Fu, B. Du, and L. Zhang, "Hybrid-context-based multi-prior entropy modeling for learned lossless image compression," *Pattern*

- Recognition*, vol. 155, art. Numb. 110632, Nov. 2024, <https://doi.org/10.1016/j.patcog.2024.110632>
- [17] J. Lu, L. Zhang, X. Zhou, M. Li, W. Li, and S. Gu, "Learned Image Compression with Dictionary-Based Entropy Model," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 10-17 June 2025, pp. 12850–12859. <https://doi.org/10.1109/CVPR52734.2025.01199>
- [18] S. Li, J. Lu, Y. Hu, L. S. Mattos, and Z. Li, "Towards scalable medical image compression using hybrid model analysis," *J Big Data*, vol. 12, art Numb. 45, Feb. 2025, <https://doi.org/10.1186/s40537-025-01073-1>
- [19] P. V. Bindu and A. Jabeena, "ROI and Non-ROI Image Compression Using Optimal Zero Tree Wavelet and Enhanced Convolutional Neural Network for MRI Images," *SN COMPUTER. SCIENCE.*, vol. 5, art. numb. 38, Nov. 2023, <https://doi.org/10.1007/s42979-023-02335-6>
- [20] B. Perumal and M. P. Rajasekaran, "A hybrid discrete wavelet transform with neural network back propagation approach for efficient medical image compression," in *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, Pudukkottai, India, 24-26 February 2016, pp. 1–5. <https://doi.org/10.1109/ICETETS.2016.7603060>
- [21] R. Mostafiz, M. S. Uddin, N.-A.- Alam, Md. Mahfuz Reza, and M. M. Rahman, "Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, Part B, Jun. 2022, pp. 3226–3235, <https://doi.org/10.1016/j.jksuci.2020.12.010>
- [22] M. Beladgham, Y. Habchi, M. Ben Aissa, and A. Taleb-Ahmed, "Medical video compression using bandelet based on lifting scheme and SPIHT coding: In search of high visual quality," *Informatics in Medicine Unlocked*, vol. 17, p. 100244, 2019, <https://doi.org/10.1016/j.imu.2019.100244>
- [23] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in Biology and Medicine*, vol. 109, pp. 218–225, <https://doi.org/10.1016/j.combiomed.2019.05.002>
- [24] S. S. Magar and B. Sridharan, "Hybrid image compression technique using oscillation concept & quasi fractal," *Health Technol.*, vol. 10, an. 2020, pp. 313–320, <https://doi.org/10.1007/s12553-018-00282-4>