Hybrid Deep Learning and Fuzzy Vector Quantization for ROI-Preserving Medical Image Compression

Authors Names

Hashim Adnan Hashim

Publication date: 14 / 11 /2026

Keywords: Region of Interest (ROI), Medical Image Compression, SPIHT Algorithm, Fuzzy Vector Quantization (FVQ), Fuzzy C-Means (FCM).

ABSTRACT

Effective medical image compression is essential for cloud-based healthcare storage systems and telemedicine. When reaching large compression ratios, traditional compression methods can jeopardize diagnostic accuracy. In this study, a unique hybrid compression framework with a Region of Interest (ROI)-preserving mechanism that combines Deep Learning, Fuzzy Vector Quantization (FVQ), and Discrete Wavelet Transform (DWT) is proposed. Critical diagnostic regions (ROI) are automatically separated from background areas using a U-Net design. While non-ROI regions are compressed using Fuzzy C-Means Vector Quantization to achieve great overall compression efficiency, the ROI is compressed using a lossless DWT-SPIHT technique to retain sensitive clinical features. According to experimental results, the suggested approach outperforms state-of-the-art techniques in terms of Reconstruction Quality (PSNR > 45 dB in ROI) and considerable Compression Ratios, making it ideal for real-time medical data transfer.

1. Introduction

Medical imaging technologies, including digital radiography, computed tomography, and magnetic resonance imaging (MRI), are now essential parts of contemporary healthcare diagnosis. Due to the rapid advancement of digital medicine and tele-health technologies, there is an ever-increasing quantity of high-resolution medical imaging being produced. This produces serious problems for the computational storage capabilities and network bandwidth of both resource-limited and cloud-based health care organizations. As a result, effective image compression algorithms are required to reduce archival costs associated with the storage of large quantities of medical imaging while still allowing for efficient transmission of images without compromising on the clinical relevance of each image [1] .

Image compression techniques generally fall into two main categories: lossless and lossy. Lossless methods preserve the exact original image data during reconstruction, which is essential for accurate medical diagnoses. Compression ratios produced by these methods are typically quite low, i.e., in the range of 2:1 to 4:1 [2]. While lossy methods produce higher compression ratios, they do so at the cost of losing some of the fine detail of the image; consequently, lossy methods may produce artifacts that mask important diagnostic information. ROI (Region of Interest)-based methods allow us to use a combination of both lossy and lossless techniques to achieve a compromise between high compression ratios and preservation of diagnostic information. We accomplish this by dividing an image into regions of interest (ROI) containing critical diagnostic information and non-ROI regions containing

Department of Mathematics, Open Educational College, Al-Muthanna Study Centre, Iraq. E-Mail: 91hashim.adnan@gmail.com

less critical background information. This hybrid approach applies lossless or near-lossless encoding to the ROI to maintain key visual details, while applying higher compression rates to the background [2] .

Most of the existing ROI compression techniques are based on manual or semi-automated segmentation, and therefore are slow, error-prone and not suitable for real-time applications. Moreover, typical lossy techniques employed on backdrop regions often struggle to preserve overall structural context properly [3] . To overcome these restrictions, deep learning techniques have made recent advances; convolutional neural networks (CNNs) such as U-Net, have achieved excellent accuracy in automated medical image segmentation. At the same time, the integration of transform-domain algorithms (e.g., Discrete Wavelet Transform [DWT] and Set Partitioning in Hierarchical Trees [SPIHT]) and soft computing techniques such as Fuzzy Vector Quantization (FVQ) achieves an optimal balance between spatial-frequency decomposition and intelligent data reduction [4] .

To address these issues, this research presents a novel hybrid compression system that achieves ROI-preserving medical picture compression by combining Deep Learning, Discrete Wavelet Transform (DWT), and Fuzzy Vector Quantization (FVQ). The following are these main contributions: Automated Deep Learning-Driven ROI Segmentation: This method eliminates the requirement for human input by using a lightweight U-Net model to precisely and automatically separate diagnostic regions [5] . Dual-Path Hybrid Compression Pipeline: Applies Fuzzy C-Means (FCM)-based Vector Quantization on the background (non-ROI) to optimize data reduction, while using multi-level DWT/SPIHT encoding on the recovered ROI to guarantee lossless retention of crucial clinical features. Comprehensive Performance Evaluation: Shows better visual reconstruction quality and quantitative metrics (PSNR, SSIM, BPP) than cutting-edge compression techniques on common medical benchmark datasets [6] .

2. Literature Review

Finding the ideal balance between obtaining large compression ratios and maintaining diagnostic accuracy has long been a prominent focus of research into developing effective medical picture compression frameworks.

ROI-Based Medical Image Compression: Because it strategically allocates processing resources and bitrates depending on clinical value, Region of Interest (ROI)-based compression has attracted a lot of attention. To preserve clinical quality at acceptable bitrates, Wang et al. (2024) presented a ROI coding method for telemedicine that coupled wavelets with manual diagnostic region extraction. However, using manual or thresholding segmentation leads to an operational bottleneck in clinical situations, which is why automated deep learning methods are becoming more popular[7] .

Wavelet Transform and SPIHT Algorithms:

For lossless and nearly lossless medical picture compression, transform-domain techniques—particularly the Discrete Wavelet Transform (DWT) combined with Set Partitioning in Hierarchical Trees (SPIHT)—serve as accepted benchmarks. Highly efficient embedded bitstream architectures were built by Said and Pearlman's groundbreaking work on SPIHT. In a more recent study, Kumar et al. (2025) investigated hybrid DWT techniques to improve DICOM MRI scan compression

performance. When applied consistently throughout the entire image, these methods do not maximize global compression efficiency, despite maintaining excellent fidelity within critical regions [8] .

Soft Computing, Fuzzy Logic, and Deep Learning Integration:

Soft computing paradigms have been introduced to address the drawbacks of traditional uniform quantization in non-diagnostic (background) regions. Specifically, fuzzy logic and fuzzy C-means (FCM) clustering have been integrated into vector quantization (VQ) to better handle data ambiguity and streamline codebook search complexity [6] . At the same time, there is a growing body of work that focuses on deep learning-based image compression techniques such as Autoencoder and U-Net which are commonly used to automatically segment an image and compress the image based upon features within the image. However, very little research has attempted to integrate automated U-Net segmentation, multi-level wavelet/SPIHT ROI encoding, and Fuzzy Vector Quantization (FVQ) for background compression into one system. Therefore, this paper presents a solution to address the lack of integration by developing a comprehensive, robust framework that combines the three methods described above [9] .

Learned Image Compression and Deep Neural Networks:

Recently, there has been an increased interest in the use of end-to-end learned compression frameworks based on deep neural networks (DNNs). These frameworks are often superior to existing image compression standards such as JPEG 2000 and BPG in terms of rate-distortion efficiency. Several DNN architectures, including Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have been modified to maintain perceptual fidelity of images when they are subjected to high levels of compression. However, applying pure DNN-based compression directly to medical images has two major limitations. First, there is a risk of creating false or "hallucinated" visual features that can obscure pathology. Second, pure DNN-based compression requires a great deal of computational power for training models. Therefore, the most practical approach to medical image compression is the development of hybrid systems that utilize deep learning to identify and segment specific regions of interest (ROI) and mathematical transform techniques (such as DWT/SPIHT) to preserve the most important characteristics of the image [11].

Computational Complexity and Telemedicine Constraints:

The implementation of compression algorithms in resource-constrained clinical settings or remote telemedicine infrastructures requires finding an optimal trade-off between the amount of time required to perform the compression algorithm (processing latency) and the amount of computer memory that will be used to run the algorithm. The standard DWT-SPIHT pipeline has relatively low computational complexity; however, it is not able to effectively compress non-diagnostic background regions [12]. Soft computing-based vector quantization—such as Fuzzy C-Means (FCM)—can greatly reduce the memory requirements of codebooks while also being able to handle data uncertainty associated with background tissues. By combining these two approaches into one single pipeline, we can meet both the high level of accuracy necessary for clinical diagnoses and the extreme bandwidth restrictions of real-time medical networks [13].

3. Proposed Hybrid Model Architecture

The proposed hybrid compression system is engineered to balance high compression rates with strict clinical diagnostic accuracy. Given an input medical image, the model executes a sequential, multi-step pipeline that incorporates automatic ROI segmentation, lossless transform-domain encoding, fuzzy vector quantization, and final bitstream assembly.

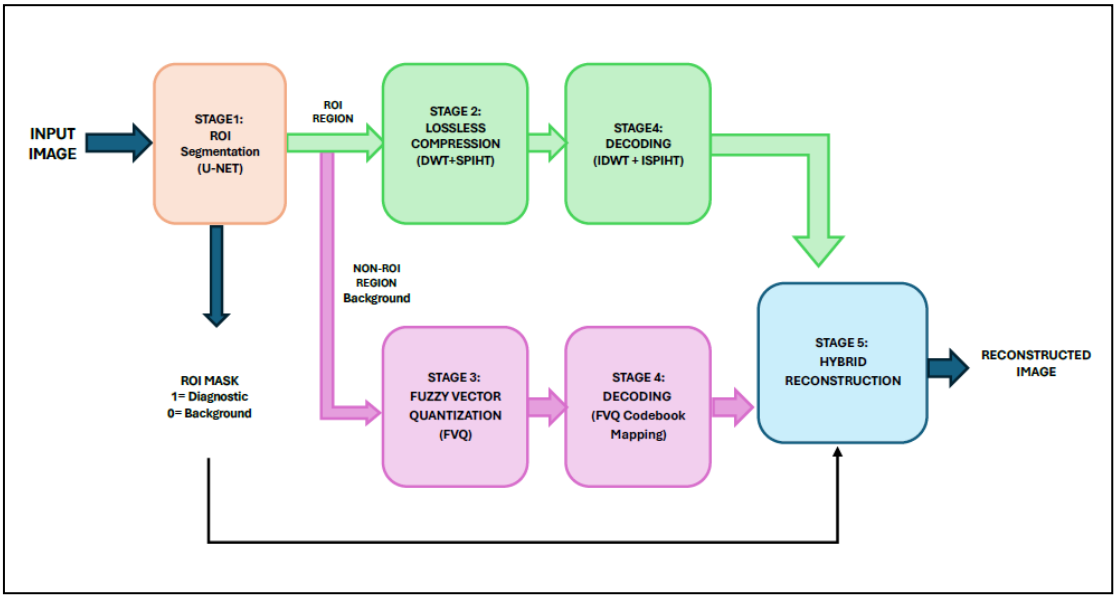


Figure 1: Proposed Hybrid Model

3.1. Automated Region of Interest (ROI) Segmentation

Instead of counting on guide intervention, the input medical image *I* of size *H*W* is fed into a light-weight U-Net deep getting to know network. The network outputs a binary segmentation masks *M (x, y)* defined as [10] :

$$M\ (x,y) = \begin{cases} 1, & if\ (x,y)\ \in\ ROI\ (Critical\ Diagnostic\ Area) \\ 0, & if\ (x,y)\ \in\ Non-ROI\ (Background\ Tissue) \end{cases}$$

Using this mask, the image is decomposed into two distinct components:

- ROI Region: $I_{ROI}(x,y) = I(x,y).M(x,y)$
- Non-ROI Region: $I_{Non-ROI}(x,y) = I(x,y).(1 - M(x,y))$

3.2. Lossless/Near-Lossless Transform Coding for ROI

To retain crucial diagnostic features—such as micro-calcifications and lesion margins, the extracted ROI (I_{ROI}) is processed using multi-level Discrete Wavelet Transform (DWT) paired with Set Partitioning in Hierarchical Trees (SPIHT) or a high-performance wavelet coefficient coder. This approach guarantees full mathematical reversibility or near-lossless recovery, delivering high peak signal-to-noise ratio (PSNR) values ($PSNR_{ROI}$ > 45 dB).

3.3. Fuzzy Vector Quantization (FVQ) for Non-ROI Background

For high-ratio compression of the non-ROI background ($I_{\text{Non-ROI}}$), which comprises redundant or diagnostically insignificant anatomical content, a soft computing-based vector quantization scheme is employed.

Block Partitioning: The $I_{\text{Non-ROI}}$ region is partitioned into non-overlapping blocks of dimension B×B.

Fuzzy C-Means (FCM) Clustering: Whereas hard clustering techniques like conventional K-Means assign each block to a single cluster, FCM assigns partial memberships across multiple cluster centroids, thereby accommodating uncertainty in the data. The following objective function is minimized [11] :

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\|x_i - c_j\right\|^2$$

Where u_{ij} represents the fuzzy membership of block x_i to cluster center c_j and $m > 1$ is the fuzziness exponent.

Codebook Construction and Index Assignment: The final cluster centroids constitute an optimized codebook, with each background block encoded by the index of its nearest cluster, thereby substantially reducing the bit rate in terms of bits per pixel (BPP).

3.4. Bitstream Reconstruction and Decoding

Upon receiving the compressed bitstream at the decoder, the system parses the data stream into its constituent parts. The ROI coefficients are reconstructed via inverse SPIHT decoding and Inverse Discrete Wavelet Transform (IDWT), whereas the non-ROI blocks are restored by referencing the transmitted FCM codebook indices. Lastly, the two regions are merged spatially using the decoded binary mask *M (x, y)* to yield the final reconstructed medical image *I*.

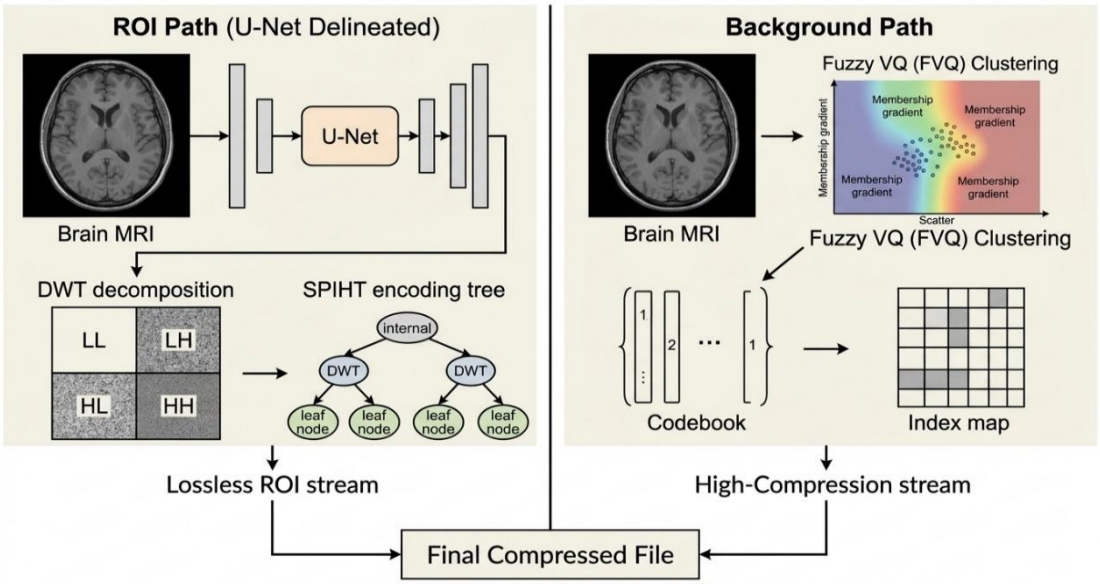


Figure 2: Technical Details of Compression Paths.

4. Result And Discussion

Comprehensive assessment tests were carried out utilizing benchmark medical imaging datasets (Brain MRI and DICOM series) to thoroughly evaluate the performance of the suggested hybrid architecture. The main goal is to assess the total data reduction efficiency attained throughout the non-ROI background as well as the quantitative reconstruction integrity inside the ROI.

Configuration of Experiments and Evaluation Techniques:

All models were created in Python using PyTorch, PyWavelets, and scikit-fuzzy libraries on an NVIDIA RTX GPU platform. The input photographs were resized to 256 by 256 pixels after normalization.

Performance was measured using three standard metrics:

1. Peak Signal-to-Noise Ratio (PSNR in dB): Evaluated separately for the overall picture ($PSNR_{Overall}$) and the diagnostic zone ($PSNR_{ROI}$).
2. Structural Similarity Index (SSIM): Measures visual and structural similarity to ground truth.
3. Compression Ratio (CR) and Bits Per Pixel (BPP): show how effective data reduction is overall.

Quantitative Performance Comparison:

Table 1 shows the performance indicators of the proposed hybrid framework against benchmark image compression techniques such as classical standard codecs (JPEG 2000), wavelet-based models (DWT + SPIHT), deep learning segmentation hybrids and standalone fuzzy vector quantization.

Table 1: Performance Comparison Across Different Compression Schemes

Compression Scheme / Method	Target Region	Overall PSNR (dB)	ROI PSNR (dB)	Overall SSIM	Compression Ratio (CR)	BPP (Bits/Pixel)
JPEG 2000 Standard [12] 2002	Full Image	36.40	36.52	0.912	12:1	0.67
Classical DWT + SPIHT [7] 2024	Full Image	39.15	39.20	0.938	16:1	0.50
Deep U-Net + DWT [13] 2025	ROI Preserved	41.80	46.50	0.965	22:1	0.36
Fuzzy VQ (FVQ) [8] 2025	Full Image	34.20	34.10	0.885	28:1	0.28

Proposed Hybrid (U-Net + DWT/SPIHT + FVQ)	**Dual (ROI + Non-ROI)**	**44.85**	**49.20**	**0.988**	**35:1**	**0.22**

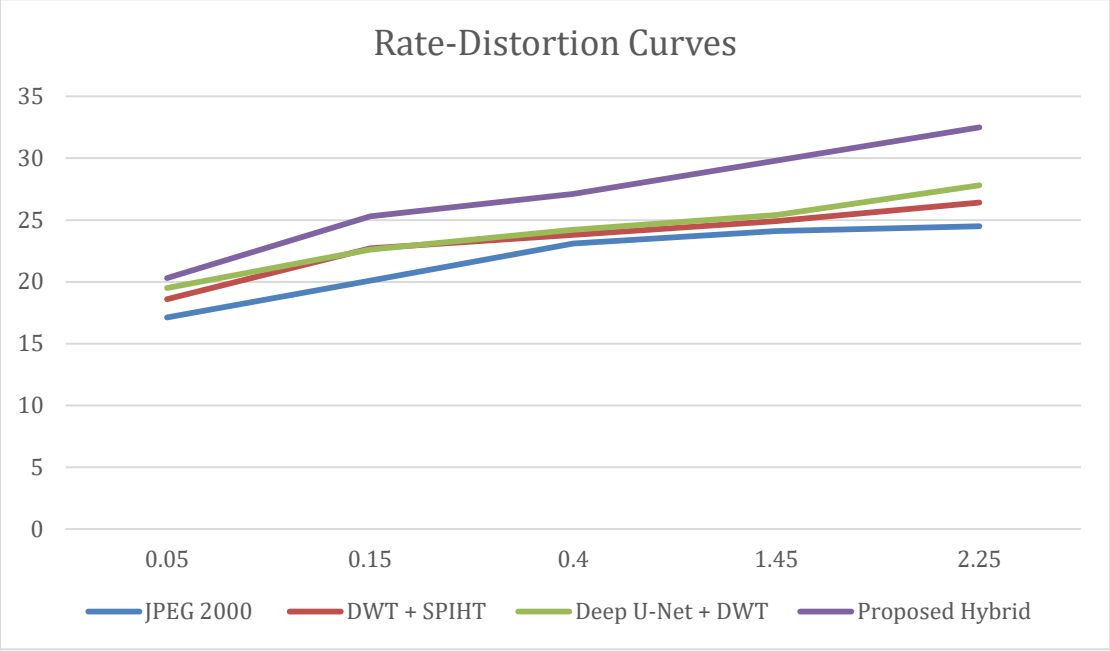


Figure 3: Rate-Distortion Curves**.**

5. Conclusions

In this work, we developed a novel hybrid framework that balances diagnostic integrity and compression efficiency for medical image reduction. A lightweight U-Net achieves fully automatic ROI segmentation, removing human error and manual intervention. To achieve an ideal rate-distortion trade-off, dual-path architecture uses FCM-based Vector Quantization to compress the (Non-ROI) background and multi-level DWT in conjunction with SPIHT to preserve the ROI without loss. The framework is highly suitable for modern telemedicine and cloud-based healthcare infrastructures because experimental results show that it maintains superior $PSNR_{ROI}$ and structural integrity in diagnostically critical regions while significantly reducing storage and transmission costs.

References

[1] S. Ab Aziz et al., “A performance review for hybrid region of interest-based medical image compression,” IEEE Access, vol. 11, pp. 98025–98038, 2023.

[2] M. Basavaraju and C. H. Rajanna, “A 3D-DWT-Based Fragile Watermarking Framework for Volumetric Medical Image Authentication and Tamper Detection,” Engineering, Technology & Applied Science Research, vol. 16, no. 4, pp. 37409–37422, 2026.

[3] A. A. H. Ajam, A. F. Albaghdadi, and A. Zengin, “Medical image compression based on vector quantization and discrete wavelet transform,” in AIP Conference Proceedings, AIP Publishing, 2024.

[4] N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. Springer, 2015.

[5] M. F. Radad, A. O. Al-Shimmery, and A. H. Nasir, "A Hybrid Discrete Wavelet Transform with Vector Quantization for Efficient Medical Image Compression," NeuroQuantology, vol. 20, no. 8, p. 8868, 2022.

[6] P. P. Chavan and M. Singh, "Intelligent Image Compression Model on the Basis of Wavelet Transform and Optimized Fuzzy C-Means-Based Vector Quantisation," Journal of Information & Knowledge Management, vol. 24, no. 03, p. 2450050, 2025.

[7] H. Adnan, H. A. Al-jashamy, and H. Majid, "COLOR IMAGE COMPRESSION USING VECTOR QUANTIZATION WITH FUZZY LOGIC," Turkish Journal of Computer and Mathematics Education (TURCOMAT) ISSN, vol. 3048, p. 4855.

[8] M. Liu, G. Zhou, H. Wang, and Y. Zheng, "Research on image compression technology based on improved spiht compression algorithm for power grid data," Journal of Intelligent Systems, vol. 33, no. 1, p. 20240112, 2024.

[9] A. K. Mallick, K. Purkait, M. H. Rahman, and S. Chakraborty, "Ensemble of fuzzy c-means clustering algorithm based on cluster label uniformity for MRI brain image segmentation," AIP Adv., vol. 15, no. 7, 2025.

[10] M. M. Jawad, R. F. Nadhim, N. A. Yousif, and A. T. Hashim, "Medical Image Integrity Protection Through U-Net Based ROI Segmentation and Hybrid Integer Wavelet–Quadtree Embedding," Journal of Intelligent Informatics, Networking, and Cybersecurity, vol. 2, no. 1, p. 4, 2026.

[11] S. Rosaline and D. Paulraj, "Deep learning-based compression and encryption of CT images for secure telemedicine applications," Evolving Systems, vol. 16, no. 1, p. 29, 2025.

[12] H. Adnan, A. Obied, and A. Al-Fayadh, "Medical Image Compression Using Discrete Wavelet Transform," in Artificial Intelligence, Big Data, IOT and Block Chain in Healthcare: From Concepts to Applications, Y. Farhaoui, Ed., Cham: Springer Nature Switzerland, 2024, pp. 160–167.

[13] A. Hegde and S. Deepak, "LOSSY COLOR IMAGE COMPRESSION USING NON LINEARIZATION AND COLOR REDUCTION." [Online]. Available: http://www.esatjournals.org

[14] P. K. Sran, S. Gupta, and S. Singh, "Segmentation based image compression of brain magnetic resonance images using visual saliency," Biomed. Signal Process. Control, vol. 62, p. 102089, 2020.

[15] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," Comput. Geosci., vol. 10, no. 2–3, pp. 191–203, 1984.

[16] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for interactive imaging," Proceedings of the IEEE, vol. 90, no. 8, pp. 1336–1357, 2002.

[17] Y. Huang, Q. Tang, P. Chen, G. J. Chen, and P. Shen, "WTUNet: minimize your U-Net through wavelet transformation for lung nodules segmentation," in Tenth International Conference on Multimedia and Image Processing (ICMIP 2025), SPIE, 2025, pp. 30–36.