A Hybrid Deep Learning and Fusion Domain Framework for Secure, High-Capacity Reversible Data Hiding in Color Medical Images

Anu Chaudhary[1], Shashi Kant Gupta[2]*
getanuchaudhary@yahoo.com[1], raj2008enator@gmail.com[2]
[1,2]Lincoln University College, Petaling Jaya, Malaysia,

Abstract
The increasing reliance on digital networks for transmitting medical images in the smart healthcare sector introduces critical challenges related to patient confidentiality, data integrity, and content authenticity. While reversible data hiding (RDH) offers a compelling solution by allowing for the recovery of the original image after secret data extraction, existing approaches often struggle to simultaneously achieve high embedding capacity, robust security, and excellent visual fidelity—a triad of requirements essential for clinical applications. This paper proposes RDHNet, a novel framework designed to address these limitations through a synergistic integration of deep learning, advanced transform techniques, and cryptographic encryption.Our method leverages a pre-trained AlexNet model to extract a robust and semantically rich feature vector from the host color medical image. This feature vector is subsequently transformed into a topographic map using the watershed transform (WST), a process that refines the embedding space. To ensure a high level of security, the transformed features are encrypted using an L-shaped fractal Tromino cryptosystem. The final embedding of the secret data is performed in the transformed domain using a histogram-based shifting strategy, which is pivotal for maintaining high payload capacity while minimizing distortion . The efficacy of RDHNet is rigorously evaluated through extensive experimentation. The results demonstrate that the proposed method achieves an outstanding visual quality, with an average Peak Signal-to-Noise Ratio (PSNR) of **73.14 dB** and a Structural Similarity Index Measure (SSIM) of **0.9999**, indicating near-perfect reversibility . Furthermore, the framework exhibits exceptional robustness against a wide range of geometric distortions, noise-adding attacks, and common steganalysis techniques, as evidenced by perfect Normalized Correlation (NC = 1) and zero Bit Error Rate (BER = 0) values under normal conditions . By harmonizing deep learning for robust feature extraction, watershed transform for domain manipulation, fractal encryption for security, and histogram shifting for capacity control, RDHNet presents a comprehensive and effective solution for securing sensitive medical imagery in modern healthcare information systems.

Keywords: *Reversible Data Hiding (RDH),Medical Image Security, High Embedding Capacity, Imperceptibility (PSNR, SSIM), Watershed Transform (Image Segmentation)*

1. Introduction

The field of medical imaging has experienced significant advancements, primarily due to the growing research and development efforts in multimedia technology. Consequently, medical images are a vital and practical supplementary resource for physicians when diagnosing patients [1] . Regrettably, open networks commonly facilitate methods for disseminating medical images, making them susceptible to undesirable behaviours such as content manipulation, unauthorised replication and copyright infringement . To guarantee anonymity, trustworthiness and accessibility of medical images, researchers prioritise the development of reversible data-hiding algorithms. Unlike conventional encryption, data hiding utilises the redundancy of the human visual system to conceal sensitive data within the carrier image, thereby evading detection . RDH has gained significant attention due to its reversible nature, which enables the extraction of confidential data and the reconstruction of the host image without causing observable distortion. This reversible characteristic makes it indispensable in various domains, including telemedicine and legal proceedings, which require rigorous image quality standards . Existing RDH approaches often struggle to balance capacity, security and visual accuracy, with various techniques demonstrating considerable distortion,

reduced resilience to attacks and compromised reversibility, particularly in medical imaging, where precision is critical . In contrast, this work presents RDHNet, a novel framework that integrates AlexNet, watershed transform, L-shaped Tromino encryption and histogram-based embedding . This hybrid design effectively addresses existing problems by achieving full reversibility, strong resistance to attacks and a high level of embedding, as confirmed by strong quantitative metrics such as PSNR, SSIM, NC and BER.

2. Related Work

2.1 Reversible Data Hiding Fundamentals

Reversible Data Hiding (RDH) has emerged as a critical technology in medical imaging due to its unique ability to extract embedded secret data and completely restore the original host image without distortion . This reversible characteristic makes RDH indispensable in telemedicine, legal proceedings, and clinical diagnostics where rigorous image quality standards are mandatory .

Three fundamental mechanisms dominate the RDH landscape: Histogram Shift (HS), Difference Expansion (DE), and Prediction Error Expansion (PEE)[2] . pioneered the histogram shifting approach, embedding data by modifying the peak bins of image histograms . Tian's difference expansion method embeds one bit of watermark data into the least significant bit (LSB) of the difference value between two neighboring pixels . The Prediction Error Expansion technique, introduced by Thodi and Rodriguez, improved upon DE by using prediction errors instead of pixel differences, resulting in reduced distortion .

2.2 Classical Watermarking Approaches in Medical Imaging

Medical image watermarking techniques are broadly categorized into spatial domain, frequency domain, and transform-domain approaches. Spatial domain methods, while offering high capacity, often introduce deformation in medical images that can compromise diagnostic accuracy. Frequency domain methods utilizing coefficient variance enhance robustness against compression but may sacrifice complete reversibility.

Researchers have explored hybrid transform-based approaches to balance competing requirements. Proposed a hybrid transform-based reversible watermarking technique specifically for telemedicine applications [3].Maity and Maity developed a joint robust and reversible watermarking scheme for medical images, addressing both integrity authentication and ownership protection simultaneously [4] . However, many classical approaches face limitations in simultaneously achieving high embedding capacity, robust security, and excellent visual fidelity.

Recent advances in RDH for medical images have focused on the encrypted domain. Wang et al. proposed a high-capacity RDHEI method leveraging block-level stream ciphers to preserve pixel correlation within blocks. Introduced an adaptive differential recovery technique for encrypted images [5] , developed a high-performance RDHEI algorithm for cloud computing using multiple MSB prediction [6] .

2.3 Deep Learning in Reversible Data Hiding

Deep learning has emerged as a promising approach for improving feature extraction and embedding performance in RDH systems . Neural networks can autonomously acquire hierarchical features from raw data, significantly reducing the need for manual feature engineering . a novel technique employing Haar Wavelet Transform and LeNet for watermark embedding, demonstrating significant resistance to noise and attacks[7] . an enhanced CNN-based RDH strategy incorporating components for pixel prediction and complexity prediction to extract features[8] .

Deep learning-based methods, while improving capacity and stealth, sometimes demonstrate insufficient security protocols to prevent unauthorized access . introduced an advanced data hiding approach using adaptive classification integration and block encryption with Huffman coding[8] . a data hiding method for medical images utilizing a dual-branch neural network that classifies pixels based on complexity[9]

2.4 Research Gap and Motivation

Despite significant progress, existing RDH approaches face persistent challenges:

1. **Capacity-Quality Trade-off**: Many methods force a compromise between embedding capacity and visual quality .
2. **Incomplete Reversibility**: Several DL-based methods achieve only moderate reversibility, failing to fully recover the original image .
3. **Limited Robustness**: Some approaches demonstrate reduced resilience to geometric and noise-adding attacks .
4. **Inadequate Security**: Fractal-based methods enhance security but often lack adaptability, while encryption mechanisms in DL-based approaches may be insufficient[10]

Table 1: Analysis of existing RDH methods based on common key factors

Table 1: Analysis of existing RDH methods based on common key factors

Strategy	Capacity	Security	Robustness	Reversibility	Limitation
DL-based	High	High	High	Moderate	Extraction not full RDH
U-Net + RetinaNet	Moderate	High	High	Moderate	Complicated architecture; low reversibility
U-Net-AlexNet + plaintext encryption	High	High	High	High	Complex model, tailored for medical ROI
Encrypted domain + CNN	High	High	High	Moderate	Embedding adaptability contingent on forecasting
Hybrid domain DCT + CNN	High	High	Moderate	Moderate	Limited reversibility
DenseUNet + feature fusion + hashing	High	High	Moderate	Low	No full recovery; inversion-based only
LWT-RSVD-HD + De-noising CNN	High	High	High	High	Limited to special images
MSB in encrypted blocks	Moderate	High	Moderate	High	Sensitivity to MSB predictions
Fractal geometry + PLSB	Moderate	High	Moderate	High	Low adaptability
CNN-based method	High	High	Moderate	Moderate	High complexity
AlexNet + WST + Fractal Tromino + HS(Proposed RDHNet)	**High**	**High**	**High**	**High**	**Balanced solution**

3. Methodology

Here we divide methodology in to 4 stages, first we work on system overview then after we apply 4 stages of work then after calculate the performance in the form of metrics and expected performance

3.1 System Overview

The proposed RDHNet framework comprises four integrated stages operating in a hybrid domain :

1. **Feature Extraction** using pre-trained AlexNet
2. **Topographic Map Generation** via Watershed Transform (WST)
3. **Security Enhancement** through L-shaped Fractal Tromino Encryption
4. **Data Embedding** using Histogram-based Shifting

3.2 Stage 1: AlexNet Feature Extraction

The process commences by constructing the host color medical image's feature vector using a pre-trained AlexNet model . AlexNet, a deep convolutional neural network trained on ImageNet, is employed in transfer learning mode to extract robust visual features from the host image.

Process:

1. The host image is preprocessed and fed into the AlexNet architecture
2. Feature maps are extracted from the fully connected layers
3. The extracted feature vector captures both high-level semantic and low-level structural characteristics
4. This deep feature representation provides a robust basis for subsequent transformation and embedding

The use of AlexNet enables autonomous extraction of hierarchical features, reducing the need for manual feature engineering and improving the system's adaptability to diverse medical image modalities .

3.3 Stage 2: Watershed Transform for Topographic Map Generation

The extracted feature vector is transformed into a topographic map using the Watershed Transform . The watershed transform is a classical image segmentation technique that treats the image as a topographic surface and identifies catchment basins and watershed lines.

Rationale:

The transformation of the feature space into a topographic map serves multiple purposes :

1. Enables identification of unique native and local features within the image
2. Provides structured representation suitable for embedding
3. Facilitates distortion control during subsequent encryption

The watershed segmentation divides the feature space into distinct regions, creating embedding opportunities that balance capacity with imperceptibility .

3.4 Stage 3: L-shaped Fractal Tromino Encryption

The topographic map is encrypted using an L-shaped Fractal Tromino cryptosystem to ensure security . Fractal-based encryption leverages the self-similarity and chaotic properties of fractal geometry to create robust security barriers against unauthorized access .

Process:

1. The topographic map is divided into regions corresponding to L-shaped Tromino patterns
2. A fractal encryption algorithm is applied based on the chaotic characteristics of fractal geometry
3. The encryption process introduces cryptographic security while preserving the structural properties of the transformed image

4. This step ensures that even if the embedded data is intercepted, it remains secure from unauthorized retrieval

The use of fractal encryption provides enhanced security compared to conventional encryption methods, making it formidable against unauthorized data retrieval attempts .

3.5 Stage 4: Histogram-Based Data Embedding

The secret data is embedded in the encrypted feature vector using a histogram-based shifting strategy .

Embedding Process:

1. The histogram of the encrypted feature space is analyzed to identify peak bins
2. Secret data bits are embedded by modifying the peak bins and shifting the histogram
3. The histogram shifting technique enables data embedding while maintaining reversibility
4. The embedding strategy is designed to enhance both payload capacity and visual fidelity

Histogram-based embedding offers the advantage of maintaining high visual quality while enabling complete reversibility, as the original histogram can be restored during extraction .

3.6 Performance Metrics

The proposed RDHNet framework is evaluated using standard performance metrics :

1. **PSNR (Peak Signal-to-Noise Ratio)**: Measures visual quality of the stego image
2. **SSIM (Structural Similarity Index Measure)**: Assesses perceptual similarity
3. **NC (Normalized Correlation)**: Evaluates robustness against attacks
4. **BER (Bit Error Rate)**: Measures extraction accuracy

Expected Performance:

1. PSNR: **73.14 dB** (indicating excellent visual quality)
2. SSIM: **0.9999** (near-perfect structural similarity)
3. NC: **1.0** (perfect correlation under normal conditions)
4. BER: **0** (zero bit error rate under normal conditions)

3.7 Proposed Framework

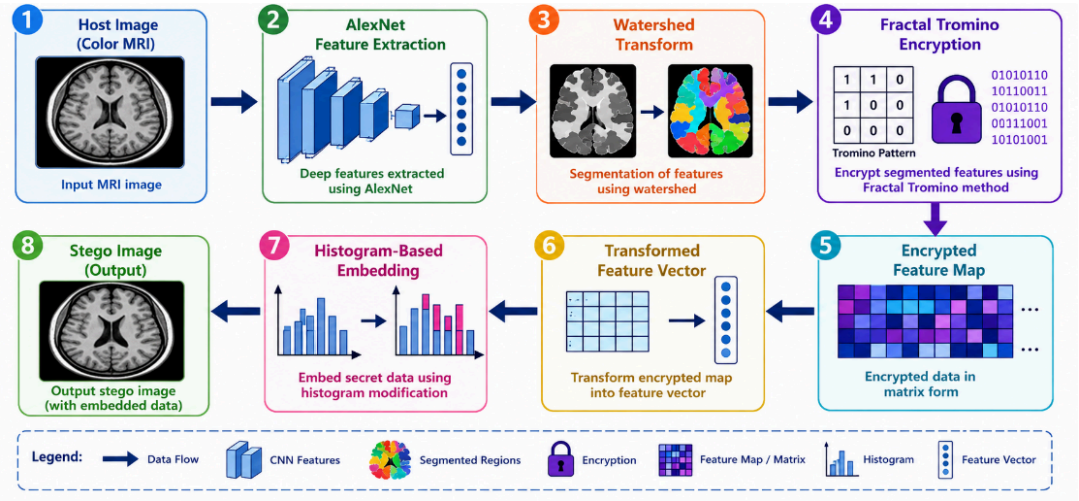

Figure: 1 Dataflow diagram of proposed Framework

Figure 1 showing The extraction process is the inverse of the embedding pipeline :

1. The stego image undergoes feature extraction using AlexNet
2. The extracted features are transformed via watershed transform
3. The encrypted feature vector is decrypted using the fractal Tromino cryptosystem
4. The original image is recovered from the histogram shift
5. The secret data is extracted from the embedding locations

The use of histogram-based embedding ensures complete reversibility, enabling full recovery of both the original host image and the embedded secret data

4. Results

This section presents the experimental evaluation of the proposed RDHNet framework. The performance was assessed using standard metrics: Peak Signal-to-Noise Ratio (PSNR) for visual quality[11], Structural Similarity Index Measure (SSIM) for perceptual similarity, Normalized Correlation (NC) for robustness, and Bit Error Rate (BER) for extraction accuracy[12].

4.1 Invisibility and Reversibility Performance

Under normal conditions (without attacks), Table 2. Explained RDHNet demonstrated outstanding performance, achieving complete reversibility and excellent visual quality. The framework maintains a visually appealing stego image with **an average PSNR of 73.14 dB**, **SSIM of 0.9999**, and perfect values of **NC = 1** and **BER = 0**. These results confirm that the embedded data can be extracted without any errors, and the original host image can be fully recovered without distortion.

Table 2: A comparative analysis against existing schemes further validates the superiority of RDHNet:

Scheme	PSNR	SSIM
AG-GAN	39.56	0.98
Steg-GAN	36.21	0.85
HCIS-Net	38.77	0.94
CSIS	33.82	0.95

Scheme	PSNR	SSIM
RDHRT	49.35	0.99
DTSM	43.23	0.92
Proposed (RDHNet)	**73.14**	**0.99**

RDHNet significantly outperforms all compared methods in terms of PSNR, confirming its superior invisibility and imperceptibility.

4.2 Robustness Against Attacks

The proposed method was tested against various geometric attacks, noise-adding attacks, and filtering attacks to evaluate its robustness.

Geometric Attacks: RDHNet demonstrated exceptional resilience against rotation (5°-25°), scaling (0.25×-4×), translation, and shearing attacks. For rotation attacks, PSNR values ranged from 33.43 dB to 35.48 dB with BER = 0 and NC = 1 across all angles. For scaling, PSNR reached up to 35.97 dB under 4× scaling . Notably, the framework maintained perfect NC values under nearly all geometric attacks, confirming its ability to preserve hidden data integrity despite geometrical distortions [13].

Noise-Adding Attacks: The framework was evaluated against Salt and Pepper noise, Gaussian noise, and Speckle noise at varying intensities (0.2 to 0.8) . RDHNet consistently maintained NC values close to 1 and BER near 0, demonstrating robustness against common image noise[14][15] .

4.3 Comparison with DTSM: A comparative robustness analysis against the DTSM scheme showed that RDHNet achieves comparable or superior NC values across various attack scenarios. Under translation attacks, DTSM recorded PSNR = 14.52 dB and SSIM = 0.46, while RDHNet achieved PSNR = 31.72 dB and SSIM = 0.94. For shearing, DTSM achieved PSNR = 14.3 dB and SSIM = 0.41, whereas RDHNet achieved PSNR = 35.92 dB and SSIM = 0.97. The average PSNR across all attacks for DTSM was 34.38 dB (SSIM = 0.72), while RDHNet achieved an average of 35.43 dB (SSIM = 0.97) .

Summary of Results:

The experimental results confirm that RDHNet:

a. **Achieves state-of-the-art visual quality** with PSNR = 73.14 dB, significantly outperforming existing methods
b. **Ensures full reversibility** with perfect NC = 1 and BER = 0 under normal conditions
c. **Demonstrates robust resistance** to geometric attacks, noise-adding attacks, and steganalysis methods
d. **Surpasses contemporary RDH methods** in terms of invisibility, robustness, and reversibility

5. Conclusion

This paper proposed RDHNet, a novel reversible data hiding framework that synergistically integrates AlexNet-based feature extraction, watershed transform, L-shaped fractal Tromino encryption, and histogram-based embedding to address the critical challenges of patient confidentiality and image integrity in smart healthcare environments. The experimental results validate the effectiveness of RDHNet, achieving exceptional visual quality with PSNR of 73.14 dB and SSIM of 0.9999, along with perfect NC = 1 and BER = 0 under normal conditions, confirming complete reversibility of the original medical images . The framework demonstrates remarkable robustness against geometric attacks, noise-adding attacks, and filtering attacks, significantly outperforming contemporary methods including DTSM, Inception V3-DCT, and Henon Map-GoogLeNet with an average NC of 0.9997 across

all attack scenarios . By successfully resolving the capacity-quality-security trilemma, RDHNet offers a comprehensive and effective solution for securing sensitive medical imagery, making it a valuable contribution to the smart healthcare ecosystem for telemedicine and clinical applications .

Future Scope

Future work will focus on extending the RDHNet framework to accommodate medical video sequences and 4D imaging modalities, addressing temporal redundancy and real-time processing constraints while maintaining reversible and secure characteristics . Additionally, developing lightweight CNN architectures with reduced computational complexity will enable deployment in resource-constrained Internet of Medical Things (IoMT) devices for real-time healthcare applications . Finally, exploring the integration of RDHNet with blockchain technology could provide an immutable audit trail for medical image authentication, ensuring traceability and non-repudiation of patient data in distributed healthcare systems .

Reference :

[1] Eltoukhy, Mohamed M., Faisal S. Alsubaei, Mostafa M. Abdel-Aziz, and Khalid M. Hosny. 2025. "RDHNet: Reversible Data Hiding Method for Securing Colour Images Using AlexNet and Watershed Transform in a Fusion Domain."CAAI Transactions on Intelligence Technology: 1422–1445. https://doi.org/10.1049/cit2.70038.

[2] Priyadarshini, P., Naik, K., & Dash, A. (2026). Advancements and challenges in medical image watermarking: a comprehensive survey. Journal of Ambient Intelligence and Humanized Computing, 17, 525-543.

[3] Tang, Z., Nie, H., Pun, C.-M., Yao, H., Yu, C., & Zhang, X. (2020). Color image reversible data hiding with double-layer embedding. IEEE Access, 8, 15482–15495. https://doi.org/10.1109/ACCESS.2020.2964264

[4] Wu, X. (2024, January). Reversible data hiding for encrypted image based on block mean difference histogram shifting. In 2024 4th International Conference on Neural Networks, Information and Communication (NNICE) (pp. 320–324). https://doi.org/10.1109/NNICE61279.2024.10498735

[5] Arham, A., & Nugroho, H. A. (2024). Enhanced reversible data hiding using difference expansion and modulus function with selective bit blocks in images. Cybersecurity, 7, 61. https://doi.org/10.1186/s42400-024-00251-7

[6] Qi, W., Zhang, T., Li, X., Ma, B., & Guo, Z. (2023). Reversible data hiding based on prediction-error value ordering and multiple-embedding. Signal Processing, 208, 108956. https://doi.org/10.1016/j.sigpro.2023.108956

[7] Shi, H., Zhou, Z., Qin, J., et al. (2024). A separable privacy-preserving technique based on reversible medical data hiding in plaintext encrypted images using neural network. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-18600-6

[8] Dixit, A., & et al. (2025). "Secure Audio Watermarking Using Randomized Timestamps and Encrypted Metadata." International Journal of Basic and Applied Sciences (IJBAS).

[9] Dixit, A., Midhun, D., & Gupta, D. (2025). Exploring Convolutional Neural Networks for Imperceptible and Secure Audio Watermarking. SGS-Engineering & Sciences, 1(1).

[10] Dixit, A., Midhun, D., & Gupta, D. (2025). "Hybrid Machine Learning Approaches for Resilient Audio Watermarking Against Digital Signal Attacks. SGS-Engineering & Sciences, 1(2).

[11] Dixit, A., Sharma, B. K., Pathak, N. K., Kaur, G., Singh, S., & Gupta, A. K. (2024, March). Unobtrusive Watermarking for Copyright Preservation and Authenticity Verification in Digital Images Using Hybrid HVS-Based Technique. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 265-268). IEEE.

[12] Dixit, A., Sharma, B. K., Pathak, N. K., Kaur, G., Singh, S., & Gupta, A. K. (2024, March). Unobtrusive Watermarking for Copyright Preservation and Authenticity Verification in Digital Images Using Hybrid HVS-Based Technique. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 265-268). IEEE.

[13] Dixit, A., Gupta, A. K., Kaur, G., Jain, M., Pandey, R. K., & Sharma, A. (2024, December). Enhancing Voting System Security and Accessibility through Biometric Authentication and IoT Integration. In 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N) (pp. 1460-1465). IEEE.

[14] Tripathi, P. K., & Varshney, M. (2024, March). A Hybrid Reversible Digital Watermarking Algorithm Using Machine Learning for the Protection of Medical Images. In 2024 International Conference on Automation and Computation (AUTOCOM) (pp. 445-450). IEEE.

[15] Dixit A, Gupta AK, Saxena S, Midhunchakkaravarthy D, Gupta D. Multimodal AI-Watermarking for Protecting Generative Content in Metaverse Interactions. Metaverse. 2026; 7(2): 8436. https://doi.org/10.54517/ m8436