International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

SecureDICOM: Dual Watermarking for ROI-Aware Medical Image Authentication and Sharing

Mrs. B. Jyothi[1], Mohan Siva Sanjay Bhimaneni[2]
[1]Assistant Professor(c), [2]Student
Department of Computer Science and Engineering
University College of Engineering Narasaraopet, JNTUK
Narasaraopet, Andhra Pradesh, India.

***Abstract*—Clinical image authentication must meet three concurrent requirements: imperceptibility, reliable ownership verification, and protection of diagnostically critical regions. SecureDICOM addresses these needs through a dual-layer strategy that combines robust frequency-domain watermarking with fragile block-level integrity checks under ViT-guided ROI handling and RBAC-controlled backend execution. The pipeline is realized as five embedding stages and four extraction stages, with selective writeback preserving ROI pixels. On the SIIM CT benchmark used in this study [19], the method achieves 74.31 dB mean PSNR, near-unity SSIM, and 99.28% average blind extraction confidence. These outcomes indicate that SecureDICOM is a practical choice for clinical, provenance-aware image sharing.**

***Index Terms*—medical image watermarking, DICOM security, blind extraction, fragile watermarking, ROI-aware embedding, tamper localization, ViT segmentation, healthcare image authentication**

I. Introduction

Digital imaging underpins modern diagnosis, telemedicine, and follow-up care. As DICOM studies move across PACS/HIS boundaries, cloud services, and research exchanges, risks increase around untracked reuse, undetected edits, and unclear ownership history. Cryptographic controls are essential for storage and transmission, but they do not always provide post-decryption provenance and localized integrity evidence during routine clinical use.

Watermarking can fill that gap, but the medical setting imposes stricter conditions than consumer multimedia. Diagnostic ROI regions must be preserved, perceptual distortion must remain negligible, and verification outputs must be trustworthy enough for operational decisions. Many existing methods emphasize either robustness or tamper sensitivity, and many assume manual ROI delineation.

SecureDICOM is designed to close this integration gap. It combines ROI-aware robust ownership embedding, fragile tamper localization, and deployment-focused backend controls in one workflow. ViT-based ROI estimation, blind robust extraction, and RBAC-governed APIs are treated as parts of one system rather than separate components.

A. Main Contributions

1) **Dual-layer watermarking architecture**: A phase-wise framework that unifies FFT-domain robust ownership embedding with block-level fragile integrity marking under one ROI-constrained pipeline.
2) **ViT-assisted ROI preservation**: Integration of Vision Transformer feature extraction and clustering for automatic clinically sensitive region identification, enabling pixel-level protection during embedding and extraction.
3) **Blind robust extraction with voting**: Redundancy-based majority decoding for confident ownership recovery without requiring the original image, coupled with extraction confidence metrics suitable for clinical decision support.
4) **Block-wise fragile verification**: Cryptographic digest-based tamper localization at block granularity, enabling clinicians to visualize and interpret suspected modification regions.

II. Related Work

A. Transform-Domain Robust Watermarking

Robust medical watermarking has largely evolved around transform-domain embedding. Foundational work such as Cox et al. [2] demonstrated spread-spectrum strategies in frequency space. Later DFT [1], DWT, and DCT methods placed payload energy in carefully selected bands to balance invisibility and resistance to compression or mild signal perturbation. Recent MD-FFT results [1] further indicate strong robustness with high visual fidelity on medical imagery.

Hybrid and SVD-augmented approaches [10] improved stability by mixing spatial and transform constraints. Even so, many pipelines still treat ROI preservation as a secondary add-on and rarely couple robust embedding with a dedicated fragile verification channel.

B. ROI-Aware Segmentation and Protection

Early ROI-aware methods (for example, Zain and Clarke [4]) relied on manual annotation or hand-crafted rules, often approximating ROI by excluding background. With deep learning, U-Net-style models [11] became the default for medical segmentation tasks.

Transformer-based models, including UNETR [13] and TransUNet [14], improved global context modeling and boundary consistency. They enable data-driven ROI detection with less manual intervention. However, full segmentation models are still infrequently embedded into end-to-end watermarking systems, where many papers continue to assume fixed masks.

In this work, ViT is used as a feature extractor for unsupervised ROI clustering rather than full dense semantic labeling.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

TABLE I: Comparison of SecureDICOM with Related Work

Method/Aspect	Transform	Blind	ROI	Robust	Fragile	Deploy
Cox Spread-Spectrum Model [2]	Spread Spectrum	Yes	No	Yes	No	No
Zain RONI Model [4]	Spatial	No	Yes	No	Yes	Limited
DWT variants	DWT	Yes	No	Yes	No	Limited
SVD hybrids [10]	SVD+Spatial	Yes	No	Yes	No	Limited
QDFT [7]	QDFT	No	No	No	Yes	Limited
MD-FFT (2025) [1]	MD-FFT	Yes	Yes	Yes	No	API-level
FFT + ROI-Aware (Proposed)	**FFT**	**Yes**	**Yes**	**Yes**	**Yes**	**Full**

This design keeps the pipeline lightweight while retaining anatomically meaningful masking behavior for watermarking.

C. Fragile Watermarking and Tamper Detection

Fragile watermarking complements robust ownership embedding by prioritizing edit sensitivity. Classical fragile schemes [3] intentionally break under modification, producing a clear tamper/no-tamper signal. Semi-fragile variants relax this behavior to tolerate benign operations while still surfacing meaningful edits.

Block-digest methods [7] strengthen interpretability by hashing local regions (e.g., MD5 or SHA-256) and storing digest evidence through carrier mapping or LSB placement. Verification then compares extracted and recomputed digests per block, yielding spatial tamper cues rather than only a global verdict.

The main open issue is calibration: thresholds must reduce false alarms from benign transformations without missing subtle malicious changes. SecureDICOM retains this challenge as an explicit limitation and future optimization target.

D. Secure Deployment and DICOM Governance

Clinical adoption requires more than algorithmic quality. Practical systems must fit HIS/PACS workflows, remain standards-compliant with DICOM [15], and support governance expectations such as HIPAA [18] and RBAC [16]. SecureDICOM addresses this by coupling watermark operations with role-gated APIs and audit-ready verification artifacts.

E. Summary of Gaps This Paper Addresses

Table I highlights the central gap addressed here: combining blind extraction, ROI awareness, robust ownership recovery, and fragile tamper evidence inside one deployable architecture.

III. Fundamental Concepts

A. DICOM Security Context

Digital Imaging and Communications in Medicine (DICOM) [15] defines how medical images and associated metadata are stored and exchanged. A valid watermarking workflow must preserve this structure: the file must remain parsable, metadata must stay intact, and pixel data must remain clinically interpretable.

B. ROI and RONI Separation

Let $I(x, y)$ denote the grayscale image intensity at pixel (x, y), and $M(x, y)$ a binary mask:

$$M(x,y) = \begin{cases} 1, (x,y) \in \Omega_{\text{ROI}} \\ 0, (x,y) \in \Omega_{RONI} \end{cases} \quad (1)$$

Here, Ω_{ROI} and Ω_{RONI} denote, respectively, the sets of diagnostically relevant and non-diagnostic pixel coordinates; ROI pixels are protected while RONI pixels are candidate embedding locations. The watermarking constraint is:

$$I'(x,y) = \begin{cases} I(x,y), if\ M(x,y) = 1 \\ f_{embed}(I(x,y), W), if\ M(x,y) = 0 \end{cases} \quad (2)$$

where I' is the watermarked image and f_{embed} is the embedding function parameterized by watermark payload W.

C. Frequency-Domain Embedding Principle

For a patch $P \in \mathrm{R}^{n \times n}$, the 2D discrete Fourier transform is:

$$F(u,v) = \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} P(x,y) e^{-j2\pi(ux+vy)/n} \quad (3)$$

In this expression, P (x, y) is the spatial-domain pixel value at location (x, y), F (u, v) is the complex Fourier coefficient at frequency index (u, v), and n is the patch width/height. The coefficient F (0, 0) is the DC component, which represents the average intensity (global brightness) of the patch; this project avoids embedding in this component to reduce visible bias and preserve luminance stability.

This project modifies selected mid-band frequency magnitudes (frequency indices 2–6 after DC component) according to payload bits, then applies inverse FFT to reconstruct the patch:

$$P'(x,y) = \text{IFFT}\big(F'(u,v)\big) \quad (4)$$

Bit repetition (redundancy factor = 3) and majority voting during extraction improve robustness to noise and compression.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

D. Fragile Digest Localization

For each block B_i, compute a cryptographic digest (SHA-256):

$$g_i = \text{SHA256}(B_i) \tag{5}$$

Embed g_i using LSB carrier mapping over non-ROI pixels. During verification, recover the digests and compare:

$$\delta_i = I(g_i \neq \hat{g}_i \tag{6}$$

The tamper map is $\Delta = \{\delta_i\}$, indicating which blocks were modified.

IV. Proposed Method

A. System Architecture Overview

The SecureDICOM architecture is shown in Figure 1. The system consists of three major components:

1) **Frontend**: DICOM input processing and user interface for embedding/verification requests.
2) **Core Watermarking Engine**: Dual-layer embedding/extraction operations integrated with ROI segmentation.
3) **Backend Governance**: RBAC, audit logging, secure storage, and API routing.

Authentication is handled in a stateless manner using JWT claims (e.g., role and organization). Each watermarking request is authorized against policy, and denied actions are recorded for auditability.

B. Phase 1: DICOM Preparation

1) Load DICOM file and extract pixel array.
2) Normalize pixel values to floating-point range [0, 1] using percentile-based scaling (2nd and 98th percentiles) to suppress outliers.
3) Validate image is 2D grayscale; handle 3D/multi-frame by selecting middle slice.
4) Store original bit-depth and dynamic range for post-processing.

C. Phase 2: ROI Estimation via ViT

1) Extract ViT patch embeddings (768-dim) from input image.
2) Perform unsupervised clustering (K-means, $k = 2$) on embedding space to separate foreground (ROI) from background (RONI).
3) Generate binary mask and apply morphological closing to remove small holes.
4) Validate mask for clinical plausibility (ROI ratio in expected range for medical images, typically 30–70%).

For this implementation, ViT backbones were fine-tuned on modality-specific medical resources, including MIDI-B checkpoints [20] and RIDER lung CT segmentation checkpoints [21]. The resulting features are clustered with a lightweight unsupervised stage to produce ROI masks.

D. Phase 3: Robust Watermark Embedding

Algorithm 1 outlines robust embedding:

Algorithm 1 Robust Watermark Embedding

1: **Input:** Image I, ROI mask M, payload W, strength s
2: **Output:** Watermarked image I'
3: $P_{\text{list}} \leftarrow$ partition(I, size=8, stride=8)
4: $P_{\text{eligible}} \leftarrow$ [patches with $< 100\%$ ROI coverage]
5: **for** each patch P in P_{eligible} **do**
6: $F \leftarrow \text{FFT}(P)$
7: **for each** bit $b \in W \times \text{REDUNDANCY} = 3$:
8: Select mid-band frequency $f \in [2, 6]$
9: $|F(f)| \leftarrow |F(f)| \cdot (1 + s \cdot b)$ // Amplitude modulation
10: $P' \leftarrow \text{IFFT}(F)$
11: **Selective writeback:** for (x, y) in P':
12: if $M(x, y) = 0$: $I'(x, y) = P'(x, y)$
13: else: $I'(x, y) = I(x, y)$ // Preserve ROI
14: **end for**
15: **return** I'

Key Innovation: Lines 11–13 enforce ROI protection via per-pixel selective writeback. Even if a patch overlaps ROI, only RONI pixels are updated.

E. Phase 4: Fragile Watermark Embedding

1) Partition watermarked image into 8×8 non-overlapping blocks.
2) For each block B_i, compute $g_i = \text{SHA256}(B_i)$.
3) Insert g_i through LSB carrier mapping only within the LSB plane of non-ROI pixels.
4) Store mapping metadata for verification.

Carrier Mapping Strategy: Use deterministic pseudo-random bit positions (seeded by block index) to scatter digest bits across the block, reducing detectability.

F. Phase 5: Secure Persistence and Access Control

1) Apply RBAC check: verify user role permits embedding operation.
2) Persist watermarked DICOM with metadata (watermark ID, timestamp, embedding strength, ROI mask hash).
3) Log embedding event (user ID, file hash, operation type, timestamp).
4) Return watermarks ID and verification codes to user for later extraction.

G. Extraction and Verification Pipeline

Blind extraction is performed without access to the pristine source image. The decoder revisits the embedding frequency bands, aggregates redundant bit observations across eligible patches, and resolves each payload bit via majority vote. A confidence value is then derived from vote consistency. In parallel, fragile verification recomputes local digests and compares them with stored or extracted references to generate a block-level tamper map.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

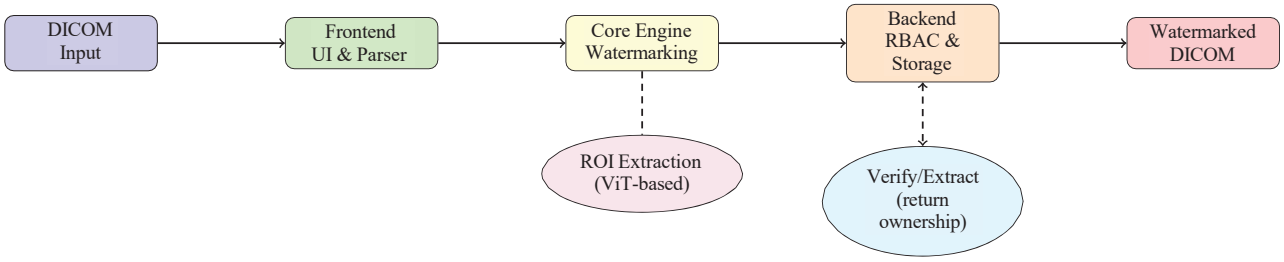

Fig. 1: SecureDICOM system architecture: DICOM input flows through frontend processing, core watermarking with ViT-based ROI extraction, backend governance (RBAC, storage), and outputs watermarked DICOM with metadata.

V. Experimental Results

A. Evaluation Setup

Evaluation was conducted on the SIIM Medical Images dataset [19]. For this study, 100 CT DICOM scans were sampled to capture variation in patient age (39–83 years), anatomical presentation, and acquisition protocol (including contrast and non-contrast studies).

B. Metrics and Methodology

1) Perceptual Quality: **Peak Signal-to-Noise Ratio (PSNR)**:

$$\text{PSNR} = 10 \log_{10} \frac{L^2}{\text{MSE}} \text{ dB} \quad (7)$$

where L is the dynamic range (typically 1 for normalized images) and MSE is mean squared error.

Structural Similarity Index (SSIM):

$$SSIM(I, I') = \frac{(2\mu_I\mu_{I'}+c)(2\sigma_{I,I'}+c_2)}{(\mu_I^2+\mu_{I'}^2+c_1)(\sigma_I^2+\sigma_{I'}^2+c_2)} \quad (8)$$

SSIM is more aligned with human perception than PSNR; values near 1.0 indicate imperceptibility.

1) Robust Extraction: Confidence score quantifies consistency of majority voting. Higher confidence indicates stable recovery despite image transformations.

$$\text{Confidence} = \frac{\text{Number of correctly decoded votes}}{\text{Total number of votes}} \quad (9)$$

For n payload bits with redundancy factor 3 and c consistent votes, confidence is computed as $c/(3n)$.

2) ROI Preservation: ROI Coverage Ratio:

$$\rho_{ROI} = \frac{|\Omega_{ROI}|}{H \, X \, W} \quad (10)$$

Extracted ROI areas are compared against ground-truth (ViT-based) masks.

1) Tamper Detection: Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

where TP = true positive tampering detections, FP = false positives. Synthetic tampering was simulated by XORing random bits in selected regions.

C. Results: Perceptual Quality

Across the 100-image SIIM cohort, perceptual distortion remains visually negligible, as reflected by high PSNR and SSIM values close to 1.0.

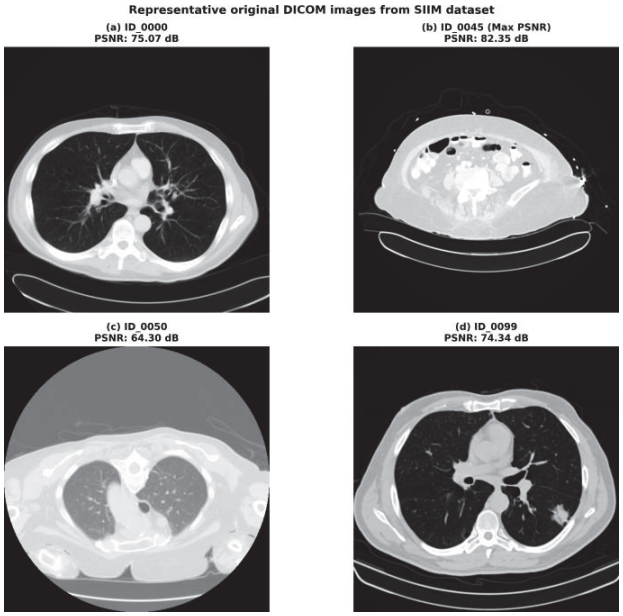

Fig. 2: Representative original DICOM images: (a) ID_0000, (b) ID_0045 (max PSNR), (c) ID_0050, (d) ID_0099. All show imperceptible watermarking visually.

D. Results: Robustness and Verification

The ROI module produced 54.98% mean coverage, and blind extraction reached 99.28% mean confidence (minimum 97.14%), indicating stable ownership recovery without source-image access. Under synthetic bit-flip tampering, fragile detection precision was 38.55%.

E. Results: ROI Extraction and Visualization

Qualitative examples show that extracted masks align with clinically relevant anatomy and preserve expected ROI boundaries prior to watermark insertion.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

TABLE II: Quantitative Results Summary (100 SIIM CT Images)

Metric	Mean	Std Dev	Min	Max
PSNR (dB)	74.31	5.49	58.99	82.35
SSIM	0.999998	0.000008	0.999896	1.0
Robust Extraction Confidence	0.9928	—	0.9714	1.0
ROI Coverage Ratio	0.5498	—	—	—
Tamper Detection Precision	0.3855	—	—	—

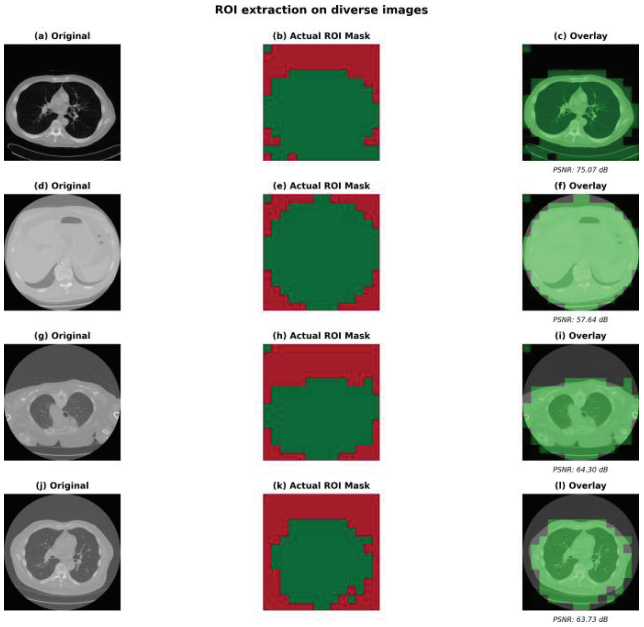

Fig. 3: ROI extraction on diverse images: (a, b, c) original, binary mask, and overlay for four representative samples, demonstrating ViT-based segmentation generalization.

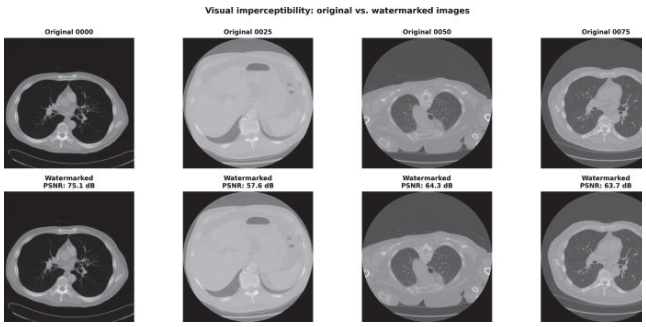

Fig. 4: Visual imperceptibility: original vs. watermarked images are indistinguishable, confirming transparent embedding even at high PSNR values.

F. Discussion

1) PSNR and SSIM Analysis: The mean PSNR is 74.31 dB, far above common clinical transparency targets (often around 40 dB). SSIM values close to 1.0 are consistent with negligible visual distortion, indicating that embedding remains visually imperceptible for diagnostic reading.

2) Robust Extraction Confidence: Blind extraction confidence remains high at 99.28% on average (97.14% minimum), which indicates consistent recovery under the tested conditions and supports dependable ownership checks in routine verification workflows.

3) ROI Segmentation Coverage: ViT-driven clustering produced an average ROI coverage of approximately 55%, which is consistent with expected clinically salient area proportions in this dataset. Because mask generation is based on unsupervised clustering over learned features, the method remains practical without per-study manual contouring.

4) Tamper Localization Limitations: Tamper localization is the weakest subsystem, with precision at 38.6%. This reflects the known trade-off between sensitivity to subtle edits and robustness against benign operations such as compression. The current evaluation uses synthetic tampering and would benefit from:

1) Evaluation on realistic attack models (e.g., lesion removal, measurement alteration).
2) Stronger block carrier mapping to reduce false positives.
3) Adaptive threshold tuning based on image contrast and texture.

This improvement area is explicitly identified for future work.

5) Clinical Applicability: From a deployment perspective, SecureDICOM addresses the following practical requirements:

1) **Transparency**: PSNR/SSIM imperceptibility ensures diagnostic accuracy is not compromised.
2) **Verifiability**: Blind extraction allows on-demand ownership verification without storing pristine references.
3) **Accountability**: RBAC integration supports compliance auditing and role-based operation policies.
4) **Graceful Degradation**: If watermark verification fails, the image remains clinically usable; watermarking is not a blocking gate.

VI. Limitations and Future Work

A. Limitations

1) **Tamper Precision**: Fragile embedding currently reaches 38.6% precision under synthetic tampering. Digest mismatches are sensitive to benign local intensity variation (compression, interpolation, and texture-heavy noise), which increases false positives.
2) **Computational Cost**: ViT-based ROI extraction and fragile digest operations add latency (about 2–5 seconds per image on GPU and 10–15 seconds on CPU). Workflow-level optimization is still required for high-throughput settings.
3) **Reversibility**: The watermark is not fully reversible after lossy compression. For example, JPEG processing

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

can prevent exact recovery of the original embedded signal, which is a known limitation of frequency-domain embedding.

4) **Multi-Layer Interaction**: Behavior of robust and fragile layers under stronger compression has not been fully characterized. Layer interaction may reduce extraction confidence in some settings.

B. Future Work

1) **Improved Fragile Embedding**: Explore algebraic reconstruction and information-theoretic carrier design to raise tamper precision while preserving imperceptibility.
2) **Real-World Attack Models**: Extend evaluation beyond JPEG artifacts to clinically relevant modifications, adversarial perturbations, and synthetic forgery scenarios.
3) **Hardware Acceleration**: Optimize ROI and watermarking kernels for near real-time deployment on medical imaging infrastructure.
4) **Integration with PACS**: Package SecureDICOM as a PACS-facing microservice with transparent ingest-time embedding and retrieval-time verification.
5) **Privacy-Preserving Verification**: Investigate zero-knowledge style verification so ownership can be validated without revealing payload contents.

VII. Conclusion

SecureDICOM presents a system-level approach to DICOM authentication that combines robust FFT-domain ownership watermarking, fragile block-level integrity evidence, and ViT-guided ROI preservation in one operational pipeline.

On 100 SIIM CT images, the method preserves imperceptible quality (74.31 dB mean PSNR and SSIM near 1.0), attains strong blind extraction confidence (99.28% mean), and delivers practical ROI masks (54.98% mean coverage). Integration with RBAC-enabled backend services further supports deployment beyond stand-alone algorithmic testing.

The primary remaining challenge is fragile tamper precision (38.6% under synthetic attacks). Closing this gap will require improved carrier mapping and broader validation under clinically realistic edit patterns.

Overall, this study contributes a reproducible and deployment-aware blueprint for medical image watermarking that bridges algorithm design and healthcare system constraints.

Appendix A
Abbreviations

TABLE III: List of Abbreviations

Abbreviation	Meaning
ROI	Region of Interest
RONI	Region of Non-Interest
DICOM	Digital Imaging and Communications in Medicine
FFT	Fast Fourier Transform
ViT	Vision Transformer
RBAC	Role-Based Access Control
PACS	Picture Archiving and Communication System
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index

References

[1] M. W. Abo El-Soud, M. M. Eltoukhy, M. M. Abdel-Aziz, A. Alourani, and K. M. Hosny, "Robust Blind Watermarking to Secure Color Medical Images Using Multidimensional-FFT Fusing LFSR-Encryption and LZW Compression," *IEEE Access*, vol. 13, pp. 46054–46069, 2025, doi:10.1109/ACCESS.2025.3547614.

[2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997, doi:10.1109/83.650120.

[3] I. J. Cox and J.-P. M. G. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 587–593, 1998, doi:10.1109/49.668971.

[4] J. M. Zain and M. Clarke, "Reversible region of non-interest (RONI) watermarking for authentication of DICOM images," *International Journal of Computer Science and Network Security*, vol. 7, no. 9, pp. 226–235, 2007.

[5] A. Soualmi, L. Laouamer, and A. Alti, "A blind watermarking approach based on hybrid imperialistic competitive algorithm and SURF points for color images authentication," *Biomedical Signal Processing and Control*, vol. 84, art. 105007, 2023, doi:10.1016/j.bspc.2023.105007.

[6] A. Soualmi, A. Benhocine, and I. Midoun, "Artificial bee colony-based blind watermarking scheme for color images alter detection using BRISK features and DCT," *Arabian Journal for Science and Engineering*, vol. 49, no. 3, pp. 3253–3266, 2024, doi:10.1007/s13369-023-07958-8.

[7] J. Ouyang, J. Huang, X. Wen, and Z. Shao, "A semi-fragile watermarking tamper localization method based on QDFT and multi-view fusion," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 15113–15141, 2023, doi:10.1007/s11042-022-13938-1.

[8] S. Mehraj *et al.*, "Spatial domain-based robust watermarking framework for cultural images," *IEEE Access*, vol. 10, pp. 117248–117260, 2022, doi:10.1109/ACCESS.2022.3217920.

[9] Y. Bai, L. Li, S. Zhang, J. Lu, and M. Emam, "Fast frequency domain screen-shooting watermarking algorithm based on ORB feature points," *Mathematics*, vol. 11, no. 7, art. 1730, 2023, doi:10.3390/math11071730.

[10] Q. Su, X. Zhang, and H. Wang, "A blind color image watermarking algorithm combined spatial domain and SVD," *International Journal of Intelligent Systems*, vol. 37, no. 8, pp. 4747–4771, 2022, doi:10.1002/int.22738.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.

[12] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.

[13] H. Hatamizadeh *et al.*, "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 574–584, doi:10.1109/WACV51458.2022.00063.

[14] J. Chen *et al.*, "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[15] DICOM Standards Committee, "Digital Imaging and Communications in Medicine (DICOM)," https://www.dicomstandard.org/.

[16] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *IEEE Computer*, vol. 29, no. 2, pp. 38–47, 1996, doi:10.1109/2.485845.

[17] M. Jones, J. Bradley, and N. Sakimura, "JSON Web Token (JWT)," RFC 7519, IETF, May 2015.

[18] U.S. Department of Health and Human Services, "Health Insurance Portability and Accountability Act (HIPAA)," https://www.hhs.gov/hipaa/index.html.

[19] SIIM, "SIIM Medical Images Dataset," https://www.kaggle.com/datasets/c1stcc/siim-medical-images, 2023.

[20] The Cancer Imaging Archive (TCIA), "MIDI-B," https://www.cancerimagingarchive.net/collection/midi-b-test-midi-b-validation/.

[21] The Cancer Imaging Archive (TCIA), "RIDER Lung CT," https://www.cancerimagingarchive.net/collection/rider-lung-ct/.