MedSecure: Watermarking Technique for Medical Images for Authentication and Copyright Protection

Archana Tiwari * ARCHANAT@IITBHILAI.AC.IN
Sudev Kumar Padhi* SUDEVP@IITBHILAI.AC.IN
Umesh Kashyap UMESHK@IITBHILAI.AC.IN
Sk. Subidh Ali SUBIDH@IITBHILAI.AC.IN
Department of Computer Science and Engineering, Indian Institute of Technology Bhilai, Durg, Chhattisgarh, India, 492002

Editors: Under Review for MIDL 2025

Abstract

The surge in telemedicine adoption has underscored the critical need for secure medical image transmission and storage. However, existing techniques struggle to balance imperceptibility, resilience to acceptable image manipulations, and robustness against adversarial threats. We propose a deep learning-based dual watermarking framework that embeds a perceptual hash for copyright protection and a cryptographic hash for integrity verification. By incorporating deep learning, our method ensures robustness against surrogate model and content-preserving attacks while preserving diagnostic fidelity. The experimental results demonstrate imperceptibility (PSNR: 40.23 dB, SSIM: 0.98) and with an accuracy of 95.4% against adversarial manipulations, which set a new benchmark for secure medical image authentication in telemedicine.

Keywords: Deep Learning, Watermarking, Copyright Protection, Authentication, Telemedicine.

1. Introduction

Telemedicine has revolutionized healthcare by enabling remote consultations and diagnoses, which are highly dependent on medical imaging for accurate evaluations. However, secure transmission and storage of medical images pose challenges, as unauthorized modifications and adversarial attacks threaten patient confidentiality and diagnostic reliability (Amrit et al., 2024). Digital watermarking offers a solution by embedding secure, imperceptible information within images while preserving diagnostic quality. Conventional watermarking remains vulnerable to distortion and overwriting. Recent deep learning-based methods improve robustness but are still susceptible to surrogate model attacks. To overcome these issues, we propose a deep learning-based dual watermarking technique for telemedicine. Our framework ensures both content authentication and copyright protection while countering surrogate model and overwriting attacks. It integrates perceptual and cryptographic hash-based watermarks within a deep learning framework, leveraging the sensitivity of cryptographic hashes for robust authentication.

* Contributed equally

2. Methodology

A perceptual hash (W_1) is first embedded to verify digital ownership, followed by a cryptographic hash (W_2) for integrity verification. These watermarks remain independent, ensuring robust extraction. The perceptual hash, calculated using a DCT-based method (Kalker et al., 2001), is transformed into a tensor via a neural network and embedded in the medical image (I) using an encoder (E_1), producing a visually similar watermarked image (W_1). To further ensure authenticity, a SHA-3 cryptographic hash is calculated from the latent vector of I_1, converted into an image, and embedded via a second encoder (E_2), resulting in the final watermarked image (I_2) (Tancik et al., 2020).

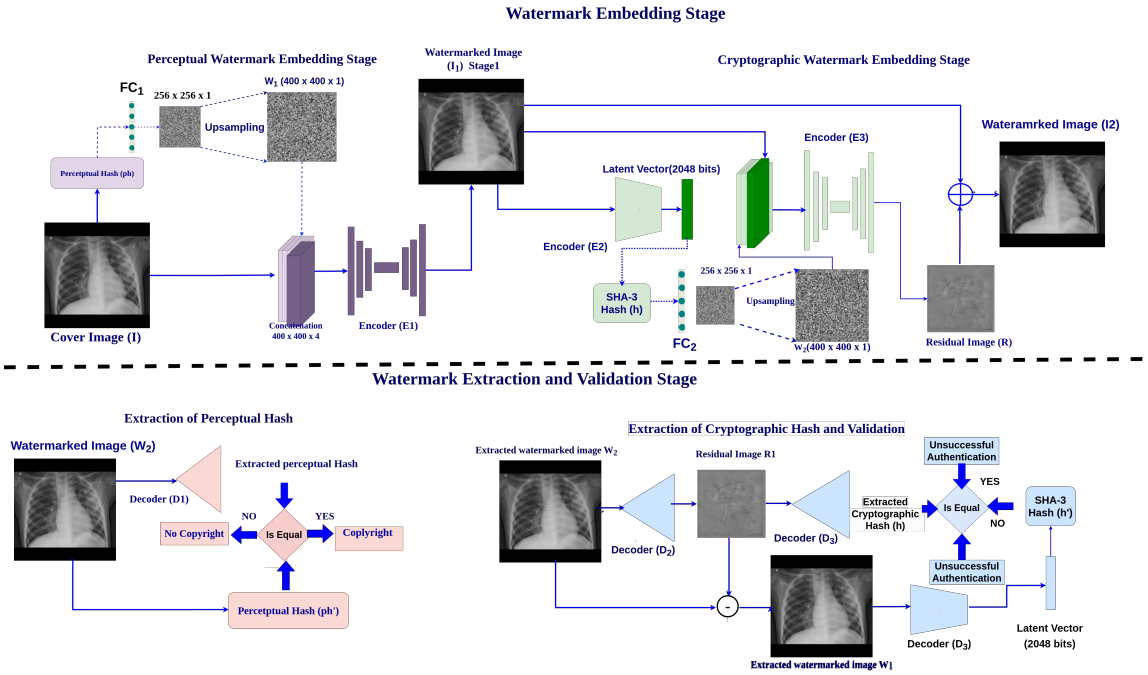

Figure 1: Overview of the proposed Deep Learning-based Dual Watermarking for medical image authentication and copyright protection.

During extraction, the cryptographic hash is recovered and verified using decoders (D_2, D_3), while the perceptual hash validates image integrity and source authentication. If both hashes match, the medical image is authenticated for secure telemedicine transmission, ensuring its integrity and provenance. The detailed methodology is illustrated in Figure 1. This approach improves tamper resistance, ensures reliable source verification, and strengthens the security of remote healthcare diagnostics.

3. Experimental Results

To evaluate the effectiveness of our approach through experiments on diverse medical datasets, we used MedIMeta (Woerner et al., 2024), which encompasses 19 medical imaging datasets in 10 distinct domains. The model was trained in 80 k images and tested on 40 k images, ensuring a comprehensive evaluation of different medical imaging modalities. The

evaluation focused on three key metrics: Peak Signal-to-Noise Ratio ($PSNR$) for watermark imperceptibility, Structural Similarity Index ($SSIM$) for image quality preservation, and attack resilience against adversarial manipulations.

Our method achieved a $PSNR$ of 40.23 dB, ensuring minimal perceptual distortion, and an $SSIM$ of 0.98, confirming high image fidelity. Furthermore, the model demonstrated 95.4% accuracy, showing strong robustness against various attacks, including surrogate model attacks and content-preserving manipulations Table 1. Table 2 presents the performance metrics and Figure 2 illustrates the quality of watermarked images for the glaucoma and skin images. These results establish the proposed approach as a secure and reliable solution for preserving the integrity of medical images in telemedicine.

Table 1: Accuracy of copyright protection when different content-preserving image manipulation attacks are performed on the medical images.

Attacks	**Pneumonia**	**Glucoma**	**Mammography**	**Dermatoscopy**
Rotation (45°, 90°and 180°)	100	100	100	100
Vertical and Horizontal Flip	100	100	100	100
Gamma Correction (0.5)	100	100	100	100
Histogram Equlization	100	100	100	100
$JPEG$ **compression**	90.5	91	93	91.8
Gaussian Blur (7 × 7)	100	100	100	100
Mean Filtering (3 × 3)	100	100	100	100
Median Filtering (3 × 3)	100	100	100	100
Brightness (50)	100	100	100	100
Contrast (50)	100	100	100	100
Salt and Pepper noise (0.01)	97.5	98	99.2	98.5
Gaussian Noise (0.5)	100	100	100	100
Poisson noise (0.08)	100	100	100	100
Speckle noise (0.01)	98.3	99	99.5	99
Cropping	97.9	98	97.7	98.3

Table 2: **Performance Metrics**

Metric	**Value**
PSNR (dB)	40.23
SSIM	0.98
Attack Resilience	95.4%

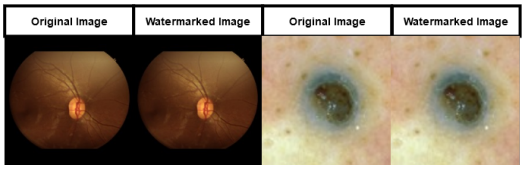

Figure 2: Perceptual Quality.

4. Conclusion

We proposed a deep learning-based dual watermarking technique which embeds a perceptual hash for copyright verification and a cryptographic hash for content and source authentication. By incorporating two independent watermarks, our approach ensures robust security against content-preserving manipulations while preserving diagnostic quality. Carefully designed loss functions and a structured training strategy further enhance its effectiveness. Experimental results demonstrate that our technique maintains high imperceptibility and exhibits strong attack robustness (95.4%). These findings highlight its reliability in securing medical images for telemedicine applications.

Acknowledgments

This research was conducted at the Machine Intelligence and Security of Things (MIST) Lab, IIT Bhilai. We gratefully acknowledge the support and funding provided by iHub Anubhuti-IIITD Foundation

References

Preetam Amrit, Naman Baranwal, Kedar Nath Singh, and Amit Kumar Singh. Convnet-hide: Deep learning-based dual watermarking for healthcare images. *IEEE MultiMedia*, 2024.

Ton Kalker, Jaap Haitsma, and Job C Oostveen. Issues with digital watermarking and perceptual hashing. In *Multimedia Systems and Applications IV*, volume 4518, pages 189–197. SPIE, 2001.

Matthew Tancik, Ben Mildenhall, and Ren Ng. Stegastamp: Invisible hyperlinks in physical photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2117–2126, 2020.

Stefano Woerner, Arthur Jaques, and Christian F Baumgartner. A comprehensive and easy-to-use multi-domain multi-task medical imaging meta-dataset (medimeta). *arXiv preprint arXiv:2404.16000*, 2024.