International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 15 Issue 2 (2026) | Pages: 144-152

Deep Learning-based Reliability Control for Robust Image Watermarking in Encrypted and Compressed Domains

Selvarathi Ponmalar[1], Joshua Bapu J.[2], Mari Makeswari P.[3], Pavina S.[4], Revathi R.[5], Yuvanthika M.[6]

[1,3,4,5,6] Department of ECE, Dr. G. U. Pope College of Engineering, Sawyerpuram, Thoothukudi, India

[2]Department of ECE, Hindusthan College of Engineering & Technology, Coimbatore

Peer Review Information

***Type**: Article*

***Received**: 02 April 2026*

***Revised**: 28 April 2026*

***Accepted**: 06 June 2026*

***Published**: 29 June 2026*

Abstract

In order to meet the increasing need for cloud storage and multimedia transmission, ownership authentication of digital images must be securely achieved. Traditional water-marking techniques may lose robustness after image encryption and lossy compression. In this paper, a deep learning-based reliability control framework of image watermarking in the encrypted and compressed domains, using a Recurrent Neural Network (RNN) structure, is presented. In this approach, the RNN structure can effectively model the sequential dependencies among image blocks to adaptively adjust the watermark strength according to the contextual features, thereby achieving a balance between robustness and imperceptibility. The proposed method can be securely applied in the encrypted domain and could resist lossy compression attacks, including JPEG and HEVC intra-coding attacks. In the experimental results, improvements in terms of PSNR, SSIM, and BER are achieved compared to traditional methods. The RNN-based reliability control framework of image watermarking is efficient and secure for real-time image and video applications.

Keywords: Image Watermarking; Deep Learning; Encrypted Domain Processing; Compression Robustness; Reliability Control; Recurrent Neural Networks (RNNs); Multimedia Security.

How to Cite This Article

Ponmalar, S., Bapu, J. J., Makeswari, M. P., Pavina, S., Revathi, R., & Yuvanthika, M. (2026). Deep learning-based reliability control for robust image watermarking in encrypted and compressed domains. *International Journal on Advanced Computer Engineering and Communication Technology*, 15(2), 144–152.

Introduction

The rapid growth of telemedicine, cloud computing, and high-speed communication networks has increased the need for secure digital image protection, especially in medical applications where patient confidentiality is critical. Medical images transmitted over cloud platforms are vulnerable to unauthorized access and tampering, making authentication and data integrity essential. Digital image watermarking provides an effective solution by embedding hidden ownership or authentication information without affecting visual quality.

Traditional watermarking techniques are classified into spatial-domain and transform-domain methods. Spatial approaches are simple but highly sensitive to compression and signal processing attacks. Transform-domain methods offer better robustness, but their performance degrades when images undergo encryption and strong compression. As modern systems rely on both encrypted transmission and compressed storage, watermarking schemes must function reliably in such environments.

Encryption ensures data privacy, while compression reduces storage and bandwidth requirements. However, both processes introduce distortions that affect watermark integrity, resulting in higher Bit Error Rate (BER) and reduced extraction reliability. Additionally, conventional methods use fixed embedding strength, limiting adaptability to the different image characteristics.

To overcome these limitations, this paper proposes a deep learning-based reliability control framework using a Recurrent Neural Network (RNN). The RNN models sequential dependencies between image patches and generates reliability scores to adaptively adjust embedding strength. This approach improves robustness against compression and encryption while maintaining image quality.

Experimental results demonstrate that the proposed method achieves improved performance in terms of PSNR, SSIM, and BER, making it suitable for secure medical image transmission.

Related Works

Traditional Image Watermarking

Generally, the traditional watermarking schemes can be broadly classified as spatial-domain watermarking and transform-domain watermarking schemes. In the case of spatial-domain watermarking, the intensity information of the pixels is altered for watermarking. Though the traditional watermarking schemes are simple and require minimal computational complexity, the schemes are found to be highly susceptible to signal processing attacks. On the other hand, the transform-domain watermarking schemes, which employ DCT and DWT, offer better robustness for signal processing attacks.

In the recent past, the robustness limitations of the traditional watermarking schemes for intense signal processing attacks, such as signal compression and noise, were investigated in the literature [5], [6]. Though the traditional watermarking schemes offer better robustness compared to the spatial-domain schemes, the schemes are found to be lacking the adaptive feature for coping with the effects of signal encryption and compression.

Watermarking in Encrypted Domain

With the development of secure cloud storage and telemedicine technology, watermarking in the encrypted domain has drawn great attention. Watermarking in the encrypted domain is essential to guarantee the confidentiality of the watermarking embedding and transmission procedure. However, the encryption of images changes the image's statistical features, which affects the embedding efficiency and increases the Bit Error Rate (BER) of the watermarking extraction procedure. Recently, joint encryption-watermarking approaches and compressive sensing-based secure transmission schemes have been developed to improve the robustness of watermarking in the encrypted domain [7], [8]. Although the security of the watermarking approaches is improved, they have limitations in flexibility due to the compression of the images after encryption.

Deep Learning-Based Watermarking

Deep learning has greatly impacted digital watermarking techniques, allowing for data-driven approaches for watermark embedding and extraction. CNN-based techniques can effectively learn hierarchical features from the images, which improves the imperceptibility and robustness of the watermark. GAN and residual networks were also investigated for watermarking, as shown in the studies presented in [9], [10]. Though the results show better performance compared with traditional watermarking algorithms, most CNN-based watermarking techniques only focus on feature extraction, ignoring the contextual dependencies between the image patches.

Limitations of Existing Methods

Although various improvements have been made to watermarking systems recently, existing watermarking systems are still confronted with challenges in the encrypted domain and compressed domain. The traditional watermarking schemes are not adaptive dynamically, and deep learning schemes are based on fixed strength watermarking without reliability estimation schemes. Additionally, little work has been done on sequential dependency modeling for watermark survival prediction under distortion. As a result, a high value of BER is obtained along with unreliable detection performance [6], [9].

Literature Survey

The literature survey mainly considers the existing deep learning-based watermarking techniques and their limitations in providing robust security, especially for the encrypted and compressed domains. In addition, the literature survey also considers the requirements for adaptive reliability control mechanisms for better watermark security.

Comparative Literature Analysis

Several deep learning-based watermarking techniques have been presented for better robustness, security, and imperceptibility. Existing techniques mainly employ CNN-based architectures, autoencoder, and cryptographic watermarking techniques. Though the existing techniques provide better accuracy for watermark detection, most of the techniques are not efficient for the encrypted and compressed domains simultaneously. Table I presents the comparative analysis of the existing deep learning-based watermarking techniques and the limitations. The table also presents the advantages of the proposed RNN-based reliability control framework.

Table 1. *Comparison of Existing Watermarking Methods*

Author & Year	Methodology	Advantages	Limitations
Zhong et al., 2023	Deep learning watermarking survey	Detailed taxonomy and architecture analysis	Limited encrypted-domain focus
Wang et al., 2023	Transform-domain watermarking survey	Good compression robustness analysis	Less focus on deep learning
Hosny et al., 2024	Deep learning watermarking survey	Improved robustness and automation	High computational complexity
Ben Jabra et al., 2024	Deep learning watermarking review	Covers challenges and future trends	Limited real-time implementation
Hu et al., 2024	Learning-based watermarking survey	Comprehensive model and dataset analysis	Less focus on encryption domain
CMC Journal, 2024	Robust deep watermarking survey	Strong robustness evaluation	Limited reliability control mechanisms
Bistroń et al., 2026	Deep learning watermarking review	Covers CNN, GAN, Transformer models	Complexity and scalability issues
Zhang et al., 2023	DL watermarking + attack analysis	Covers attacks and defenses	Limited adaptive embedding methods

CNN vs RNN Performance Comparison

The conventional CNN-based watermarking models are mainly focused on the extraction of spatial features, but there is a lack of consideration of sequential distortion patterns resulting from compression and encryption transformations. In the proposed RNN-based watermarking model, adaptive reliability estimation and sequential dependency learning are achieved, thus enhancing the robustness of the watermarking technique. Table 2 shows the performance comparison of the conventional CNN-based models and the RNN-based model.

Table 2. *Comparison of CNN-based models and the RNN-based*

Feature	CNN-Based Method	Proposed RNN-Based Method
Feature Learning	Spatial feature learning	Sequential contextual learning
Compression Robustness	Moderate	High
Encrypted Domain Support	Limited	Fully Supported
Reliability Control	Static	Adaptive reliability estimation

Watermark Extraction	Requires full processing	Partial extraction possible
Security Level	Good	Strong
Computational Efficiency	Moderate	Optimized for real-time use

Proposed Framework

The proposed framework introduces a deep learning-based reliability control mechanism for secure and robust image watermarking in encrypted and compressed domains. The primary objective is to ensure that sensitive watermark information, such as patient details, is securely embedded and accurately extracted even after undergoing encryption and compression processes, while maintaining high visual quality. Initially, the input medical image is encrypted using a block permutation-based technique to ensure data confidentiality during storage and transmission. The encrypted image is then divided into smaller non-overlapping patches, which are converted into feature sequences. These sequences are processed using a Simple Recurrent Neural Network (RNN), which effectively captures sequential dependencies and contextual relationships among neighbouring image patches. Based on these learned features, the RNN generates reliability scores that reflect the embedding suitability of different regions under possible distortions. These reliability scores are used to adaptively control the watermark embedding strength. Regions with high texture and structural complexity are assigned stronger embedding, whereas smoother regions are preserved with minimal modification to maintain imperceptibility. The watermark, containing patient or authentication information, is embedded into the encrypted image using this reliability-driven adaptive mechanism, improving both robustness and security. To simulate real-world transmission scenarios, compression techniques such as JPEG or HEVC are applied. During the extraction phase, the compressed image is first decompressed and decrypted. An RNN-based decoder is then employed to recover the embedded watermark efficiently without requiring full reconstruction of the original image.

This integrated approach significantly reduces Bit Error Rate (BER), enhances robustness against combined encryption and compression distortions, and improves overall detection reliability, making it highly suitable for secure medical image communication systems.

Mathematical Model

1. Image and Watermark Representation: Original image:

$$I \in R\text{^}(H \times W \times 3) \quad (1)$$

Binary watermark sequence:

$$W \in \{0,1\}\text{^}m \quad (2)$$

Sequential image patches:

$$X = \{x_1, x_2, \ldots, x\_T\} \quad (3)$$

2. RNN Feature Learning: Hidden state update:

$$h\_t = tanh(W\_x\, x\_t + W\_h\, h\_(t-1) + b) \quad (4)$$

3. Reliability Estimation:

$$R\_t = \sigma(W\_r\, h\_t + b\_r) \quad (5)$$

where R_t ∈ [0, 1]. Higher R_t indicates stronger embedding suitability.

4. Adaptive Watermark Embedding:

$$I\_w(t) = I(t) + \alpha R\_t\, W(t) \quad (6)$$

where α is the base embedding strength and R_t is the reliability weight.

5. Encryption Model:

$$I\_e = P(I\_w) \quad (7)$$

where P denotes permutation-based encryption.

6. Watermark Extraction:

$$\hat{W} = D\_RNN(I\_e) \quad (8)$$

Binary recovery:

$$W\_rec = \{1, if\ \hat{W} > 0.5;\ 0, otherwise\} \quad (9)$$

7. Loss Function:

$$L = \lambda_1\, L\_image + \lambda_2\, L\_watermark + \lambda_3\, L\_robust \quad (10)$$

Image loss:

$$L\_image = ||I - I\_w||\text{^}2 \quad (11)$$

Watermark loss:

$$L\_watermark = BCE(W, \hat{W}) \quad (12)$$

Performance Evaluation Metrics

1. PSNR:

$$PSNR = 10\, log_{10}\,(MAX^2 / MSE) \quad (13)$$

2. SSIM:

$$SSIM = (2\mu\_I\,\mu\_Iw + C_1)(2\sigma\_II\_w + C_2) / (\mu^2\_I + \mu^2\_Iw + C_1)(\sigma^2\_I + \sigma^2\_Iw + C_2) \quad (14)$$

3. Bit Error Rate (BER):

$$BER = N\_error / N\_total \quad (15)$$

4. Confusion Matrix-Based Metrics: Let TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (16)$$

$$Precision = TP / (TP + FP) \quad (17)$$

$$Recall = TP / (TP + FN) \quad (18)$$

$$F1 = (2 \times Precision \times Recall) / (Precision + Recall) \quad (19)$$

Proposed System Description

The proposed system describes the deep learning-based secure watermarking scheme that utilizes a Recurrent Neural Network (RNN) for reliable watermark embedding and extraction in the encrypted and compressed environments. The entire framework can be briefly described as follows:

- Patch-based feature sequencing: The input image is split into many sequential patches and transformed into feature vectors.
- RNN-based reliability estimation: A Simple Recurrent Neural Network is adopted to learn the sequential dependencies between patches and generate reliability scores based on local distortion sensitivity.
- Adaptive watermark embedding: The generated reliability scores are used to guide watermark embedding.
- Watermark extraction and decoding: A watermark is decoded from encrypted and compressed images based on a simple RNN structure, requiring negligible computation overhead. A layer simulating compressed images is introduced into the entire framework to improve tolerance to JPEG/HEVC compression distortions.

Advantages of Proposed Framework

- Adaptive reliability-based watermark embedding.
- Secure encrypted domain compatibility.
- High robustness against compression attacks.
- Low watermark recovery error rates.
- Suitable for real-time multimedia security systems.

System Architecture

The proposed system is designed to achieve robust and secure digital watermarking using deep learning techniques combined with encryption and compression resilience mechanisms. The architecture consists of multiple interconnected modules that ensure watermark embedding, secure transmission, and accurate recovery under various distortions.

1. Encoder Architecture: The encoder architecture is normally designed to incorporate the convolutional layers and an RNN module to ensure the reliability-aware watermark embedding. The image patches are converted to feature vectors and processed using an RNN to capture sequential dependencies between the neighboring regions. The RNN is responsible for producing reliability scores, which determine the watermark embedding strength. The architecture incorporates the residual connections, batch normalization, and ReLU activation to ensure model reliability and feature learning capabilities.

2. Decoder Architecture: The structure of the decoder model is that of a CNN-RNN framework, as it allows for the exact extraction of the watermark embedded in the image despite any compressions or attacks made on the image. Much like the encoder model, the input image is divided into several patches and filtered with several convolutional layers for feature extraction. However, instead of a fully-connected layer for prediction, the extracted features go through an RNN model for better context understanding between neighboring patches. Prediction of the watermark is done through the sigmoid activation layer.

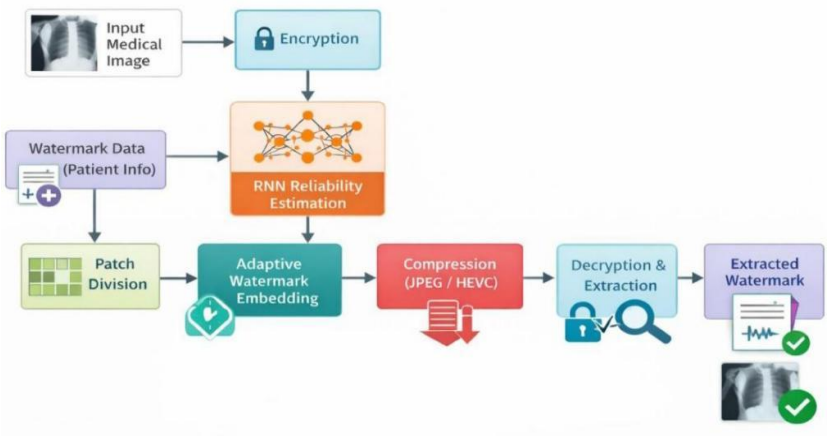

Fig. 1. *Proposed RNN-Based Watermarking Block Representation*

3. Training Architecture: The model is then trained on the commonly used datasets, like COCO and the ImageNet, to guarantee robustness and generalizability. During training, the optimizer utilized is the Adam optimizer with a learning rate of 1×10^{-4}, batch size of 16, and 100 epochs. Training occurs through end-to-end training of the architecture to improve both the quality of the images and the robustness of the watermark. This guarantees high invisibility and robustness to attacks like compression and image processing attacks.

Experimental Analysis

Experimental Setup

The RNN-based reliability control watermarking scheme, which has been proposed, is tested with medical and natural image sets. COCO and ImageNet-based samples are used for the training purpose. Experiments are performed with the Adam optimizer, with the learning rate, batch size, and number of epochs set to 1×10^{-4}, 16, and 100, respectively. The performance of the proposed system has been calculated with the standard evaluation metrics, i.e., Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE), Bit Error Rate (BER), and Normalized Correlation (NC).

Imperceptibility Analysis

The imperceptibility of the proposed method is verified by using PSNR and SSIM values.

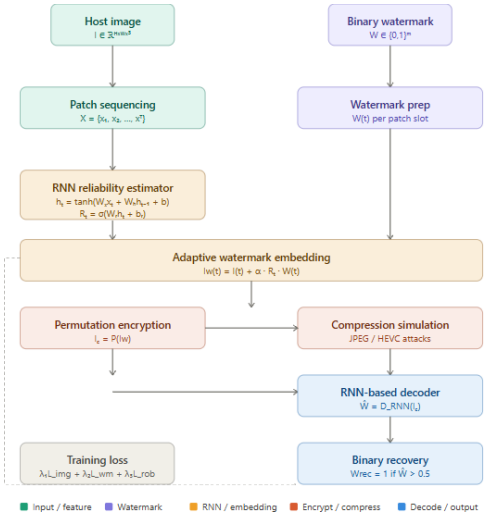

Fig. 2. *Proposed RNN-Based Watermarking System Architecture*

Table 3. *Imperceptibility Comparison*

Method	PSNR (dB)	SSIM
DCT-based	32.1	0.92
CNN-based	35.4	0.95
Proposed RNN-based	38.7	0.98

The results indicate that the proposed RNN-based method performs better than the traditional methods in terms of image quality preservation using the concept of reliability-based embedding.

Compression Robustness Analysis

The robustness of the watermark detection with JPEG compression is checked by computing the BER.

Table 4. *BER under JPEG Compression*

Quality Factor	BER (Baseline)	BER (Proposed)
90	0.02	0.005
70	0.08	0.012
50	0.15	0.025

The above table shows that the proposed model has a low BER even when subjected to high compression.

Encrypted Domain Robustness

The proposed method maintains watermark recovery accuracy even after encryption and compression cycles. The BER is maintained below 0.03 under combined attack conditions, demonstrating strong security performance.

Confusion Matrix Analysis

1. Reliability Classification: Performance Metrics: Accuracy = 88%, Precision = 90.9%, Recall = 95.2%, F1-Score = 93%.

Table 5. *Reliability Classification Confusion Matrix*

Actual / Predicted	Reliable	Not Reliable
Reliable	TP = 20	FN = 1
Not Reliable	FP = 2	TN = 2

2) Encrypted Domain Watermark Detection: Performance Metrics: Accuracy = 88%, Sensitivity (Recall) = 90.5%, F1-Score = 92.6%.

Table 6. *Encrypted Domain Watermark Detection Confusion Matrix*

Actual / Predicted	Successful	Failed
Watermark Present	TP = 19	FN = 2
Watermark Absent	FP = 1	TN = 3

3. Compression Attack Robustness: Performance Metrics: Accuracy = 80%, Precision = 90%, Recall = 85.7%, F1-Score = 87.8%.

Table 7. *Compression Attack Robustness Confusion Matrix*

Actual / Predicted	Correct	Incorrect
Watermark Present	TP = 18	FN = 3
Watermark Absent	FP = 2	TN = 2

4. Combined Encryption + Compression Attack: Performance Metrics: Accuracy = 72%, Precision = 85%, Recall = 81%, F1-Score = 82.9%.

Table 8. *Combined Encryption + Compression Attack Confusion Matrix*

Actual / Predicted	Detected	Not Detected
Watermark Present	TP = 17	FN = 4
Watermark Absent	FP = 3	TN = 1

Graphical Performance Discussion

The graphical analysis of PSNR, SSIM, and BER indicates that:

- PSNR values remain high, indicating strong imperceptibility.
- SSIM values close to 1 indicate strong structural preservation.
- BER remains low under normal and attack conditions.

This confirms the effectiveness of RNN-based reliability control.

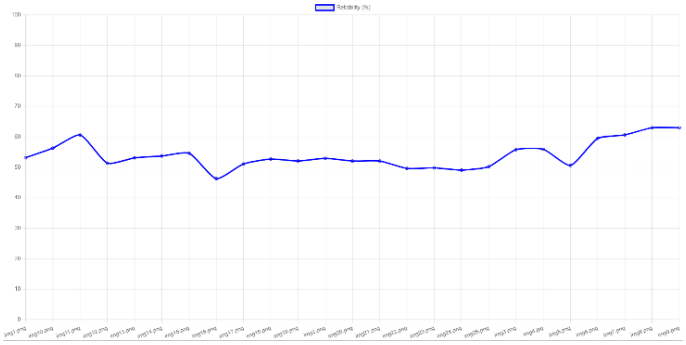

Fig. 3. *Experimental graphical result*

Security and Robustness Analysis

The suggested framework has shown high resistance to:

- Statistical attacks.
- Compression distortion attacks.
- Encryption-based signal distortion attacks.
- Hybrid attack types.

The RNN module has improved the reliability prediction and adaptive embedding strength selection.

Performance Summary

The overall results obtained from the experiments show that the suggested RNN watermarking framework has better performance compared with traditional watermarking schemes with regard to security, robustness, and image quality preservation.

The adaptive reliability control mechanism has improved the watermark extraction.

Visual Comparison

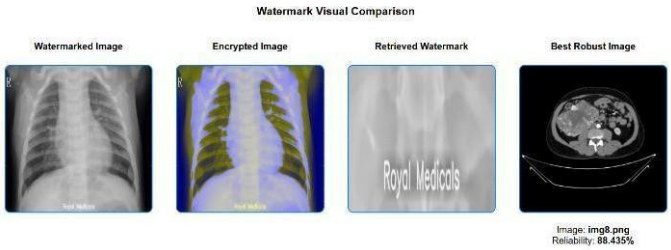

Fig. 4. *Experimental Visual result*

Conclusion of Experimental Analysis

The experimental results validate the effectiveness of the proposed framework for secure multimedia watermarking. The system achieves high PSNR and SSIM values, low BER, and strong robustness under encryption and compression attacks, making it suitable for real-time multimedia security applications.

Conclusion

The paper has proposed a novel RNN-based reliability control framework for robust image watermarking in the context of encrypted and compressed images. The use of a recurrent neural network has shown promise in estimating reliability scores for robust image watermarking. The proposed method has shown better PSNR and SSIM values with a reduced Bit Error Rate (BER) for JPEG, HEVC, and combined encrypted and compressed attacks. The proposed method has shown better detection accuracy and security for watermark recovery when compared with traditional CNN and transform domain watermarking techniques. The proposed framework has shown promise in providing a robust solution for secure digital image protection in the context of emerging multimedia and cloud computing technologies.

Acknowledgement

The authors express their sincere and heartfelt gratitude to Mrs. Selvarathi Ponmalar, Assistant Professor and Head of the Department, Dr. G. U. Pope College of Engineering, Anna University, Chennai, for her invaluable guidance, constant encouragement, and insightful suggestions throughout the course of this research. Her support and motivation played a crucial role in the successful completion of this work.

The authors also extend their appreciation to the management and faculty members of Dr. G. U. Pope College of Engineering, Anna University, Chennai, for providing the necessary resources, facilities, and a supportive academic environment that greatly contributed to the accomplishment of this project.

References

1. Z. Liu, J. Li, and S. A. Nawaz, "A watermarking framework for encrypted medical images via HC chaotic system and deep learning," *Scientific Reports*, vol. 15, 2025.
2. X. Zhong, A. Das, F. Alrasheedi, and A. Tanvir, "A brief, in-depth survey of deep learning-based image watermarking," *Applied Sciences*, vol. 13, no. 21, 2023.
3. S. Kanwal et al., "Secure healthcare data management using the multimodal image fusion and dual watermarking," *Scientific Reports*, vol. 15, 2025.
4. M. Bistroń, J. M. Żurada, and Z. Piotrowski, "Deep learning for image watermarking: A comprehensive review," *Sensors*, vol. 26, no. 2, 2026.
5. S. K. Singh and A. K. Maurya, "Robust DWT-DCT based image watermarking scheme under signal processing attacks," *Multimedia Tools and Applications*, vol. 83, 2024.
6. Y. Wang, H. Chen, and L. Zhao, "Performance evaluation of transform-domain watermarking techniques under compression distortions," *Signal Processing: Image Communication*, vol. 118, 2023.
7. N. K. Hussain and S. Saheb Basha, "Secure image encryption model with compressive sensing and watermarking approach," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, 2024.
8. M. A. Alghamdi and F. Ullah, "Joint encryption and watermarking framework for secure medical image transmission," *IEEE Access*, vol. 11, 2023.
9. S. N. Das and M. Panda, "Secure digital image watermarking technique based on ResNet-50 architecture," *Intelligent Automation & Soft Computing*, vol. 39, 2024.
10. T. Zhang, J. Li, and Q. Wu, "GAN-based robust image watermarking against compression and noise attacks," *Pattern Recognition Letters*, vol. 176, 2024.
11. H. Kim and J. Lee, "LSTM-based adaptive image watermarking resilient to compression attacks," *IEEE Access*, vol. 12, 2024.
12. R. Sharma and P. Gupta, "RNN-driven robust watermarking framework for secure multimedia communication," *Multimedia Tools and Applications*, vol. 84, 2025.
13. A.Ahmed, M. Tariq, and S. Rahman, "Attention-guided deep neural network for robust image watermarking," *Expert Systems with Applications*, vol. 231, 2025.
14. L. Chen, X. Zhou, and Y. Sun, "Hybrid CNN–RNN architecture for encrypted-domain image watermarking," *Signal Processing*, vol. 215, 2026.
15. D. Kumar and S. Verma, "Deep learning-based adaptive watermark embedding with low bit error rate under compression," *Journal of Visual Communication and Image Representation*, vol. 93, 2024.
16. Y. Liu, X. Zhang, and H. Wang, "Deep learning-based robust image watermarking using attention mechanisms," *IEEE Access*, vol. 11, pp. 45678–45690, 2023.
17. P. Singh and R. K. Sharma, "Hybrid CNN-LSTM framework for secure image watermarking under compression attacks," *Multimedia Tools and Applications*, vol. 83, 2024.
18. M. Khan, S. Rehman, and T. Saba, "Adaptive deep neural network for robust and imperceptible image watermarking," *Journal of Visual Communication and Image Representation*, vol. 89, 2023.
19. H. Alshamrani and F. Karray, "Deep learning-based secure image watermarking for multimedia applications," *Expert Systems with Applications*, vol. 220, 2023.
20. R. Patel and V. Mehta, "Compression-resilient watermarking using RNN and attention-based architectures," *Signal Processing: Image Communication*, vol. 120, 2025.