INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Secure Data Embedding in Medical Images Using an Optimized CNN–GWO Based Invisible Watermarking Approach

[1]Dr.D.Menaka, [2]Ms.L.Anju, [3]Ms.S.Kalyani

[1]Associate Professor, [2]Assistant Professor, [3] Assistant Professor
[1]Electronics and Communication Engineering,
[1]Sri Venkateswara College of Engineering , Chennai, India

Abstract : **Guaranteeing the security and integrity of medical images is imperative in the health sector. In this research, an imp erceptible watermarking scheme through a Convolutional Neural Network (CNN) tuned with the Grey Wolf Optimizer (GWO) for strong and hidden data embedding is introduced. Discrete Wavelet Transform (DWT) and Randomized Singular Value Decomposition (RSVD) are employed in filtering the image, while PSNR and BER based optimal locations for embedding are chosen by GWO. It is trained and tested on medical images with CNN-based models demonstrating high accuracy, minimal distortion, and high resistance to tampering. Applications are in secure telemedicine, forensic imaging, and data protection of medical information.**

IndexTerms **- Medical image security, CNN, GWO, watermarking, DWT,RSVD.**

INTRODUCTION

Medical imaging is a key aspect of clinical diagnosis, surgical planning, and medical research. With increasing dependence on digital platforms for storing and transferring sensitive patient data, maintaining the security, authenticity, and integrity of medical images has become an essential challenge. Compromise in data privacy or tampering with medical images can result in misdiagnosis, insurance fraud, legal disputes, and erosion of trust in healthcare systems. To safeguard such sensitive information, digital watermarking has proven a viable technique. It makes it possible to embed ownership or authentication data within the picture directly without lowering the quality of the image too much. The traditional watermarking methods, however, tend to be vulnerable to image processing attacks (such as compression, noise, and cropping), and could sometimes fail to offer the desired balance between imperceptibility and robustness in medical applications. This project presents a smart invisible watermarking method for medical images that is secure and diagnostically safe. The system's foundation is constructed with a Convolutional Neural Network (CNN) to learn smart features and patterns and the Grey Wolf Optimizer (GWO) to determine the best embedding points based on performance measures.

The CNN facilitates feature learning and adaptability, whereas GWO provides a nature-inspired optimization approach that emulates the grey wolves' leadership structure and hunting pattern to converge at optimal solutions. In addition, the system utilizes Discrete Wavelet Transform (DWT) to transform images into frequency sub bands with emphasis on the low-frequency (LL) band for strong watermark embedding. Randomized Singular Value Decomposition (RSVD) is used to increase security by selecting the most stable and globally informative singular values. To validate the method, various CNN-based architectures were utilized. The models were tested and trained on MRI datasets, with assurance that the watermarking will not lose any diagnostic quality of the images but will have high resistance to attacks. The uses of this work cover various key domains: • Telemedicine – Confidential transmission of patient information between healthcare centers. • Medical Research – Defense against misuse of research data. • AI Model Protection – Avoiding medical images from being used in unauthorized training datasets.

LITERATURE SURVEY.

The increasing trend of digital healthcare and telemedicine has required the construction of safe methods of protecting medical image information. New watermarking strategies have been put forward over the years, fusing conventional image processing methods with contemporary artificial intelligence techniques to provide robustness and invisibility. Vijay Krishna Pallaw (2023) proposed a strong medical image watermarking scheme based on nature-inspired optimization algorithms to provide better performance for telemedicine services. Although the method provided enhanced attack resistance, it was not adaptive and learning enabled, as required in real-time dynamic situations. This work extends this by adding deep learning with CNN to make the watermarking scheme intelligent and adaptive. Alireza Tavakoli and Zahra Honjani (2022) suggested a CNN-based watermarking scheme based on Discrete Wavelet Transform (DWT). The method had good imperceptibility and robustness. Yet, it employed fixed embedding positions and lacked optimization methods, hence being less adaptable to different types of images. This project, on the other hand, integrates CNN with the Grey Wolf Optimizer (GWO), enabling dynamic embedding position selection according to a fitness function that optimizes PSNR and BER. Ling-Yuanhsu and Hwai-TsuHsu (2023) used GWO with a DnCNN-based approach and QDCT for blind color image watermarking. While effective in general-purpose image watermarking, their approach was not specifically designed for medical images, which need to maintain diagnostic quality. This research fills this specific gap by targeting MRI images from the MRI dataset and assessing clinical usability. Most of the current methods either sacrifice image quality—making the images diagnostically unacceptable— or do not work in attacks like noise, compression, or cropping. Also, there is no integration between evolutionary optimization and machine learning models that has hindered the intelligence and flexibility of past systems. To bypass these constraints, your solution utilizes DWT + RSVD for robust feature extraction, GWO for optimal watermark placement, and CNN-based architectures to learn embedding patterns and generalize across challenging medical datasets. Such a hybrid approach guarantees the watermark to be imperceptible, robust, and secure, without affecting clinical interpretation.

PROPOSED METHODOLOGY

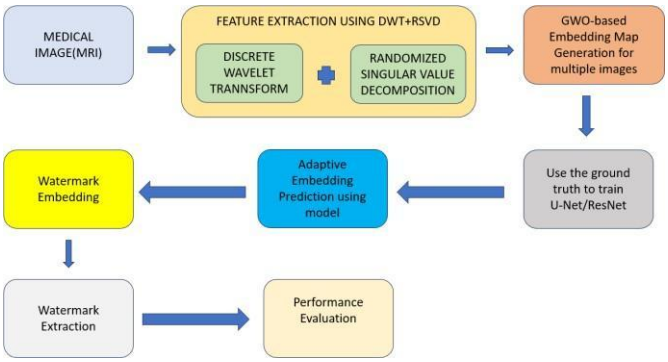

Fig 1:Workflow

The proposed watermarking system is a CNN-GWO- based adaptive embedding solution designed for secure medical image authentication is shown in Fig 1. It targets MRI data and predicts optimized embedding maps using a U-Net model trained on GWO-generated targets. The system employs DWT and RSVD for robust and imperceptible watermark embedding. Its modular, learning-based design makes it adaptable, interpretable, and effective for medical cybersecurity research and prototype deployments.

Feature Extraction using DWT + RSVD
Feature extraction is important in the determination of areas that are favorable for strong watermark embedding. In the proposed method, Discrete Wavelet Transform (DWT) and Randomized Singular Value Decomposition (RSVD) are employed to yield high-energy, low-frequency components that carry most of the image information.

Discrete Wavelet Transform (DWT)
DWT decomposes an image I(x,y) into four sub-bands: LL, LH, HL, and HH, where LL (low-low) contains the most important image features.
DWT(I)→ {LL, LH, HL, HH}

The LL sub-band is used for watermark embedding due to its resilience to compression and noise.

Randomized Singular Value Decomposition (RSVD)
RSVD reduces the dimensionality of the LL sub-band while preserving critical information.
Steps:

- Generate random Gaussian matrix Ω
- Compute sample matrix: $Y = A \times \Omega$
- Orthonormal basis: Q from QR decomposition
- Reduced matrix: $B = Q^t A$
- Perform SVD: $B = U\Sigma V^t$
- Approximation: $A \approx QU\Sigma V^t$ (1)

Where Σ contains dominant singular values representing the core structure of the LL band.

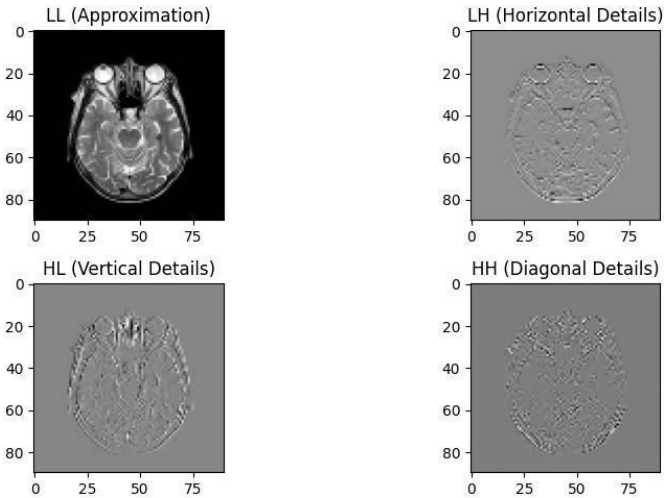

Fig 2:Single-level Discrete Wavelet Transform (DWT)of a brain MRI scan

The result of applying a single-level Discrete Wavelet Transform (DWT) to a brain MRI scan is diaplayed in Fig 2, decomposing the image into four subbands: LL (Approximation), LH (Horizontal Details), HL (Vertical Details), and HH (Diagonal Details). These subbands represent different frequency components, enabling the analysis of both structural and textural features within the image.

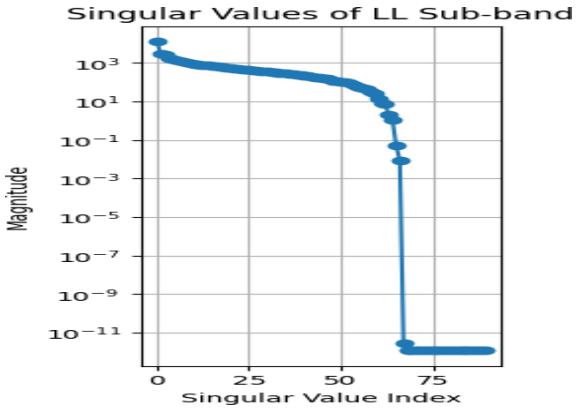

Fig 3:Singular values of LL band

The singular values of the LL sub-band from DWT is shown in Fig 3, with most image energy concentrated in the first few values, indicating the leading components hold key image information.

Grey Wolf Optimizer (GWO)

The Grey Wolf Optimizer is a metaheuristic algorithm inspired by the leadership hierarchy and hunting behavior of grey wolves (α, β, δ, ω). It is used in the proposed system togenerate adaptive embedding maps for optimal watermark placement.

Hunting behaviour modelling

Wolves update their positions based on the top three solutions (α, β, δ) to converge toward the best fitness.

Fitness function

The system uses a weighted fitness function to balance image quality and watermark robustness:

Fitness = $\alpha \times PSNR - \beta \times BER$ (2) Where:

PSNR (Peak Signal-to-Noise Ratio) measures imperceptibility

BER (Bit Error Rate) measures watermark robustness

α and β are weight coefficients tuned for optimal performance

This fitness guides the GWO to evolve embedding maps that improve watermark quality and security. Table 1 compares PSNR and BER values across four algorithms (GWO, PSO, GA, SA), showing GWO with the highest PSNR but highest BER

Table 1: Comparison of various algorithms

ALGO / METRICES	PSNR	BER
GWO	46.86	0.201
PSO	28.93	0.1124
GA	28.43	0.1108
SA	31.28	0.0721

CNN-Based Prediction of Embedding Maps

To automate adaptive watermarking, a Convolutional Neural Network (CNN) model is trained to predict optimal embedding maps based on the GWO-generated targets.

Model Architecture

A U-Net architecture is employed for its ability to capture both spatial context and fine details. The model takes MRI images as input and outputs pixel-level embedding maps.

Training Dataset

The training dataset consists of MRI images paired with embedding maps generated by the GWO algorithm. These maps highlight regions with optimal trade-offs between imperceptibility and robustness.

Training Process

The U-Net model is trained using supervised learning to minimize the difference between predicted and GWO-generated maps.

- Loss Function: Combined Mean Squared Error (MSE) and Structural Similarity Index (SSIM) loss to ensure spatial accuracy and perceptual quality.
- Optimization: Adam optimizer with learning rate scheduling.
- Normalization: Inputs and targets are normalized between 0 and 1.

This approach enables the model to learn GWO behaviour and generalize to unseen MRI images for fast, adaptive embedding during deployment.

Architecture of U-Net

U-Net is a CNN architecture designed for image segmentation but is also suitable for spatial tasks like watermark embedding. It has an encoder-decoder structure with skip connections that preserve spatial detail.

Key Characteristics:

Encoder: Downsampling through convolution and max- pooling.

Decoder: Upsampling with transposed convolutions.

Skip connections: Concatenate encoder and decoder features for precise localization.

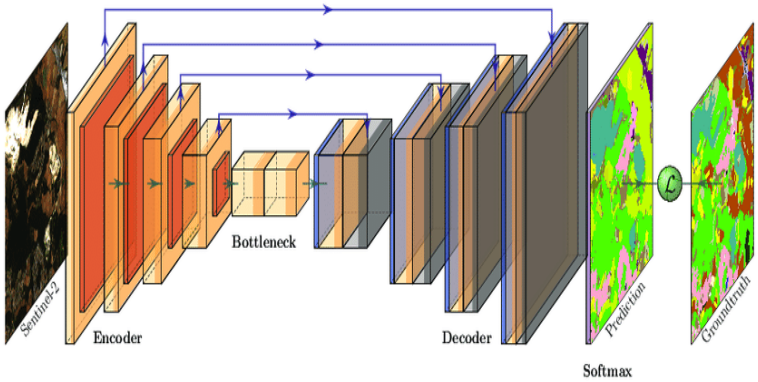

Fig 4:U-Net architecture

Figure 4 shows the U-Net architecture with an encoder- decoder structure and skip connections, which combine spatial and contextual features for precise image segmentation. Adapted from Ronneberger et al.'s original work, available at https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/.

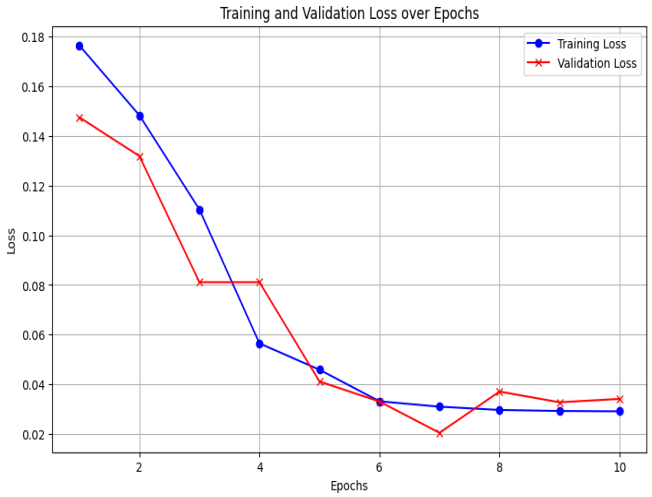

Fig 5: Training and validation loss of U-Net

Figure 5 shows the training and validation loss of U-Net steadily decreasing over 10 epochs, indicating good model convergence. Figure 6 shows the U-Net results with an original MRI and watermark, their watermarked MRI, and the extracted watermark.

Fig 6: U-Net results

Architecture of U-Net++

U-Net++ enhances U-Net by introducing nested skip pathways with dense connections and deep supervision.
Improvements to the U-Net:

- Better gradient flow and feature fusion
- Increased accuracy on small regions (important in watermarking).

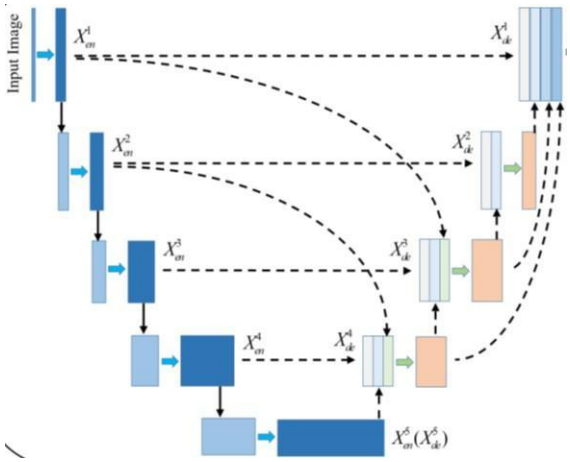

Fig 7: UNet++ architecture

Figure 7 depicts the UNet++ architecture, which preserves high-resolution features via parallel multi-scale branches and repeated information exchange.

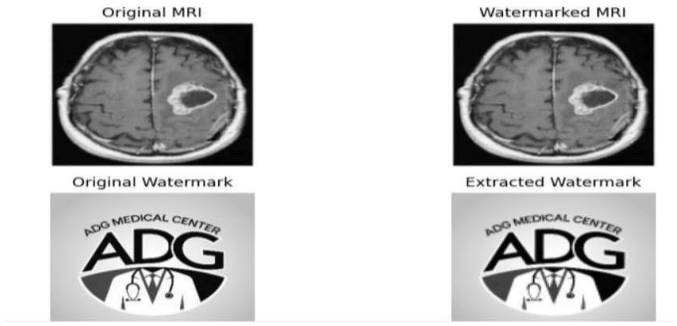

Fig 8:Watermarking process using U-Net++

Figure 8 shows a watermarking process using U-Net++ on MRI scans, where the watermark is embedded and later accurately extracted without visible loss of image quality

Architecture of ResNet

ResNet (Residual Network) introduces skip connections that allow gradients to flow directly through identity mappings, enabling deeper network training without degradation.

Key Benefits for Watermarking:

- Improved feature learning through residual blocks
- Better generalization and robustness on complex MRI textures
- Effective at capturing fine-grained embedding patterns

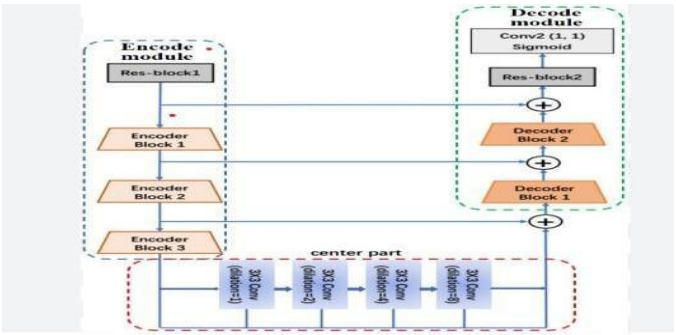

Fig 9:ResNet-based encoder-decoder architecture

Figure 9 illustrates a ResNet-based encoder-decoder architecture, employing residual blocks and skip connections toenhance feature extraction and reconstruction,The training and validation loss of a ResNet model shown in rapidly decreasing over 10 epochs, indicating fast and effective convergence.

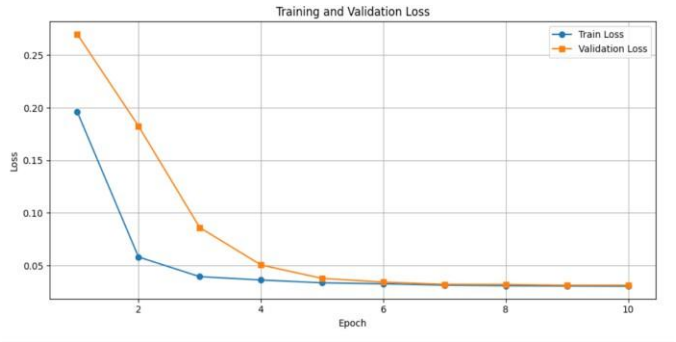

Fig 10: Training and validation loss of Resnet model

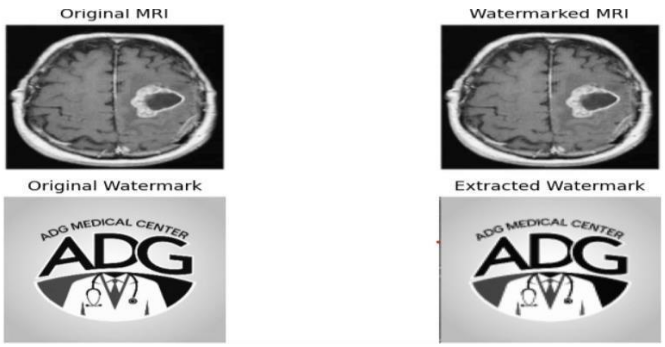

Fig 11: ResNet-based medical image watermarking system

The image demonstrates a ResNet-based medical image watermarking system, where a watermark is embedded into and accurately extracted from an MRI scan without visual degradation

Table 2: Comparison of different architectures

METRICES / ARCHITECTURE	PSNR	SSIM	BER	MSE	NCC	UIQI
UNET	45.6134	0.9942	0.2465	2.0651	0.9622	0.7933
RESNET	47.72	0.9980	0.0802	1.1735	0.9539	0.9443
UNET++	49.33	0.9973	0.0002	0.758	0.9999	0.9999

The table 2 shows that UNET++ outperforms UNET and RESNET across all metrics, achieving the best image quality and watermarking performanceU-Net++ outperforms the other architectures in nearly all metrics, indicating that it generates watermarked images with the best quality, most accurate watermark extraction, and least distortion.
ResNet follows closely behind, showing good performance in terms of image quality (PSNR, MSE) and structural similarity (SSIM).U-Net has the lowest performance, especially in BER, NCC, and MSE, suggesting that while it's still effective, it might not perform as well as U-Net++ or ResNet for watermarking tasks.

Limitations

The proposed method, although effective for MRI-based watermarking, is currently limited to a single modality. This restricts its generalizability to other medical images such as CT or X-ray scans. Additionally, the process of generating GWO- optimized embedding maps is computationally intensive and may pose a challenge for large-scale training. While the CNN model learns to approximate GWO behavior, its predictions occasionally show reduced contrast in fine-detail areas, which may impact precise embedding control in sensitive regions.

Future Work

Future enhancements could include extending the model to support multi-modal medical datasets, enabling broader applicability in healthcare. Advanced neural architectures such as transformer-based networks may be explored to improve spatial understanding and prediction accuracy. Integrating lightweight encryption mechanisms alongside watermarking could provide additional layers of security. Moreover, optimizing the GWO implementation through GPU-based parallelism would significantly reduce preprocessing time and make the system more practical for real-time scenarios.

Conclusion

In this study, a CNN-GWO-based adaptive watermarking framework is introduced for secure and imperceptible embedding in medical images. The approach leverages the strengths of Discrete Wavelet Transform and Randomized SVD for effective feature extraction, while a U-Net++ model is trained using GWO-generated maps to learn optimal embedding patterns. The system demonstrates strong robustness and high visual quality, validated by metrics such as PSNR, SSIM, and BER. The results confirm that the proposed method is both reliable and adaptable for medical cybersecurity applications.

References

[1] S. M. Shedole and S. V, "A robust dual watermarking using Grey Wolf Optimization, selective encryption and fast flexible de-noising convolution neural network," Journal of Machine and Computing, vol. 4, no. 3, pp. 820–829, 2024.
[2] H. K. Singh and A. K. Singh, "Digital image watermarking using deep learning," Multimedia Tools and Applications, vol. 82, no. 12, pp. 17289–17311, 2023.
[3]Z. Lin et al., "A CNN-based Robust Watermarking Scheme," IEEE Access, vol. 8, pp. 56323–56333, 2022
[4] A. Tavakoli, Z. Honjani, and H. Sajedi, "Convolutional neural network-based image watermarking using discrete wavelet transform," arXiv preprint arXiv:2210.06179, pp. 1– 10, 2022.
[5] Z. Liu, J. Li, Y. Ai, Y. Zheng, and J. Liu, "A robust encryption watermarking algorithm for medical images based on ridgelet-DCT and THM double chaos," Journal of Cloud Computing: Advances, Systems and Applications, vol. 11, no. 60, pp. 1–20, 2022.
[6]B. Madhu and G. Holi, "CNN approach for medical image authentication," Indian Journal of Science and Technology, vol. 14, no. 4, pp. 351–360, 2021
[7] A. Kaur and S. Singh, "Medical Image Watermarking: A Review," Multimedia Tools and Applications, vol. 78, no. 3, pp. 3051–3083, 2019.
[8] S. Maloo, M. Kumar, N. Lakshmi, and N. K. Pareek, "Robust digital image watermarking based on hybrid GWO- DWT technique," International Journal of Pure and Applied Mathematics, vol. 119, no. 12, pp. 12969–12976, 2018

[9] O. Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI, 2015.
[10]S. Mirjalili et al., "Grey Wolf Optimizer," Advances in Engineering Software, vol. 69, pp. 46–61, 2014.
[11]K. Kuppusamy and K. Thamodaran, "Optimized image watermarking scheme based on PSO," Procedia Engineering, vol. 38, pp. 493–503, 2012.
[12] X. Luo, F. Chen, and H. Li,"Deep learning-based robust image watermarking against geometric attacks,"
IEEE Transactions on Multimedia, vol. 24, pp. 3215–3227, 2022.
[13] M. Al-Haj and A. Mohammad,"Hybrid DWT-SVD based robust watermarking technique for medical images,"
Multimedia Tools and Applications, vol. 81, no. 5, pp. 6571–6593, 2022.
[14] Y. Qin, Z. Wang, and X. Zhang,"Deep neural network-based adaptive watermarking framework,"
Signal Processing: Image Communication, vol. 99, 2021, Art. no. 116437.
[15] S. Roy and A. K. Pal,"Optimisation-based digital watermarking using evolutionary algorithms: A survey,"
Expert Systems with Applications, vol. 186, 2021, Art. no. 115761.
[16] H. T. Huynh, M. T. Nguyen, and T. H. Le,"Robust and imperceptible medical image watermarking using CNN and chaotic encryption,"Computers in Biology and Medicine, vol. 140, 2022, Art. no. 105060.
[17] J. Zhang, C. Li, and S. Wang,"Transformer-based image watermarking network for robust medical image protection,"
IEEE Access, vol. 11, pp. 42115–42128, 2023.
[18] A. Rehman and M. Wang,"End-to-end optimized deep watermarking for secure image transmission,"Information Sciences, vol. 608, pp. 1204–1220, 2022.
[19] R. K. Singh and P. Kumar,"Medical image authentication using hybrid CNN and metaheuristic optimization,"
Biomedical Signal Processing and Control, vol. 84, 2023, Art. no. 104746.
[20] T. Chen and W. Zhao,"Robust blind watermarking using multi-scale wavelet transform and deep neural networks,"
Neurocomputing, vol. 517, pp. 72–85, 2023.