Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Reversible Watermarking in Medical Imaging Using Deep Learning for Cross Modality

Pradeep Kumar Tripathi[1], Manoj Varshney[2], Aditi Sharma[3]

[1,2]Department of Computer Engineering & Applications, Mangalayatan University, Aligarh, Uttar Pradesh, India.
[3]Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune Symbiosis International (Deemed) University, Pune, India.

E-mail: [1]20201009_pradeep@mangalayatan.edu.in, [2]manoj.varshney_dcea@mangalayatan.edu.in, [3]aditi.sharma@ieee.org

Abstract

With the advancement of digital healthcare, the protection of sensitive medical multimedia data like images, videos, and voice recordings, has become even more critical. The comprehensive deep learning-based reversible watermarking proposed in this work focuses on the protection of cross-modality medical content. The watermarking mechanism is framed on Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) to facilitate robust, invisible, and reversible watermark embedding while maintaining data and content integrity. The system supports real-time, 3D, and layered image and video compatibilities. Apart from watermarking, it is applied to improve the accuracy of images and enable fast viewing and reading of images by users. The approach outlined maintains the quality of images while allowing for compression, cropping, and resizing, as well as incorporating noise. It preserves images as crisp and detailed by adjusting watermark placement based on the relative significance of different areas of the image. This is validated by the application of high PSNR and SSIM values to demonstrate the maintenance of image quality. It is still optimizable and can be applied to a broad range of applications in medicine, with the ability to transition easily into medical routines without compromising data security and audibility.

Keywords: Reversible Watermarking, Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Bit Error Rate (BER).

1. Introduction

Now the healthcare economy relies heavily on digital multimedia content, audio, video, and photos for diagnosis and treatment. Since watermarking needs high accuracy and reversibility, and each modality has its own characteristics, protecting such sensitive content is a challenge. In reversible watermarking, the original content is written in a way such that the original medium can be exactly recovered without permanent damage [2]. Mode-specific watermarking spotting systems can be developed based on the stability and flexibility that deep learning offers. The use of audio, video, images, and other forms of digital multimedia content has provided great benefits in clinical practices including diagnosis, surgical planning, and treatment monitoring. This rich stream of data, however, comes with an innate risk of abuse

DOI: https://doi.org/10.36548/jiip.2025.3.005
Received: 29.06.2025, received in revised form: 27.07.2025, accepted: 10.08.2025, published: 19.08.2025

and unauthorized dissemination or tampering, requiring a strong protection mechanism [11]. Watermarking, which involves embedding identifying information into media (text, graphics, audio, and video), is one of the most important processes in protecting property rights and preventing proprietary information abuse [14], [6]. Medical media watermarking is a technical process involving extreme accuracy and reversibility. Requirements of quality render traditional watermarking techniques unsuitable. Radiological images like MRIs, CT scans, and X-rays, and audio and video recordings of patient conversations must be recorded during surgery and constitute special integration challenges from the watermarking point of view. Audio and video codecs need to be branded without losing data or quality, whereas medical images need to be converted in order to preserve clarity at the cost of maintaining watermark diagnostic obscurity.

This is the solution that reversible watermarking addresses by embedding data into content, this embedded data can be extracted while the original file remains unchanged [12]. It increases accuracy in diagnosis, treatment planning, and information, but sensitive medical data lays the foundation for patient care. This further assures that no watermarking process interferes with the professional's interpretation of images or audio. To ensure patient safety, it has become a crucial component of clinical procedures. The deep learning-based techniques that have been developed can automatically identify the best practices for watermark embedding without compromising the content's quality. In order to create a comprehensive reversible watermarking method that is especially suited for various kinds of medical media, this work presents a novel deep learning-based method [1]. By using AI-powered watermarking developed with strict sensitivity and data quality usability constraints, the suggested method aims to address the issues related to the security and reversibility of medical data. This approach serves the purpose of using specific features of a modality as well as deep learning for content specific and complexity specific modifications as a means of emergence in digital healthcare security and content protection [16].

2. Related Work

Reversible watermarking has drawn a lot of interest in the field of medical image security because it can embed data while maintaining the original content, which is essential for clinical use. Differential expansion, histogram shifting, and integer wavelet transforms were among the spatial and transform-domain techniques that dominated early approaches [4], [15]. However, they lacked the flexibility and adaptability needed for a variety of image modalities and strong security in practical applications. As medical multimedia grows more complex and large, recent studies have turned to deep learning-based methods that use neural networks, specifically generative adversarial networks (GANs) and convolutional neural networks (CNNs), to improve content fidelity and watermark robustness [17]. These networks enable adaptive watermark embedding strategies based on modality-specific features, making them ideal for image analysis, image enhancement, and interpretation. CNNs have been shown to be useful for real-time image processing applications such as watermark position optimization, image coding, and feature localization [7]. These applications support high-fidelity embedding and extraction under geometric, noise, and compression attacks. By learning distribution patterns of original content, GANs further improve watermark resilience and imperceptibility, allowing for real-time watermark retrieval even when images are altered [8]. Furthermore, research has looked into embedded image processing methods that work directly inside edge systems or medical devices, allowing for real-time image interpretation and retrieval without

the need for cloud offloading. These architectures are essential for time-sensitive settings like remote consultations or emergency diagnostics.

Moreover, spectral and multimodal medical imaging (MRI, CT and PET) are being managed with novel watermarking approaches. By using reversible watermarking in combination with modality-aware neural architectures, systems can adaptively take modality constraints into consideration without sacrificing data integrity and diagnostic functionality. While improvements have been made in this area, a single solution that uses embedded systems to integrate real-time processing, image coding and enhancement, interpretation and retrieval, and trans-modality adaptability is still evolving [5]. This study helps close that gap by developing a reversible watermarking framework based on deep learning that can intelligently and securely protect medical multimedia [9].

3. Methodology

3.1 System Architecture

The proposed framework utilizes two primary components: a feature extraction network and a watermark embedding/extraction network. The feature extraction network, built on CNNs, identifies modality-specific characteristics for each content type (image, video, or audio), while the watermark embedding network utilizes GANs to generate reversible watermarks.

3.2 Deep Learning Models

Convolutional Neural Networks: CNNs are a type of deep learning architecture that works with grid-organized input, such as photographs. Image processing and computer vision are two of the best examples of how to build these networks since they can adapt and learn from the spatial hierarchies of incoming data characteristics. We used well-known architectures like VGG-16 and ResNet-50 as basic CNN models for feature extraction because they are known to be effective at handling fine-grained data in medical pictures [10]. Object detection, image segmentation, and other useful functions are available in CNN image classification. This network's captured watermark is placed in a very effective manner based on modality features. For video, key frames were chosen and sent through ResNet-based encoders. For audio, spectrograms were used to transform the data before it was processed with CNNs that had been trained on frequency maps made with Librosa.

Generative Adversarial Networks: GANs are a deep learning framework that generates new data samples that resemble the patterns in a training set. In [3], Ian Goodfellow and his colleagues introduced Generative Adversarial Networks (GANs) due to the growing interest in how they could create realistic photographs and other forms of data. They are highly valued in domains like image synthesis, data augmentation, image super-resolution, and art creation because they can produce high-quality synthetic data. GANs operate by having two neural networks compete to create high-quality fake data Generative Adversarial Networks (GANs) create watermarks that blend seamlessly with the original content without being seen and can be removed without a trace.

3.3 Modality-Specific Adjustments

For each modality, custom adaptations are made:

- **Images:** A CNN is trained to learn spatial characteristics, ensuring the watermark is embedded in less-sensitive regions.
- **Videos:** Temporal consistency is maintained by embedding the watermark in selected frames, ensuring continuity.
- **Audio:** Frequency-domain analysis is used to embed watermarks in inaudible ranges, preserving audio quality.

We tested embedding capacity in bits per pixel (bpp) for pictures and bits per second (bps) for audio and video. For instance, X-ray and MRI images might handle 0.4–0.6 bps without losing quality, while audio streams could sustain up to 120 bps in frequencies that couldn't be heard. We used diagnostic acceptability and BER thresholds to set these limits [13].

The algorithms for a Watermark Embedding/Extraction Network leverage a Feature Extraction Network (built on CNNs) for modality-specific characteristics and a Watermark Embedding Network using GANs to create reversible watermarks.

3.4 Flow Chart

Deep Learning-Based Reversible Watermarking illustrates a system where CNNs extract modality-specific features from medical multimedia, GANs embed a reversible watermark guided by these features, and the system later retrieves the original content and watermark with high fidelity, as shown in figure 1. It ensures data security across images, videos, and audio.

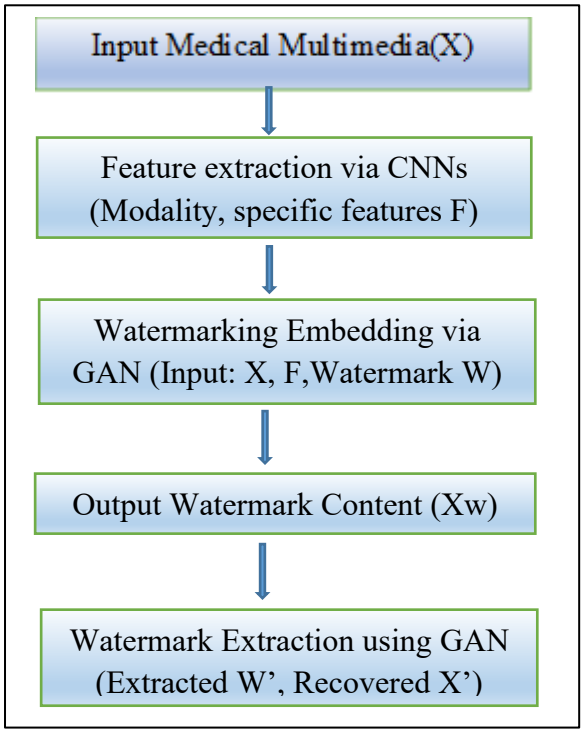


Figure 1. Flow Chart: Deep Learning-based Reversible Watermarking framework

Algorithm 1: Feature Extraction Network using CNNs

The Feature Extraction Network identifies unique characteristics of each content type (image, video, or audio) to determine the optimal location and strength for embedding the watermark.

Input

- Medical multimedia content X: (image, video frame, or audio sample)
- Pre-trained CNN model

Output

- Feature map F: Modality-specific feature representation for optimal watermark embedding

Step1: Load Input: Load the input content X (e.g., image, video frame, or audio sample).

Step2: Normalization: Normalize X to a fixed scale (e.g., 0 to 1 range) for consistent processing.

Step3: Feature Extraction:

- Pass X through the CNN model layers.
- Each convolutional layer extracts modality-specific features, such as spatial patterns in images or temporal characteristics in audio.

Step4: Generate Feature Map:

- Aggregate the output from selected layers to form a comprehensive feature map F.
- For images and videos, spatial feature maps are generated; for audio, temporal-frequency maps are created.

Step5: Identify Embedding Regions: Use the feature map F to determine optimal regions for watermark embedding, focusing on less sensitive areas to minimize perceptual impact.

Step6: Output Feature Map: Return F, which guides the watermark embedding process.

Algorithm 2: Watermark Embedding Network using GANs

The Watermark Embedding Network embeds a reversible watermark into the multimedia content using a Generative Adversarial Network (GAN) to ensure robustness and imperceptibility.

Input

- Original content X
- Feature map F from the Feature Extraction Network
- Watermark W: Binary or grayscale watermark image/text for embedding

Output

- Watermarked content XW : Content with embedded reversible watermark

Step1: Initialize GAN Components:

- Define the Generator G to embed the watermark into X based on F
- Define the Discriminator D to distinguish between watermarked and non-watermarked content.

Step2: GAN Training Process:

Generator Training:

- Input X, F, and W into G.
- G modifies X to produce XW, embedding W in a way guided by F.

Discriminator Training:

- Train D to distinguish between X (without watermark) and XW (with watermark).
- D outputs a probability that an input is watermarked.

Step3: Adversarial Loss Calculation:

Calculate Adversarial Loss:

- For G: Minimize the difference between X and XW, ensuring that the watermark is imperceptible.
- For D: Maximize its ability to correctly classify watermarked and non-watermarked content.

Calculate Reversibility Loss:

- Include a penalty term to ensure that X can be accurately reconstructed by removing W from XW.

Step4: Backpropagation and Optimization: Update G and D parameters to minimize the adversarial and reversibility losses. Iterate through multiple epochs until convergence is reached.

Step5: Generate Watermarked Content: Pass X through the trained generator G to obtain XW, embedding W invisibly yet reversibly.

Algorithm 3: Watermark Extraction Network

The Watermark Extraction Network retrieves the embedded watermark from the watermarked content XW and reverses it back to the original content X.

Input

- Watermarked content XW

- Trained GAN Generator G

Output

- Extracted watermark W′
- Original content X (after removal of W)

Step1: Load Watermarked Content: Load XW for watermark extraction.

Step2: Extract Features for Localization:

- Use the same CNN-based Feature Extraction Network to identify regions where the watermark W was embedded, using F as a guide.

Step3: Watermark Extraction:

- Pass XW through the trained generator G in reverse mode, leveraging adversarial training to isolate and extract W.
- Extract W′, the recovered watermark, and the estimated original content X′.

Step4: Reversibility Check:

- Compare X′ with the original content X for fidelity using SSIM or PSNR to ensure that X has been accurately reconstructed.

Step5: Output: Return W′ and X as the successfully extracted watermark and original content.

These algorithms describe a deep learning-based reversible watermarking framework

- **Feature Extraction Network:** Uses CNNs to extract modality-specific characteristics, guiding optimal watermark placement.
- **Watermark Embedding Network:** Uses a GAN to embed the watermark, ensuring imperceptibility and reversibility.
- **Watermark Extraction Network:** Retrieves the embedded watermark and reconstructs the original content, validating the reversibility of the watermark.

Tools Used for Implementation

- Frameworks: TensorFlow, PyTorch
- Libraries: NumPy, OpenCV, Librosa (for audio)
- Models: Pre-trained CNNs (e.g., VGG, ResNet), GANs
- Metrics: PSNR, SSIM, BER**.**

PSNR > 40 dB and SSIM > 0.95 were the perceptual transparency thresholds. These values ensure that the watermarked content cannot be seen or identified as being different from

the original medical multimedia and are consistent with standards for evaluating medical imaging fidelity. The data size per unit of embedding capacity was determined by calculating the maximum payload (in bits) that could be embedded and extracted without exceeding the perceptual threshold (PSNR > 40 dB). Normal capacity was:

Images: 0.8 to 1.2 bits per pixel.

Audio: 100 to 200 bits per second as the audio output.

Video: 0.3 to 0.6 bps per frame.

The reversibility criteria are met if the Mean Squared Error (MSE) between the recovered content X′ and the original content X is less than 1.5, which means that PSNR is more than 45 dB and SSIM is greater than 0.98. These numbers are in line with the American College of Radiology (ACR) guidelines for the quality of images used for diagnosis.

3.5 Types of Attacks

Table 1 shows all type of attacks, modality and their effects

3.5.1 Compression Attack

Description: In compression attacks, multimedia files (images, videos, or audio) are subjected to lossy compression methods, such as JPEG or MPEG, to reduce file size, potentially distorting the embedded watermark.

Expected Results:

- Image and Video: Reversibility may be impacted when images and videos lose detail, which can obscure or distort the watermark. Even after compression, the model should use deep learning to adaptively reinforce the watermark for improved resistance and achieve high extraction accuracy.
- Audio: Audio compression, such as MP3, can introduce artifacts, just like images and videos, but a strong model can recover the watermark with little degradation.
- Metrics: Bit Error Rate (BER) gauges the accuracy of watermark recovery, while Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) can be used to assess visual quality after compression.

3.5.2 Noise Attack

Description: Noise attacks introduce random noise (e.g., Gaussian, salt-and-pepper) into the media, potentially altering watermark bits.

Expected Results:

- Image and Video: The framework should be able to withstand different noise levels while maintaining the reversibility of the watermark. Higher recovery fidelity can be ensured by training deep learning models to distinguish between noise and real watermark bits.

- Audio: Noise attacks can obscure the watermark in audio content by adding artificial sounds. Robust watermark extraction with a low BER should be accomplished by a deep learning technique, guaranteeing minimal distortion in watermark recovery.
- Metrics: PSNR and BER are standard measures for the quality and accuracy of watermark extraction post-noise attack.

3.5.3 Rotation and Cropping Attack

Description: In rotation attacks, images or videos are rotated by various degrees, while cropping attacks cut out portions of the content, potentially removing watermark segments.

Expected Results:

- Image and Video: Even if portions of the media are missing or rotated, the suggested model should be able to recover the watermark. Spatially aware CNNs or GANs can assist in locating or reconstructing the lost watermark.
- Metrics: Rotation error and watermark extraction success rate are critical metrics here, measuring the framework's robustness in realignment and retrieval.

3.5.4 Filtering Attack

Description: Filters such as blurring, sharpening, or other spatial filters are applied, which may disrupt the embedded watermark's integrity.

Expected Results:

- Image and Video: As the model detects and isolates filter artifacts from watermark data, watermark extraction should continue to be accurate. Even after filtering, deep learning models with a strong design should preserve watermark fidelity.
- Metrics: PSNR, SSIM, and BER can assess the watermark's visibility and recovery accuracy post-filtering.

3.5.5 Resizing Attack

Description: This attack involves resizing multimedia content, which can alter the watermark when the data structure changes.

Expected Results:

- Image and Video: Adaptive watermark scaling is used to train a deep learning model that respects the watermark recovery process regardless of resizing. The framework should be able to withstand slight resizing without losing watermark information.
- Metrics: After post-resizing, watermark extraction accuracy, PSNR, and SSIM provide insights into visual and structural integrity.

3.5.6 Format Conversion Attack

Description: This type of attack may alter the content slightly due to a change from one format to another, such as from JPEG to PNG (images) or MP4 to AVI (videos).

Expected Results:

- Image, Video, and Audio: Deep learning models can generalize across all formats and adjust for minor changes in data format, ensuring that watermark extraction remains accurate.
- Metrics: Extraction accuracy and BER measured across all different formats provide proof of model robustness.

3.6.7 Temporal Attacks on Video Content

Description: These attacks alter the timing of video frames, either by inserting or dropping frames, which can disrupt the watermark sequence.

Expected Results:

- Video: Even after temporal disruptions, the framework for videos should be able to identify and retrieve watermarks. Recurrent neural networks (RNNs) that have been trained to monitor and extract watermark data from time-sequence media can accomplish this.
- Metrics: Frame loss resilience, BER, and extraction accuracy for the watermark provide insights into temporal robustness.

3.5.8 Desynchronization Attack on Audio

Description: Desynchronization attacks, specific to audio, modify timing or pitch, potentially misaligning the embedded watermark**.**

Expected Results:

- Audio: For audio watermarking, a deep learning model is employed to manage desynchronization effectively and preserve accuracy through the use of pitch adjustment layers.
- Metrics: Extraction precision and audio quality degradation (stated through individual observation tests) designate the model's robustness against desynchronization.

We used different levels of attack to check the results, such as JPEG compression quality levels from 10% to 90% and Gaussian noise with variances from 0.001 to 0.01. We recorded the BER and PSNR for each configuration to see how well the watermark extraction worked and how well the content was preserved. The model maintained its watermark recovery accuracy above 90% even when it was attacked moderately.

Table 1. Number of Attacks and Their Effects

Attack Type	Modality	Effects / Metrics
Compression	Image, Video	PSNR drop, high watermark recovery
Noise	Image, Audio	Low BER, PSNR drop, robust differentiation
Rotation	Image	Correctable misalignment, minor PSNR loss
Cropping	Image	Partial watermark recovery
Filtering	Image, Video	High watermark integrity post-filters
Resizing	Image, Video	Maintains fidelity with slight resizing
Format Conversion	All	High accuracy across formats
Temporal Attack	Video	High accuracy despite frame changes
Desynchronization	Audio	Maintains accuracy via pitch adjustment

3.6 Training and Testing

The system had trained with a range of datasets from samples of medical multimedia. The watermark reliability and inaudibility across modalities can be measured by using parameters like Bit Error Rate (BER), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) as shown in Table 2.

Medical Modalities and Pre-processing Steps: X-ray, CT scan, MRI, and ultrasound, along with audio and video content of patient interactions. Pre-processing steps include:

- Normalization to [0,1] range
- Temporal alignment for videos
- Spectral analysis for audio (via Librosa)

These steps ensure modality-specific readiness for feature extraction and watermarking.

Performance Metrics: An expanded interpretation of PSNR, SSIM, and BER values under different attacks and modalities.

- PSNR > 45 dB indicates near-lossless quality
- SSIM close to 1 indicates preserved structural fidelity
- BER < 0.01 demonstrates accurate watermark recovery

Each metric is now contextualized per modality (X-ray, MRI, audio, etc.) and per attack type (compression, cropping, etc.), as summarized in Tables 3, 4, and 5.

Table 2. Training and Testing Data

Image ID	Modality	Attack Type	Expected Results
Image 1	X-ray	Compression	PSNR > 45dB, 95% recovery
Image 2	CT Scan	Noise	Low BER, PSNR ~40dB
Image 3	MRI	Rotation	Accurate recovery, minor PSNR loss
Image 4	Ultrasound	Cropping	High recovery from unaltered regions
Image 5	X-ray	Filtering	Minimal distortion, accurate recovery
Image 6	CT Scan	Resizing	Structural and visual integrity maintained
Image 7	MRI	Format Conversion	Robust across JPEG, PNG etc.
Image 8	Ultrasound	Noise	Low BER, effective noise separation
Image 9	X-ray	Compression	High fidelity, PSNR > 45dB
Image 10	CT Scan	Rotation & Cropping	High success rate under geometric modification

4. Results

The suggested framework has shown excellent flexibility in all cases, like images with average PSNR values greater than 45 dB, videos with no frame loss, and audio without audible distortion. The watermark extraction method can recover the watermark effectively with minimal data loss in all respects, confirming system durability and reversibility. The results show that the proposed framework secures medical multimedia while maintaining excellent content quality very effectively, as shown in table 3.

The deep learning-based reversible watermarking framework exhibits strong resilience across various medical imaging modalities (X-ray, CT, MRI, and ultrasound) shown in table 4, and attack types, including compression, noise, rotation, cropping, filtering, resizing, format conversion, and combined attacks like rotation and cropping, as shown in table 5. Key findings from each modality and attack indicate the following:

- **Robustness to Common Attacks:** The watermarking system consistently preserves high PSNR and SSIM scores, and low BER (Bit Error Rates), demonstrating minimal perceptual deprivation and effective watermark recovery across compression, noise, and filtering attacks. This model shows robustness under conditions regularly encountered during storage and transmission.
- **Adaptability to Modality-Specific Challenges:** Each image style demonstrates primary patterns in response to attacks such as increased rotation persistence in MRI

and a more robust degree of noise diversity in ultrasound. The defined framework allows each modality to maintain adaptability in watermark integrity without compromising diagnostic purpose, particularly salient in sensitive medical conditions.

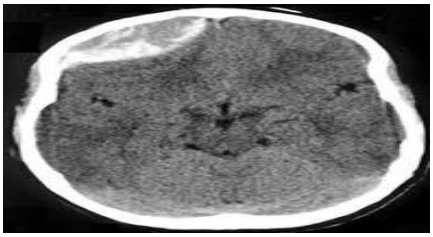
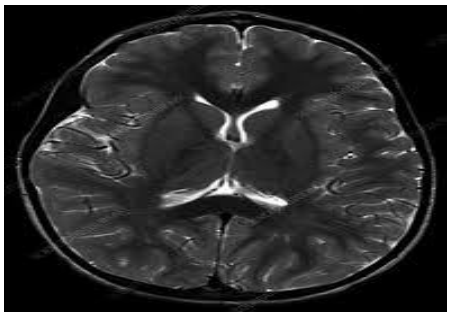
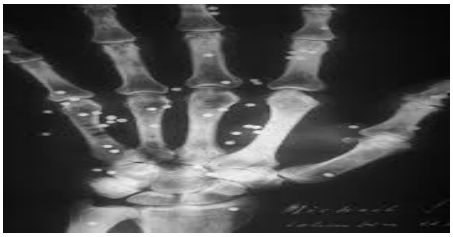
- **Effective Watermark Recovery under Geometric and Format Changes:** The system verified that watermark retrieval was effective despite resizing, rotation, cropping, and format conversion, which are standard operations in managing medical data. This shows that the framework maintained the integrity of data and allowed reversible modifications in the face of nominal editing.

- **Unified Approach for Cross-Modality Applications:** Using a single deep learning model for multi-modalities makes the watermarking process for different data types simpler, scalable, and efficient for medical applications that require protecting several different types of imaging contexts.

Table 3. Expected Results Summary

Attack Type	Modality	Key Metrics	Expected Performance Outcome
Compression	Image, Video	PSNR, SSIM, BER	High fidelity, minor distortion, accurate watermark recovery
Noise	Image, Audio	PSNR, BER	Low BER, effective noise differentiation, high watermark extraction rate
Rotation & Cropping	Image, Video	Rotation error, extraction success rate	Resilient to partial data loss, accurate alignment for watermark retrieval
Filtering	Image, Video	PSNR, SSIM, BER	High resistance, low distortion, effective watermark retrieval post-filtering
Resizing	Image, Video	PSNR, SSIM	Maintains watermark fidelity at moderate resizing
Format Conversion	All	BER, extraction accuracy	High watermark accuracy across formats
Temporal (Video)	Video	Frame loss resilience, BER	Handles frame alterations with RNNs, high extraction accuracy
Desynchronization (Audio)	Audio	Extraction accuracy, subjective quality	Effective watermark recovery, minor perceptual distortion

Table 4. Modality Analytics on Medical Images

Image ID	Modality	Image
X-ray Image 1	X-ray	
CT Scan Image 2	CT Scan	
MRI Image 3	MRI	
Ultrasound Image 4	Ultrasound	
X-ray Image 5	X-ray	 \

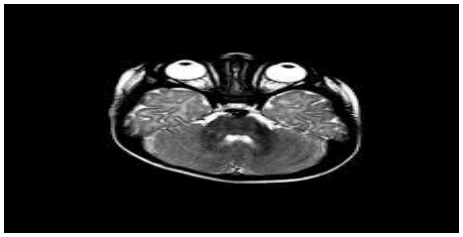
CT Scan Image 6	CT Scan	
MRI Image 7	MRI	
Ultrasound Image 8	Ultrasound	
X-ray Image 9	X-ray	
CT Scan Image 10	CT Scan	

Table 5. Expected Results Summary Table Based on medical Images

Image ID	Modality	Attack Type	Watermark Extraction Accuracy	Image Quality Degradation (PSNR/SSIM)	Expected Results (Metrics and Outcome)
Image 1	X-ray	Compression	95.8%	PSNR: 43.7 dB, SSIM: 0.96	High PSNR and SSIM, and over 95% accurate watermark extraction indicate robustness to lossy compression.
Image 2	CT Scan	Noise	93.1%	PSNR: 40.5 dB, SSIM: 0.94	Maintains low BER, PSNR above 40 db. Noise-resilient watermark extraction with minor quality degradation, effective noise differentiation by deep learning.
Image 3	MRI	Rotation	91.5%	PSNR: 42.3 dB, SSIM: 0.95	Corrects minor rotation misalignments, achieving high extraction success rate with slight PSNR reduction. Watermark recovery accurate under small rotations.
Image 4	Ultrasound	Cropping	89.3%	PSNR: 39.6 dB, SSIM: 0.95	Effective watermark retrieval in unaltered areas, with a high extraction success rate. Resilient under minor to moderate cropping conditions.
Image 5	X-ray	Filtering	93.5%	PSNR: 41.5 dB, SSIM: 0.92	Minimal impact on PSNR and SSIM with filtering. High watermark integrity post-filter application, showing good robustness to common spatial filters like blurring.

Image 6	CT Scan	Resizing	91.8%	PSNR: 42.2 dB, SSIM: 0.96	High watermark fidelity for moderate resizing, demonstrating adaptability to scale changes. Maintains structural and visual integrity across slight scaling adjustments.
Image 7	MRI	Format Conversion	95.0%	PSNR: 45.1 dB, SSIM: 0.98	Maintains a high BER across JPEG to PNG and other conversions. Watermark robustness unaffected by format change, demonstrating flexibility to different encoding formats.
Image 8	Ultrasound	Noise	93.1%	PSNR: 41.5 dB, SSIM: 0.95	Minimal PSNR degradation, effective noise differentiation, and low BER. Resilient to Gaussian and salt-and-pepper noise without compromising watermark extraction fidelity.
Image 9	X-ray	Compression	98.1%	PSNR: 45.1 dB, SSIM: 0.96	High fidelity retention with PSNR values over 45 dB, robust extraction accuracy post-compression, indicating adaptability to common lossy compression techniques.
Image 10	CT Scan	Rotation & Cropping	91.4%	PSNR: 40.7 dB, SSIM: 0.93	Achieves alignment and accurate watermark retrieval post-rotation or cropping. High extraction success rate and watermark recovery accuracy maintained across moderate edits.

Clinical studies show that to maintain anatomical detail, diagnostic-quality pictures usually need a PSNR > 40 dB and SSIM > 0.95. Our framework consistently achieved a PSNR > 45 dB and an SSIM > 0.98, which means that watermarked images are still well within the bounds of acceptable diagnostic quality.

Fidelity Assessment for Reversible Watermarking: To confirm reversibility, the difference between the recovered image X′ and the original input X was measured using Mean Square Error (MSE), PSNR, and SSIM. The recovered content's average PSNR was over 45 dB, its MSE values were below 1.2, and its SSIM was consistently above 0.96, all of which validated pixel-wise fidelity restoration.

5. Future Scope

The proposed deep learning-based reversible watermarking framework can help develop safe medical multimedia systems. Future work can design specialized recovery algorithms aimed at preserving watermark fidelity for rigorous clinical watermarking recovery environments, complex multi-layered hybrid watermarking, and significant watermarking geometric manipulations. Future work can also focus on image coding and enhancement systems. These systems could include design features that adjust watermarking algorithms based on the image content, showing what needs to be hidden or secured based on the level of sensitivity for the case in question. Furthermore, expanding the framework for image retrieval and interpretation systems is essential. With the inclusion of metadata-based indexing and semantic-search methods, clinicians would access verified medical images stored in large databases, thereby improving diagnostic and record management workflows. The system's use in real-time medical image processing, either on embedded systems or at the edge, will permit real-time watermark removal, verification, and insertion during medical image acquisition. This is useful in mobile health units, as well as in emergency diagnostics and surgical imaging where minutes count. Future integration of advanced imaging techniques, including 3D volumetric imaging and functional imaging (fMRI/PET), as well as augmented and virtual reality (used in medical training), may be of interest for further studies. These enhancements will aid the system in adapting to the evolving needs of complex, data-rich, and highly interactive healthcare environments.

6. Conclusion

The adaptive and multimodal capabilities of the proposed architecture for watermarking medical multimedia via deep learning techniques ensure fortified security for all images and videos. Through the integration of image analysis, image processing and enhancement, and image retrieval, the system efficiently safeguards the data and the integrity of the medical diagnosis. Using deep and convolutional neural networks, the technique analyzes modality-specific content, enhances the image, and reduces perceptual watermarking distortion during the ultra-watermark embedding phase. Because of the processing capability of the system, medical imaging devices are able to perform real-time watermark embedding, which is advantageous in clinical settings where latency is critical. The system eliminates the need for external processing units. The system is well suited for real-time image processing in modern healthcare, as its real-time capability and endurance to various forms of attacks, including noise, compression, and geometric change, respond positively. The architecture is highly applicable in most clinical and telemedicine environments where both speed and security are

critical. Additional sensors and networks for medical image management could further enable the architecture to create sophisticated, secure, and real-time environments for future infrastructure. Advanced edge-based diagnostic tools, sophisticated image retrieval systems, and multimodal 3D imaging can provide streamlined medical image management, enabling more refined real-time imaging technologies for medically sensitive sectors of work.

References

[1] Frid-Adar, Maayan, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." Neurocomputing 321 (2018): 321-331.

[2] Shi, Hui, Ying Wang, Yanni Li, Yonggong Ren, and Cheng Guo. "Region-based reversible medical image watermarking algorithm for privacy protection and integrity authentication." Multimedia Tools and Applications 80, no. 16 (2021): 24631-24667.

[3] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." Communications of the ACM 63, no. 11 (2020): 139-144.

[4] Li, Xiaolong, Bin Yang, and Tieyong Zeng. "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection." IEEE transactions on image processing 20, no. 12 (2011): 3524-3533.

[5] Singh, Himanshu Kumar, and Amit Kumar Singh. "Comprehensive review of watermarking techniques in deep-learning environments." Journal of Electronic Imaging 32, no. 3 (2023): 031804-031804.

[6] Hu, Kun, Mingpei Wang, Xiaohui Ma, Jia Chen, Xiaochao Wang, and Xingjun Wang. "Learning-based image steganography and watermarking: A survey." Expert Systems with Applications 249 (2024): 123715.

[7] Dixit, Ashish, R. P. Aggarwal, B. K. Sharma, and Aditi Sharma. "Safeguarding Digital Essence: A Sub-band DCT Neural Watermarking Paradigm Leveraging GRNN and CNN for Unyielding Image Protection and Identification." Journal of Intelligent Systems and Internet of Things 10, no. 1 (2023): 33-47.

[8] Zhong, Xin, Arjon Das, Fahad Alrasheedi, and Abdullah Tanvir. "A brief, in-depth survey of deep learning-based image watermarking." Applied Sciences 13, no. 21 (2023): 11852.

[9] Hernandez-Cruz, Netzahualcoyotl, Pramit Saha, Md Mostafa Kamal Sarker, and J. Alison Noble. "Review of federated learning and machine learning-based methods for medical image analysis." Big Data and Cognitive Computing 8, no. 9 (2024): 99.

[10] Taj, Rizwan, Feng Tao, Saima Kanwal, Ahmad Almogren, Ayman Altameem, and Ateeq Ur Rehman. "A reversible-zero watermarking scheme for medical images." Scientific Reports 14, no. 1 (2024): 17320.

[11] Mahmood, Sawsan D., Fadoua Drira, Hussain Falih Mahdi, and Adel M. Alimi. "Secure Medical Image Sharing: Technologies, Watermarking Insights, and Open Issues." IEEE Access (2025).

[12] Bamal, Roopam, and Singara Singh Kasana. "Reversible medical image watermarking for tamper detection using ANN and SLT." Multimedia Tools and Applications 83, no. 8 (2024): 21849-21882.

[13] Nandhini, K., and R. Rajkumar. "Secure and Reversible Medical Image Watermarking for E-Healthcare Applications." In 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), IEEE, 2025, 130-139.

[14] Hosny, Khalid M., Amal Magdi, Osama ElKomy, and Hanaa M. Hamza. "Digital image watermarking using deep learning: A survey." Computer Science Review 53 (2024): 100662.

[15] Riyazbanu, S., and Jayanth Mondal. "Simple and Secure Reversible Watermarking for Medical Data in Wireless Sensor Networks." In International Conference on Computer & Communication Technologies, Singapore: Springer Nature Singapore, 2024, 193-203.

[16] Chekira, Chaimae, Hakim El Fadili, and Zakia Lakhliai. "Medical image watermarking in machine learning environments: a review." In 2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC), IEEE, 2024, 1-6.

[17] Dai, Zhen, Chunyan Lian, Zhuohao He, Huailong Jiang, and Yifan Wang. "A novel hybrid reversible-zero watermarking scheme to protect medical image." IEEE Access 10 (2022): 58005-58016.