AGFI-GAN: An Attention-Guided and Feature-Integrated Watermarking Model based on GAN Framework for Secure and Auditable Medical Imaging Application

Xinyun Liu , Ronghua Xu * , Chen Zhao

Posted Date: 22 November 2024

doi: 10.20944/preprints202411.1711.v1

Keywords: medical images; digital watermarking; deep learning; security

 doi:10.20944/preprints202411.1711.v1

Article

AGFI-GAN: An Attention-Guided and Feature-Integrated Watermarking Model Based on GAN Framework for Secure and Auditable Medical Imaging Application

Xinyun Liu [1], Ronghua Xu [1,*] and Chen Zhao [2]

[1] Department of Applied Computing, Michigan Technological University, Houghton, MI 49931, USA
[2] Department of Computer Science, College of Computing and Software Engineering, Kennesaw State University, Marietta, Georgia 30060, USA
* Correspondence: ronghuax@mtu.edu

Abstract: With the rapid digitization of healthcare, the secure transmission of medical images has become a critical concern, especially given the increasing prevalence of cyber threats and data privacy breaches. Medical images are frequently transmitted via the Internet and cloud platforms, making them susceptible to unauthorized access, tampering, and theft. While traditional cryptographic techniques play a vital role, they are often insufficient to fully ensure the integrity and confidentiality of these sensitive images. In this paper, we present AGFI-GAN, a robust and secure framework for medical image watermarking that leverages attention-guided and feature integration mechanisms within a Generative Adversarial Network (GAN). Specifically, a Feature Integration Module (FIM) is proposed to effectively capture and combine both shallow and deep image features to facilitate multi-layer fusion with the watermark. The dense connections within the module facilitate feature reuse, boosting the system's robustness. To mitigate distortion from watermark embedding, an Attention Module (AM) is utilized, generating an attention mask by extracting global image features. This attention mask prioritizes features in less prominent and textured regions, allowing for stronger watermark embedding, while other features are downplayed to enhance the overall effectiveness of the watermarking process. The framework is evaluated based on its versatility, embedding capacity, robustness, and imperceptibility, and the results confirm its effectiveness. The study shows a marked improvement over the baseline, thus highlighting the framework's superiority.

Keywords: medical images; digital watermarking; deep learning; security

1. Introduction

Advancement in computer networking and multimedia technology have facilitated the rapid development of telemedicine, radiomics, and smart healthcare, marking a comprehensive transition of medical information exchange into the digital stage [1]. Thanks to rapid development of modern technology, medical imaging plays an increasingly important role in clinical diagnosis. Sharing digital medical images and electronic patient records (EPR) over the internet provides great convenience to doctors and patients, significantly enhancing the efficiency and accuracy of medical information transmission while promoting the optimal allocation of medical resources [2]. However, the widespread application of digital medical images also brings new challenges in information security and privacy protection [3].

The rapid advancement of modern technologies, such as the Internet of Things (IoT), cloud computing, and telemedicine, has surpassed traditional healthcare models. The growing transmission of digital medical images via the internet and cloud platforms has led to a significant rise in both active and passive cyberattacks on these data [1]. Common threats, including illegal access, tampering, deletion, and copying, pose serious risks to patient privacy and the integrity of medical images [4]. Despite the implementation of regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act [5], incidents of data theft and unauthorized access remain widespread. In addition, the digitization of

 doi:10.20944/preprints202411.1711.v1

medical images, while facilitating their storage and transmission, has also increased their vulnerability to cyberattacks [6]. Medical institutions are frequently targeted by cybercriminals, and the unauthorized access and distribution of patient data have become increasingly severe issues [7]. The legality and security of medical data directly impact the reliability of diagnoses, and the theft of patient information has emerged as a growing criminal activity, urgently necessitating stronger protection measures [8]. Thus, ensuring the security and privacy of medical images during transmission has become a critical challenge in the field of e-health services. Researchers have developed various technologies to safeguard the security and privacy of medical images, including steganography, cryptography, fingerprinting, and digital watermarking.

Digital watermarking, which embeds imperceptible information within digital content, has emerged as a powerful technique for ensuring medical data security in areas such as copyright protection, content identification, and forensics [9]. With key features like imperceptibility and robustness, digital watermarking (DW) is particularly effective for safeguarding e-health services. Unlike traditional cryptographic methods, digital watermarking enhances protection against unauthorized access and enables anti-counterfeiting traceability by embedding identifying information—such as text or images—into carriers like medical images or audio [10]. This technique exploits human sensory insensitivity and signal redundancy, allowing for the retrieval of hidden information when necessary.

In medical imaging fields, digital watermarking protects image content and verifies the authenticity and integrity of electronic patient records (EPRs), preventing unauthorized modifications and theft. As telemedicine grows, the need for secure remote transmission of medical images also increases. High-capacity, high-invisibility reversible watermarking algorithms ensure both patient privacy and undistorted recovery of medical images, providing a reliable solution for privacy protection. While modern technologies such as IoT, cloud computing, and telemedicine have increased the efficiency of medical services, they have also heightened security risks. Watermarking technology demonstrates balances in terms of embedding capacity, robustness, and visual quality. Therefore, it offers a promising mechanism to verify the authenticity and validity of medical data.

Traditional digital watermarking techniques applied to medical data are broadly categorized into spatial-domain and frequency-domain methods. Spatial-domain techniques involve directly modifying the pixel values of the host image, with the Least Significant Bit (LSB) method being a widely recognized example. In contrast, frequency-domain techniques embed the watermark within the image's frequency coefficients, which are obtained through transformations such as the Discrete Cosine Transform (DCT) [11], Discrete Wavelet Transform (DWT) [12], and Discrete Fourier Transform (DFT) [13]. Despite their effectiveness, these techniques typically require substantial prior knowledge and involve complex procedures, including both preprocessing and postprocessing steps.

Deep learning has made significant advancements in medical image processing tasks, such as image classification and segmentation, due to its powerful capacity for representing complex patterns[14,15]. Unlike traditional machine learning approaches, deep learning models can automatically extract intricate features from images, offering improved generalization across diverse scenarios by training on large-scale datasets. This capability has also been successfully applied to the field of medical image watermarking. Current deep learning models for medical image watermarking primarily utilize convolutional neural network (CNN) architectures. A notable CNN-based watermarking model integrates an iterative learning framework to improve watermark resilience, extending the frequency-domain techniques commonly used in traditional watermarking methods. Another example is ReDMark, a deep end-to-end diffusion watermarking framework that can learn watermarking models across any desired transform domain [16]. ReDMark consists of two fully convolutional networks with a residual structure, simulating various attacks through a differentiable network layer, thereby enabling end-to-end training.

Guo et al. [17] proposed a novel learning-based blind watermarking algorithm using a modified CNN, which optimizes feature extraction and embedding through self-adjustment and achieves improved robustness against common signal processing attacks. Geng et al. [18] investigate the

 doi:10.20944/preprints202411.1711.v1

integration of deep learning and blockchain technologies for copyright protection. Their work focuses on applying this fusion in photography and art studies by developing a digital image watermarking model that incorporates singular value decomposition (SVD) and deep learning techniques. The proposed model demonstrates strong scalability, robust performance in image processing, and effective decentralization, making it highly suitable for applications such as image copyright protection and post-processing. Fan et al. [19] proposed an image watermarking technique that combines CNN Inception V3 and DCT for image processing and feature extraction. This approach uses a logistic map and hash functions to scramble the watermarks. Three watermark forms were used in this method to test the algorithm's viability and generalizability.

Medical images are crucial for accurate diagnoses, but traditional watermarking algorithms compromise image quality and lack robustness against high-intensity attacks. Fan et al. [19] propose a novel watermarking algorithm using CNN Inception V3 and discrete cosine transform (DCT) to preserve image quality and improve robustness, achieving over 90% accuracy in experimental tests under various geometric attacks. Geometric attacks refer to manipulations or transformations of an image that alter its geometric properties, such as position, orientation, or shape, aiming to distort or remove the embedded watermark without significantly degrading the image quality. Zhang et al. [20] present a robust multi-watermarking algorithm for medical images, using GoogLeNet transfer learning to enhance privacy protection during transmission and storage, while improving resistance to geometric attacks and increasing watermark capacity. By fine-tuning a pre-trained GoogLeNet model and employing two-dimensional Henon chaos encryption, the algorithm achieves zero-watermark embedding and blind extraction, demonstrating their robustness and multi-watermark embedding in experimental tests. Nawaz et al. [21] propose a zero-watermarking method for medical images using discrete wavelet transform (DWT), ResNet101-DCT, and chaotic scrambling for encryption. By extracting deep features with ResNet101 and applying XOR operations, the algorithm demonstrates their robustness against conventional and geometric attacks, ensuring effective ownership verification. Nawaz et al. [22] introduce a zero-watermarking method for encrypted medical images using DWT, DCT, and an improved MobileNetV2 convolutional neural network. The technique enhances robustness by encrypting watermarks with a logistic map system and hash function, showing superior resilience and invisibility against conventional and geometric attacks in experimental results.

In addition to CNNs, GANs offer an alternative approach that can provide even greater effectiveness for deep medical image watermarking. A GAN consists of a generator and a discriminator that engage in an adversarial process [23]. This dynamic makes GANs particularly well-suited for robust image watermarking, striking a balance between capacity, invisibility, and resilience. In recent years, several GAN-based image watermarking models have emerged. Notably, a pioneering model called HiDDeN leverages the adversarial interaction between the generator and the discriminator to achieve robust watermarking [24]. HiDDeN adopts an encoder-noise-decoder architecture, where the encoder generates an imperceptible encoded image, and the decoder effectively retrieves the original watermark.

Based on the aforementioned research, we can confidently state that watermark algorithms operating directly on novel images inevitably degrade the image resolution, making watermarks vulnerable to geometric attacks. Furthermore, contemporary watermarking techniques, particularly those reliant on the extraction of image features for watermark embedding, often lack robustness and prove inadequate against image noise. A notable limitation of these models is their inability to prioritize critical image features during the learning process, which ultimately undermines the effectiveness of watermarking.

To address the challenges of achieving a balance between invisibility and robustness in digital watermarking, we introduce AGFI-GAN, an advanced model that incorporates attention mechanisms and feature integration within a Generative Adversarial Network (GAN) framework. The architecture of AGFI-GAN is structured as an encoder-noise-decoder model, allowing for end-to-end training. To mitigate the distortion commonly associated with watermarking, our model integrates an Attention-

 doi:10.20944/preprints202411.1711.v1

guided Module (AGM). This module analyzes global image features to produce an attention mask that dynamically adjusts the watermark's strength across various regions of the image. It minimizes the watermark's presence in flat or sensitive areas while amplifying it in inconspicuous or textured sections, thereby optimizing embedding. Additionally, a Feature Integration Module (FIM) is incorporated to enhance the extraction of image features by capturing both shallow and deep characteristics across multiple layers. By integrating the watermark with these deep features, the model improves its resilience to image noise, while the use of dense connections further strengthens its robustness. Furthermore, the adversarial dynamic between the encoder and the discriminator is leveraged to refine the quality of the encoded image. The discriminator plays a crucial role in distinguishing between the encoded and original images throughout the iterative training process, further supported by the AGM.

In summary, the main contributions of this paper are highlighted as follows:

- To improve watermarking invisibility and robustness, we introduce the Attention Module (AM), which computes a probability distribution among feature channels of the original image and applies spatial attention to learn spatial weights, generating an attention mask. This mask adaptively guides watermark strength across different image regions. Inconspicuous and textured areas receive higher embedding strength, while other regions are suppressed with lower strength.
- To address the limitation of extracting insufficient image features in existing deep learning-based watermarking models, we implement the Feature Integration Module (FIM) to learn and fuse shallow and deep features across multiple layers with the watermark. This multi-layer integration enhances robustness and improves resistance to various image attacks.
- We evaluate the performance of the watermarking model. Experimental results demonstrate that both the AM and FIM are crucial for the watermarking model's performance, surpassing existing models in watermarked image quality and resilience against diverse attacks.

The remainder of the paper is organized as follows: Section 2 provides background knowledge regarding DenseNet and feature integration module, and reviews attention schemes used in watermarking. Section 3 introduces system architecture of AGFI-GAN model and explains core components. Section 4 presents the prototype implementation and provides performance evaluation. Finally, Section 5 summarizes this work and briefly discuss future directions.

2. Background Knowledge and Related Work

This section describes the underlying principles of DenseNet, which has inspired the design of our feature integration module. Subsequently, we delve into a discussion of the attention mechanism, exploring its functionality and diverse applications across various domains, demonstrating its effectiveness in enhancing model performance through selective focus on relevant features.

2.1. Residual Network

Traditional neural networks face challenges as their depth increases, particularly the problem of network degradation, where accuracy improves to a certain point before rapidly decreasing. This issue arises despite the network's ability to converge. To address this, Kaiming He and colleagues from Microsoft Research proposed the Residual Network (ResNet) model in 2015 [25]. ResNets introduce residual learning, where the network learns the residual-the difference between input and output instead of a direct mapping between them. This is achieved through skip connections, which allow input to bypass intermediate layers and be directly added to the output. These skip connections offer two options: i) residual mapping, where features are processed through convolutional layers and activation functions, and ii) identity mapping, which directly adds the input to the output. Figure 1 illustrates the structure of the Residual Network. In this figure, X represents the output value from the previous layer that is fed into the neuron.

By utilizing these mechanisms, ResNets alleviate the vanishing gradient problem and enable the training of much deeper networks without degradation in performance. Once the network reaches an optimal depth, the residual mapping can be set to zero, leaving only the identity mapping, ensuring

 doi:10.20944/preprints202411.1711.v1

that network performance remains stable even as the depth increases further. ResNets, available in various configurations (e.g., ResNet-50, ResNet-101, ResNet-152), have become foundational in tasks such as image recognition and serve as the backbone for more complex models in object detection, segmentation, and other domains.

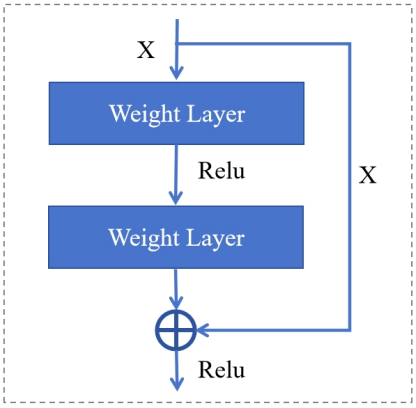

Figure 1. Residual Network.

2.2. Feature Integration Module Using DenseNet

Feature integration plays a crucial role in image processing, with significant advancements driven by the development of CNNs. DenseNet (Densely Connected Convolutional Networks), proposed by Huang in 2017, has been particularly influential in this area [26]. DenseNet features direct connections between any two layers with the same feature map size, which maximizes information flow and addresses the vanishing gradient problem. This architecture also encourages feature reuse and reduces the number of parameters compared to traditional CNNs. As Figures 2 and 3 shows, a DenseNet is composed of multiple Dense Blocks. Each dense block consists of a sequence of BN–ReLU–Conv1×1 and BN–ReLU–Conv3×3 layers, where BN–ReLU–Convj×j refers to a sequence made up of a Batch Normalization (BN) operation, a ReLU activation function, and a convolution with a j×j kernel size. The following formula determines the output of the ith dense block:

$$F_i = T_i(Concat(F_1, F_2, F_3, ..., F_{i-1})) \tag{1}$$

where T_i represents a series of nonlinear transformations of the dense block and $Concat(\cdot)$ is the concatenation operation.

By densely connecting layers, DenseNet integrates features from all previous layers, significantly enhancing the network's representational capacity. This feature reuse strategy has been effectively applied in tasks such as image super-resolution [27] and medical image analysis [28], where it improves performance by combining shallow and deep features. DenseNet's architecture, by focusing on feature reuse rather than simply deepening or widening the network, enhances resistance to image distortions and attacks, making it particularly valuable for applications like medical image watermarking.

 doi:10.20944/preprints202411.1711.v1

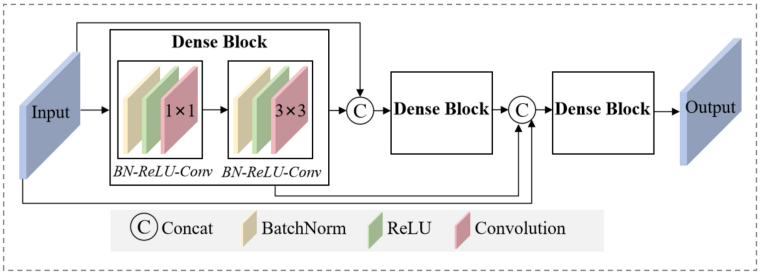

Figure 2. Architecture of DenseNet. The DenseNet extract both shallow and deep features, which are then fused with the watermark to enhance its robustness.

Layers	DenseNet
Convolution	7×7 conv, stride 2
Pooling	3×3 max pool, stride 2
Dense Block (1)	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 6$
Transition Layer (1)	1×1 conv
	2×2 average pool, stride 2
Dense Block (2)	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 12$
Transition Layer (2)	1×1 conv
	2×2 average pool, stride 2
Dense Block (3)	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 48$
Transition Layer (3)	1×1 conv
	2×2 average pool, stride 2
Dense Block (4)	$\begin{bmatrix} 1 \times 1\ conv \\ 3 \times 3\ conv \end{bmatrix} \times 32$
Transition Layer (4)	1×1 conv
	2×2 average pool, stride 2

Figure 3. Detailed architecture of DenseNet.

2.3. Attention-Guided Module

Incorporating attention mechanisms into deep learning model can enhance model performance by allowing the network to focus more precisely on critical information regions. Attention mechanisms enable the model to concentrate on meaningful areas or channels, effectively suppressing irrelevant or redundant information, helping it perform more effectively across complex tasks [15]. Attention mechanisms assign dynamic weights to various features, enabling the model to focus on essential information while downplaying less significant elements. Their use in image processing tasks has proven to be highly effective. For example, the Squeeze-and-Excitation (SE) network [29] emphasizes the value of channel-wise attention, enhancing network performance by adaptively adjusting responses across channels.

To enhance watermarking performance, attention mechanisms are integrated into the deep learning-based watermarking model. Depending on the specific task, the attention mechanism serves different purposes within the watermarking model. For improving watermark invisibility, Yu et al. [30] utilized ResNet to design an attention module that identifies inconspicuous areas for embedding the watermark. However, this model lacks robustness. Zhang et al. [31] created an attention mask based on image content to guide watermark embedding and extraction. Nevertheless, the robustness is insufficient because image distortions lead to significant discrepancies between the attention masks

 doi:10.20944/preprints202411.1711.v1

during embedding and extraction processes. To achieve robustness, Zhang et al. [32] focused on identifying stable locations by calculating an inverse gradient attention mask for watermark extraction. Despite this, the accuracy of watermark extraction significantly decreases under high-intensity noise conditions.

Although current watermarking models apply attention mechanisms to improve either invisibility or robustness, achieving both simultaneously remains a complex task [30]. Our method advances beyond prior approaches by utilizing both channel and spatial attention mechanisms, which extract essential information through max and average pooling, as depicted in Figure 4. First, the channel attention mechanism collects global statistics for each channel using two pooling methods: global max pooling and global average pooling, with the data then passed through a shared two-layer multi-layer perceptron (MLP). Spatial weights are then derived using a convolutional layer followed by a sigmoid activation, applying these weights across each spatial position on the feature map to strengthen spatial relevance. Moreover, global features generate an attention mask that emphasizes subtle, textured regions, leading to improved watermarking performance overall.

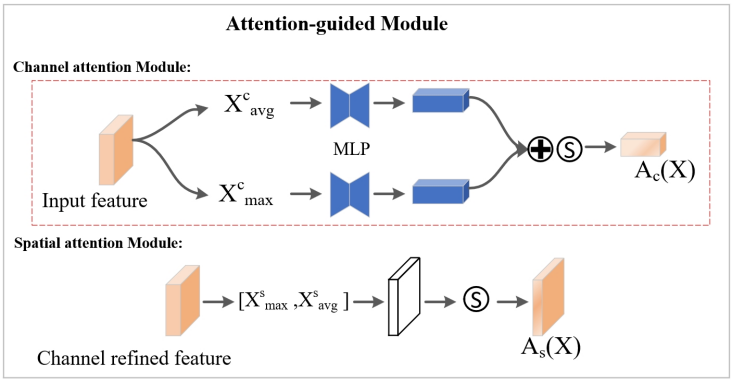

Figure 4. Architecture of Attention-guided Module, which employs both channel attention and spatial attention to enhance the performance of image feature extraction. It helps embed the watermark in less noticeable regions.

3. AGFI-GAN Digital Watermarking: Design Rationale and System Architecture

Medical images often contain complex and variable features, such as irregular shapes, occlusions, scale variations, and blurred boundaries—especially around lesions or abnormalities [33]. These factors make embedding watermarks more difficult, as they can affect both the visibility and robustness of the watermark. To address these challenges, adapting to local variations in texture, shape, and intensity is critical for a watermarking model. Therefore, we propose AGFI-GAN, a robust and secure framework for medical image watermarking that leverages attention-guided and feature integration mechanisms within a GAN. AGFI-GAN can help the model prioritizes features in less prominent and textured regions which allow for stronger watermark embedding. While other features are downplayed to enhance the overall effectiveness of the watermarking process.

3.1. System Overview

For system model, we assume that medical image generation and watermarking model rely on a secure execution environment. Thus, an adversary can neither modify original images and watermarks at the source side, nor compromise watermark verification at the end user side. We assume that image transmission and storage are not secure, therefore, an adversary can use replace tampered images with original ones. Figure 5 demonstrates the system architecture of AGFI-GAN watermarking model that guarantees the authenticity and provenance of medical images.

 doi:10.20944/preprints202411.1711.v1

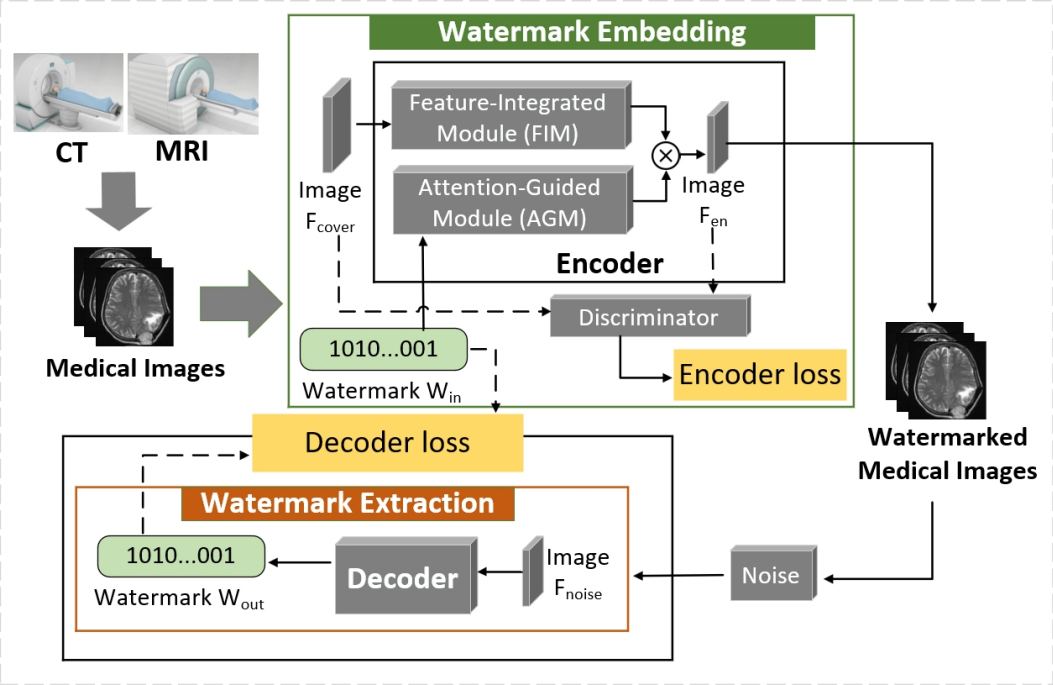

Figure 5. Overall architecture of the proposed AGFI-GAN. The AGFI-GAN is an end-to-end watermarking network designed to automatically generate watermarks with both invisibility and robustness. The key components of the model include the encoder, decoder, noise subnetwork, and discriminator.

After medical images are created by medical imaging devices (CT or MRI) at the source side, the watermark embedding process generates digital watermarks W_{in} as authenticity proofs and then hides them into medical images. Then, these watermarked medical images F_{en} can be transmitted among healthcare professionals for clinical diagnosis. At the end user side, the watermark extraction process can recover watermarks W_{out} from watermarked medical images and then verify whether medical images are forged or tampered.

To achieve the invisibility and robustness of watermarks that can be embedded in medical images, AGFI-GAN watermarking model adopts an encoder-noise-decoder framework. In watermark embedding process, the encoder extracts robust features, integrates them with the watermark, and distributes them across inconspicuous areas to maintain invisibility. The FIM module can extract robust features from original image F_{cover} that can be fused with watermark W_{in}. While AEM module can use attention mechanisms to minimize the distortion during watermark embedding process. Thus, the watermark can ideally be imperceptible to the human eye, which is accomplished by embedding it in regions of the image that are less noticeable to the human visual system, such as areas with high redundancy or low texture. Another objective of AGFI-GAN watermarking model is to ensure that the embedded watermark withstands various image manipulations, including compression, noise addition, rotation, and cropping. To achieve resilience against these attacks, the trained noise module adds the distribution of differentiable noises to F_{en} and outputs image F_{noise} that can resist various image attacks. Finally, the discriminator adopts adversarial training cycles based on encoder loss and decoder loss to improve quality of watermarked images.

Figure 5 illustrates the main components of AGFI-GAN, which include the encoder E_θ, decoder D_γ, noise layers $Noise$, and discriminator DC_ϵ. Here, θ, γ, and ϵ denote the trainable parameters for the encoder, decoder, and discriminator, respectively. These parameters are iteratively optimized during training to enhance the invisibility and robustness of the watermark. Given an input image F_{cover} with dimensions $H \times W \times C$ and a binary watermark W_{in} of length L, the encoder E_θ transforms F_{cover} into an encoded image F_{en}.

 doi:10.20944/preprints202411.1711.v1

The discriminator DC_e evaluates the likelihood that F_{en} resembles the original image, by analyzing the similarity between F_{cover} and F_{en}. This feedback guides the encoder E_θ in refining F_{en}. To improve the watermark's robustness, the noise layers *Noise* introduce various types of perturbations. Meanwhile, adversarial training is performed in which the decoder D_γ extracts the decoded watermark W_{out} from F_{noise}, aiming to make W_{out} as similar as possible to the original watermark W_{in}. The following sections will provide an in-depth explanation of each component.

3.2. Encoder

The encoder, denoted as E_θ, is specifically designed to embed a watermark into the original image F_{cover} with the goal of preserving the image's perceptual quality. This approach ensures that the embedded watermark remains unobtrusive while enhancing the overall resilience of the watermarked image against various attacks, such as noise addition, compression, and geometric transformations. To further improve the training process, a residual configuration is incorporated, which includes a global residual skip connection that effectively enhances backpropagation efficiency. This architectural choice not only accelerates convergence but also stabilizes training, helping to maintain the integrity of the embedded watermark. The mathematical expression for generating the encoded image F_{en} is as follows:

$$F_{en} = E(F_{cover}, W_{in}) \tag{2}$$

where $E(\cdot)$ represents the process of encoding, as illustrated in Figure 6. The encoder is composed of two essential modules: the Feature Integration Module (FIM) and the Attention-Guided Module (AGM).

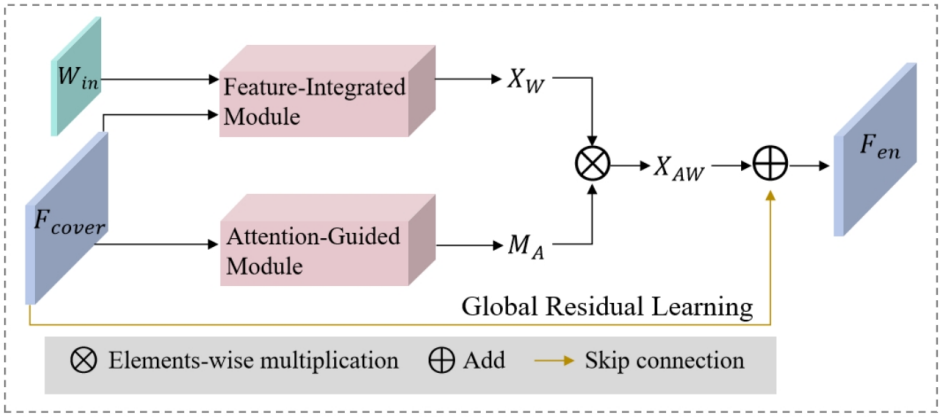

Figure 6. Architecture of encoder. The encoder includes (1) a Feature-Integrated Module (FIM) that utilizes dense connections to extract both shallow and deep features, which are then fused with the watermark to improve its robustness; (2) an Attention-Guided Module (AGM) that applies spatial attention to embed the watermark in less noticeable regions of the original image.

The FIM is designed to overcome the inherent limitation of relying solely on superficial image features, which may not be sufficient for effective watermark embedding. By capturing both shallow and deep features across multiple convolutional layers, the FIM ensures a comprehensive representation of the image's structural and textural details. This integrated approach allows for a richer feature set, significantly enhancing the encoder's capability to embed the watermark with high robustness and fidelity. The feature X_W is obtained in the following process:

$$X_W = E_{FIM}(F_{cover}, W_{in}) \tag{3}$$

where E_{FIM} represents the operation performed by the FIM. The FIM structure is built with multiple dense blocks, each comprising a series of Batch Normalization (BN), Rectified Linear Unit (ReLU) acti-

 doi:10.20944/preprints202411.1711.v1

vation functions, and convolutional layers. The dense connections within these blocks enable feature reuse, which considerably enhances the encoder's ability to represent complex patterns effectively.

Watermarks embedded in different regions of an image exert varying influences on distortion levels and robustness. Embedding in less prominent or low-texture regions can help maintain high visual quality by reducing noticeable distortions, whereas embedding in highly textured areas strengthens robustness, making the watermark more resilient against removal or degradation. To address these requirements, we developed a specialized module known as the Attention-Guided Module (AGM). This module is designed to mitigate distortion from the watermark embedding process by selectively enhancing or diminishing specific regions in the spatial domain of the feature map based on their relevance for watermarking. The AGM dynamically captures both local and global features of the image to generate an attention mask, which dictates differential embedding strengths across various image regions, thus achieving an optimal balance between watermark invisibility and resilience.

The attention mask serves a crucial role in fine-tuning the embedding process. By identifying and downscaling embedding strengths in smooth, flat, and visually sensitive areas, the AGM minimizes any perceptible impact on image quality. In contrast, regions that are inconspicuous or exhibit complex textures are assigned higher embedding strengths, thereby reinforcing watermark robustness without compromising visual stealth. Through this mechanism, the AGM ensures that the watermark remains imperceptible in visually sensitive areas while enhancing durability in parts of the image that can support stronger embedding. This carefully balanced approach enables robust watermarking that maintains the aesthetic integrity of the original image.

Specifically, the AGM utilize both channel and spatial attention mechanisms, which extract essential information through max and average pooling, as depicted in Figure 4. First, the channel attention mechanism collects global statistics for each channel using two pooling methods: global max pooling and global average pooling, with the data then passed through a shared two-layer multi-layer perceptron (MLP). Spatial weights are then derived using a convolutional layer followed by a sigmoid activation, applying these weights across each spatial position on the feature map to strengthen spatial relevance. Moreover, global features generate an attention mask that emphasizes subtle, textured regions, leading to improved watermarking performance overall.

Specifically, the AGM progressively generates a channel attention map, denoted as M_c, followed by a spatial attention map, M_s, as illustrated in Figure 4. The channel attention calculation is given by:

$$M_c(F_{cover}) = \sigma(MLP(F^c_{avg}) + MLP(F^c_{max})) \tag{4}$$

where σ denotes the sigmoid function, and MLP represents a multi-layer perceptron containing a single hidden layer. In a similar manner, the spatial attention is calculated as follows:

$$M_s(F_{cover}) = \sigma(f^{7\times7\times7}([F^s_{avg}, F^s_{max}])) \tag{5}$$

where $f^{7\times7\times7}$ denotes a convolution operation with the filter size of $7 \times 7 \times 7$. Once the channel-wise and spatial-wise attention are computed, the resulting features are represented as:

$$F'_{cover} = M_c(F_{cover}) \otimes F_{cover} \tag{6}$$

$$F^{out} = M_s(F'_{cover}) \otimes F'_{cover} \tag{7}$$

where $\otimes$ indicating element-wise multiplication, and F^{out} is the refined output of AGM. Based on the calculated F^{out}, we also need to obtain features in less prominent and textured regions, allowing for stronger watermark embedding. Let M_A represent the attention mask:

$$M_A = E_{AGM}(F_{out}) \tag{8}$$

 doi:10.20944/preprints202411.1711.v1

where $E_{AGM}()$ represents the process of generating attention mask. Next, M_A is employed to modify the distribution of X_W. Lastly, a global residual skip connection is applied to produce F_{en} by:

$$F_{en} = F_{cover} + X_W \times M_A \tag{9}$$

The encoding loss function, L_E, aims to reduce the difference between F_{cover} and F_{en} by adjusting the parameter θ. L_E consists of both the image reconstruction loss and the visual loss

$$L_E = \eta_1 MSE(F_{cover}, F_{en}) + \eta_2 SSIM(F_{cover}, F_{en}) \tag{10}$$

where $MSE()$ represents the mean-square error function, and $SSIM()$ refers to the structural similarity index metric. The weights for the image reconstruction loss and visual loss are represented by the parameters η_1 and η_2, respectively.

3.3. Watermark Extraction Decoder

The noise module, represented as $Noise$, is essential in simulating various attacks during iterative training by incorporating differentiable noise to improve watermarking robustness. This process enables the network to adaptively embed watermarks in image regions that are less prone to distortion, fostering a resilient watermarking pattern capable of withstanding both known and unexpected image noise. The resulting noised image, F_{noise}, is generated by systematically applying multiple types of noise

$$F_{noise} = Noise(F_{en}, N_{train}) \tag{11}$$

where N_{train} is a trained noise. Afterward, the decoder D_γ is utilized to extract W_{out} from F_{noise}.

$$W_{out} = D_\gamma(F_{noise}) \tag{12}$$

The process of training the decoder improves watermark robustness by minimizing the difference between W_{out} and W_{in} via adjustments to the parameter γ. The decoding loss L_D is formulated as:

$$L_D = \frac{\sqrt{(W_{in} - W_{out})^2}}{L} \tag{13}$$

Watermark extraction is fundamentally the inverse process of watermark embedding, utilizing a structure similar to the FIM to recover the watermark from the encoded image. The decoder, which is responsible for extracting the watermark, leverages deep feature extraction techniques, incorporating dense connections to ensure efficient propagation and utilization of important features throughout the network. The use of dense connections in the decoder facilitates the flow of critical information, preventing the loss of key features during the decoding process. This allows the network to concentrate on the most relevant regions of the image, even when it has undergone transformations such as compression, added noise, or mild blurring. As a result, the decoder can accurately recover the watermark, even in the presence of various distortions and noise-related artifacts.

Error tolerance is a crucial aspect of modern watermarking techniques, ensuring that the watermark remains recoverable despite noise-induced distortion. However, this tolerance has its limits. Excessive distortion, especially when introduced by aggressive noise or image degradation, can suppress the decoding process, making it challenging or even impossible to extract the watermark. To address this issue, feedback from the decoder to the encoder plays a vital role. During the training process, the decoder provides valuable insights into how well the watermark is being recovered, which in turn helps the encoder adjust its feature embedding strategy. This feedback mechanism allows the encoder to fine-tune its representation, ensuring that the features associated with the watermark are placed in regions of the image that are less likely to be affected by distortion. By doing so, the model

 doi:10.20944/preprints202411.1711.v1

can avoid fragile regions that might be vulnerable to excessive noise or transformation, thus improving the robustness of the watermark.

3.4. Discriminator

In a GAN based watermarking framework, the Discriminator plays a pivotal role in ensuring that the watermark embedded within the image is both imperceptible and robust. The primary function of the Discriminator DC_ϵ is to distinguish between real (unwatermarked) image F_{cover} and fake (watermarked) image F_{en}. During training, the Encoder(or Generator) produces watermarked images by embedding the watermark into the input image, while the Discriminator is tasked with classifying the image as either real or fake. The Discriminator's output is used to compute a loss function that guides the learning process for both the Discriminator and the Encoder. The Discriminator loss measures its ability to correctly classify real and fake images, while the Encoder loss incentivizes the Encoder to create watermarked images that are increasingly difficult for the Discriminator to identify as fake. This adversarial training process drives the Encoder to embed the watermark in a manner that minimizes perceptibility while ensuring it remains detectable under various transformations.

To enhance the invisibility of the encoded image, the discriminator aims to reduce the likelihood of incorrect classification by adjusting the parameter ϵ. The adversarial loss, L_A, is defined as:

$$L_A = \mathbb{E}_{p\sim F_{cover}}[\log(1 - P(F_{en}))] \tag{14}$$

where $P(\cdot)$ represents the probability that F_{en} contains W_{in}.

Through adversarial training process, the Discriminator forces the Encoder to refine the watermark embedding, making it robust against perceptual attacks, such as visual removal or distortion from compression and noise. The Discriminator thus helps the Encoder learn how to produce watermarked images that are both difficult to detect and resistant to various image alterations, while still allowing the watermark to be reliably extracted when needed.

4. Experimental Results and Discussions

This section begins with an overview of the experimental setup for training the watermarking model. Afterwards, we provide a detailed analysis of the AGFI-GAN watermarking model's performance. Finally, we examine the security measures of the proposed solutions, which are designed to protect against a range of potential attacks on medical data service platforms.

4.1. Prototype Implementation and Experiment Configuration

The proposed watermarking framework was evaluated using a diverse set of publicly available medical imaging datasets to cover a range of anatomical regions and imaging modalities. The datasets include Brain Tumor Detection Dataset [34], Liver Segmentation Dataset [35], Pneumonia Detection Dataset [36], Diagnosis of Coronary Artery Stenosis Dataset [37], Hand Image Dataset [38], Spine Image Dataset [39]. Each dataset was selected to provide a comprehensive evaluation of the proposed framework across various imaging modalities, supporting an extensive analysis of robustness, accuracy, and applicability in medical imaging contexts. The watermarking model is developed in PyTorch and runs on an NVIDIA Tesla V100 GPU. For consistency, all images are standardized to a resolution of $512 \times 512 \times 3$. To evaluate the watermark's invisibility, we measure the model's performance through both the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM).

To guarantee a fair and dependable performance assessment, five-fold cross-validation was applied to the above dataset, and the average results across all evaluation metrics were documented. During training, 15% of the training data was randomly designated as an internal validation set. In addition, both subjective and objective evaluations are conducted in this study to comprehensively assess the effectiveness of the watermarking model.

The subjective evaluation assesses whether the watermarked images maintain high visual quality as perceived by human observers, focusing on factors such as image clarity, overall quality, and the

 doi:10.20944/preprints202411.1711.v1

absence of visible distortions. By leveraging the synergy of the FIM and the AGM, we can obtain the encoded image F_{en} as described in Equation (2). To improve the invisibility of the watermark, we optimize the loss function in Equation (10) throughout the AGFI-GAN model training process. This approach reduces the difference between the original and watermarked images, ensuring minimal visual impact. Figure 7 illustrates the subjective invisibility achieved by AGFI-GAN by displaying both the original and watermarked images. As depicted in panels (a) and (b), the original and encoded images are nearly identical to the human eye, confirming that the watermark is embedded in visually imperceptible regions.

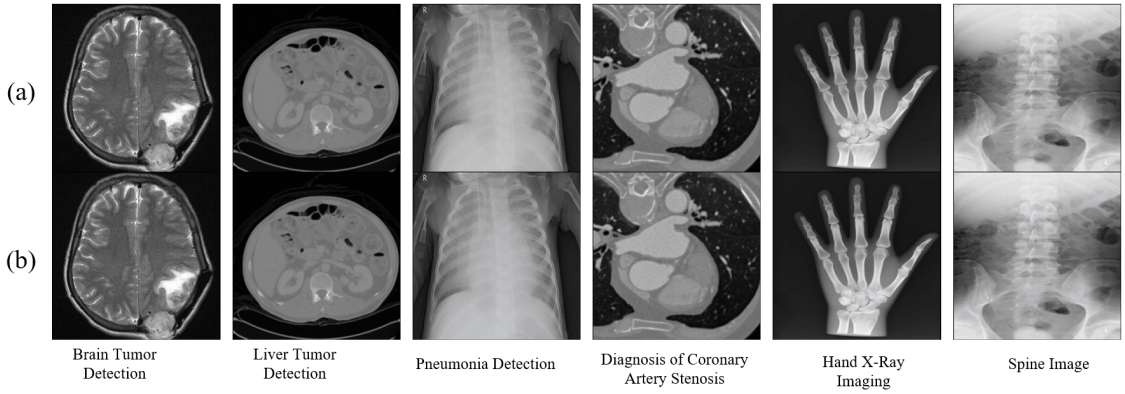

Figure 7. Watermarking performance of AGFI-GAN. (a) Original image. (b) Encoded image. The original and encoded images are visually indistinguishable, indicating that the watermark has been successfully embedded in imperceptible areas.

4.2. Subjective Evaluation of the proposed Watermarking Model

To evaluate the invisibility of the proposed approach regarding distribution of pixel value, Figure 8 provides a comparison of the histograms of the original and watermarked images. By analyzing pixel value distributions in both images, this comparison aims to measure any distortion introduced by the watermarking. The results reveal only minor pixel value deviations, suggesting that the watermarking process maintains high visual quality with minimal distortion.

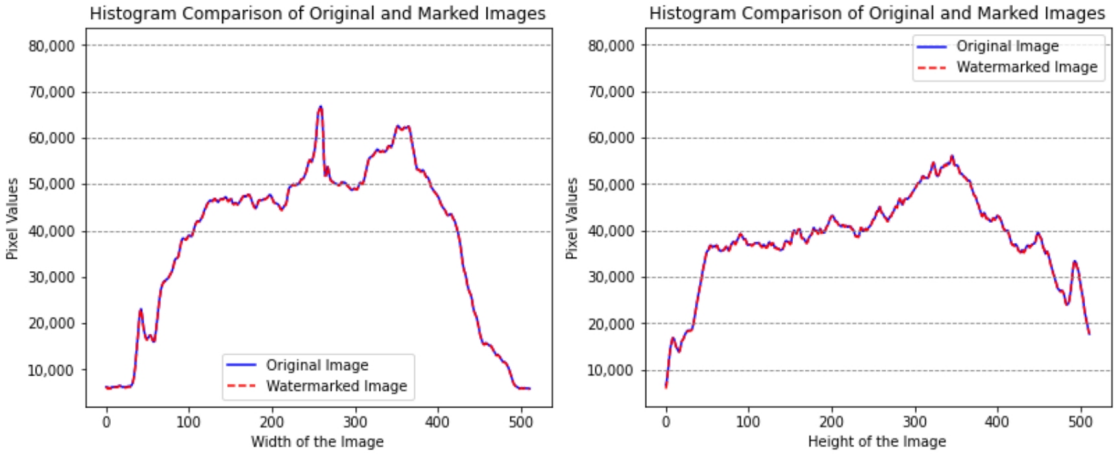

Figure 8. Subjective evaluation through histogram comparison of the original and watermarked images.

Furthermore, the Scale-Invariant Feature Transform (SIFT) technique is utilized to evaluate the effectiveness of the watermarking method by comparing the feature descriptors of both the original and watermarked images [40,41]. SIFT is a well-established computer vision algorithm that identifies and

 doi:10.20944/preprints202411.1711.v1

characterizes local features in an image, such as edges, corners, and blobs. Its strength lies in its ability to produce robust and repeatable descriptors that are invariant to variations in scale, rotation, and lighting. This makes SIFT particularly suited for evaluating how watermarking affects the underlying image content. As shown in Figure 9, the results reveal a substantial number of matching descriptors, with minimal distortion observed between the images.

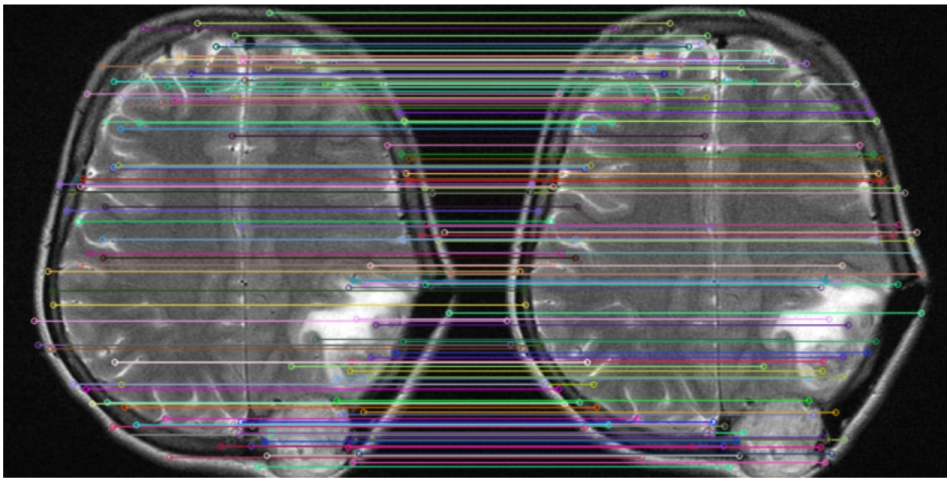

Figure 9. Subjective evaluation through SIFT feature matching between the original and watermarked images.

4.3. Objective Evaluation of the proposed Watermarking Model

Objective evaluation utilizes measurable metrics such as PSNR, SSIM, and watermark robustness analysis to rigorously assess the performance of watermarking models. These metrics provide a standardized framework for evaluating image quality and structural integrity, ensuring consistency and reproducibility in the assessment process. By employing these quantitative measures, the impact of watermarking on image fidelity can be precisely determined, enabling meaningful comparisons between different watermarking approaches.

Table 1 shows the PSNR and SSIM values for the watermarking models, which are derived by averaging results from 4,500 encoded images in the Brain Tumor Detection dataset. For comparison, we use six state-of-the-art (SOTA) watermarking models: HiDDeN [24], ReDMark [16], DA [43], TSDL [42], MBRS [44], and MIWET [45]. While we attempted to replicate the experiments from DA and TSDL, we were unable to fully reproduce their results, so we rely on the published outcomes from their respective studies for comparison. Additionally, MBRS offers a pre-trained model, and we have used their provided test results for evaluation. To ensure consistency in the comparison, the binary watermark length is fixed at $L = 30$ for all models considered.

Table 1. Comparison of different models on Brain Tumor Detection dataset.

Model	PSNR	SSIM
HiDDeN [24]	32.21	0.9287
TSDL [42]	33.50	-
DA [43]	33.70	-
MBRS [44]	35.74	0.8926
ReDMark [16]	37.86	0.9689
MIWET [45]	41.83	0.9845
AGFI-GAN	44.37	0.9902

As presented in Table 1, our AGFI-GAN model significantly outperforms the other six state-of-the-art watermarking models in terms of SSIM. Specifically, AGFI-GAN achieves a PSNR of 44.37, which is 12.16 dB higher than HiDDeN's PSNR of 32.21 and 10.67 dB higher than DA's PSNR of 33.70.

 doi:10.20944/preprints202411.1711.v1

When compared to MBRS, which has a PSNR of 35.74, AGFI-GAN shows an improvement of 8.63 dB. In terms of SSIM, AGFI-GAN scores 0.9902, surpassing HiDDeN's SSIM of 0.9287 by 0.0615. Moreover, compared to MBRS, which has an SSIM of 0.8926, AGFI-GAN's SSIM is 0.0976 higher, demonstrating that AGFI-GAN better preserves image details and visual quality. Additionally, AGFI-GAN also outperforms ReDMark in SSIM, indicating that our method more effectively retains the structural integrity of the original image after watermarking.

To assess the robustness of the proposed watermarking model, we calculate the bit accuracy (BA) by averaging the results across all the tested images. The BA is computed using W_{out} as outlined in Equation (12). To optimize performance, we adjust the loss function presented in Equation (14), where a lower loss value L_D is associated with higher BA, indicating improved robustness.

The bit accuracy (BA) results are shown in Figure 10, where a comparative evaluation of the proposed method is made against HiDDeN [24], TSDL [42], MBRS [44], and ReDMark [16] under various noise attack scenarios. The average BA values for HiDDeN, TSDL, MBRS, ReDMark, and our AGFI-GAN are 0.7997, 0.8062, 0.9377, 0.8534, and 0.9817, respectively, clearly illustrating that AGFI-GAN outperforms all other models. Specifically, AGFI-GAN shows an 18.2% improvement over HiDDeN and a 17.55% improvement over TSDL. When compared to ReDMark, AGFI-GAN achieves a 12.83% performance boost. Even though MBRS achieves a BA of 0.9377, AGFI-GAN surpasses it by 4.4%, establishing itself as the most robust model among those tested. Notably, AGFI-GAN demonstrates superior robustness against JPEG compression and Gaussian noise when compared to HiDDeN, TSDL, and ReDMark.

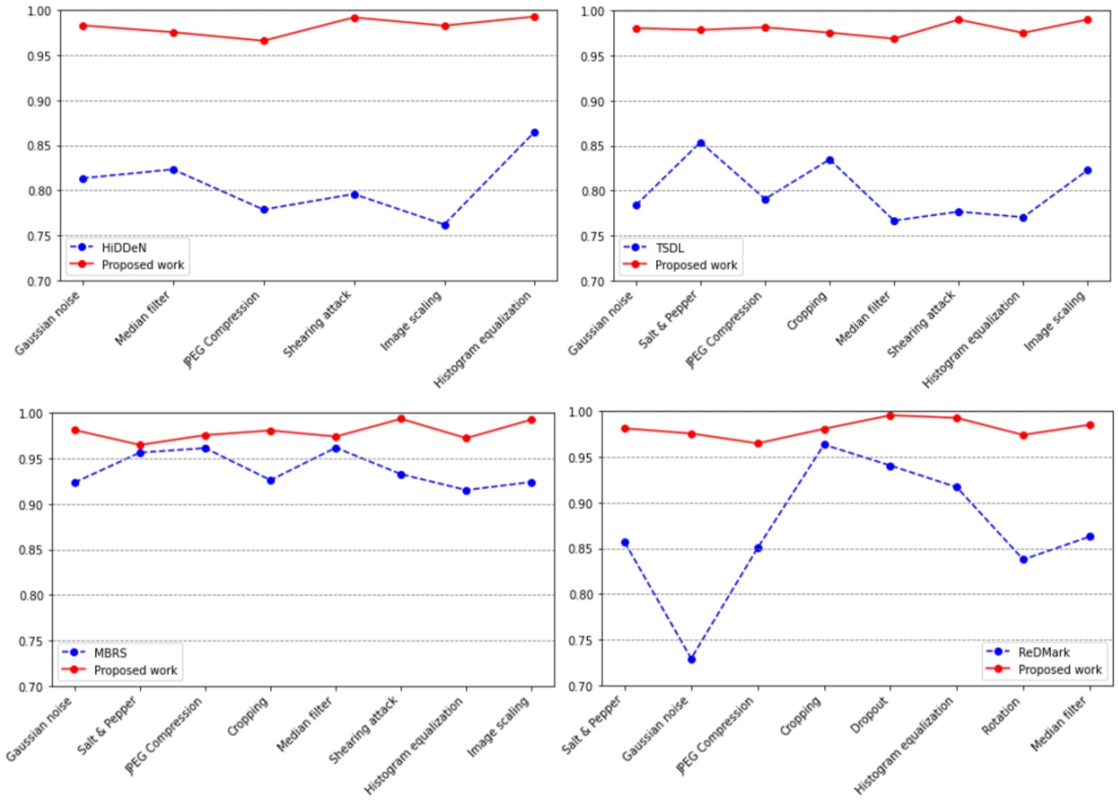

Figure 10. Comparative analysis of proposed work with HiDDeN [24], TSDL [42], MBRS [44], ReDMark [16]. The vertical axis represents bit accuracy.

For a watermarking system to be considered practical, it must demonstrate robustness against a wide range of image distortions, including those not encountered during training. To assess its ability to generalize to various image attacks, we evaluate the system using twelve types of untrained noise: Gaussian noise, JPEG compression, cropping, dropout, salt and pepper noise, rotation, median

 doi:10.20944/preprints202411.1711.v1

filtering, brightness adjustment, contrast adjustment, image scaling, shearing attacks, and histogram equalization.

As shown in Table 2, the robustness evaluation of our proposed AGFI-GAN model reveals superior performance over HiDDeN in the presence of various noise attacks. Specifically, under Gaussian noise, AGFI-GAN achieves nearly flawless performance with an accuracy of 95.61% at a noise density of 0.10, while HiDDeN's accuracy drops significantly to 75.48%. Similarly, in the case of JPEG compression, AGFI-GAN outperforms HiDDeN, with HiDDeN's accuracy reaching only 72.81% at a quality factor (QF) of 10, further highlighting the robustness of our model.

Table 2. Robustness assessment of the suggested work.

Attack	Noise density	HiDDeN [.%]	AGFI-GAN [.%]
Gaussian noise	0.001	82.96	100.00
	0.05	79.83	97.25
	0.10	75.48	95.61
JPEG Compression	QF=10	72.81	93.82
	QF=50	76.57	97.75
	QF=90	91.68	99.54
Cropping	[20, 20, 420, 420]	78.54	97.89
Dropout	0.3	85.26	98.73
Salt & Pepper	0.001	89.74	99.32
	0.05	81.92	98.41
	0.1	79.01	97.05
Rotation	45°	77.62	97.64
	90°	73.28	97.08
Median filter	[2, 2]	87.15	99.57
	[3,3]	79.34	96.49
Adjust Brightness	1.1	91.57	99.05
	1.3	79.85	96.78
Adjust Contrast	1.0	83.19	96.91
	2.0	74.63	95.26
Image scaling	0.5	74.71	98.53
	2	79.29	95.29
Shearing attack	[0.4, 0.4]	79.84	98.84
Histogram equalization	1.0	84.51	99.07

In comparison, AGFI-GAN achieves significantly better performance, with an accuracy of 93.82%. The performance differences are especially notable in several attack scenarios. Under a Dropout attack with a rate of 0.3, AGFI-GAN reaches an accuracy of 98.73%, while HiDDeN's accuracy is considerably lower at 85.26%. Similarly, during a shearing attack with parameters [0.4, 0.4], AGFI-GAN achieves an accuracy of 98.84%, far surpassing HiDDeN's accuracy of just 79.84%. For a 90° rotation, HiDDeN's accuracy drops to 73.28%, while AGFI-GAN attains a much higher accuracy of 97.08%. Under Salt and Pepper noise with a density of 0.1, HiDDeN's accuracy is 79.01%, whereas AGFI-GAN maintains a significantly better accuracy of 97.05%. Additionally, with a contrast adjustment factor of 2.0, HiDDeN's accuracy falls to 74.63%, while AGFI-GAN holds a higher accuracy of 95.26%. Overall, AGFI-GAN shows strong and consistent performance across a wide range of noise attacks, outperforming HiDDeN by a substantial margin, particularly in challenging distortion scenarios and high noise densities.

4.4. Impacts on Medical Imaging Detection

To evaluate the impact of the AGFI-GAN watermarking method on downstream tasks, we conducted a classification experiment using a watermarked Brain Tumor Detection dataset. The objective was to determine whether embedding watermarks would affect classification accuracy when employing a deep learning model. Specifically, we used a ResNet-50 model trained from scratch on the dataset over 100 epochs, with classification accuracy recorded at each epoch to detect any potential performance deviations due to watermarking.

 doi:10.20944/preprints202411.1711.v1

The results, shown in Figure 11, indicate that the test accuracy for the watermarked dataset remains consistently close to that of the original dataset across all epochs. This close alignment demonstrates that the AGFI-GAN watermarking method preserves the integrity of image features essential for classification. The negligible difference in accuracy between the two datasets supports the conclusion that our watermarking approach does not negatively impact classification performance, confirming its suitability for medical image authentication and protection without compromising the effectiveness of subsequent diagnostic tasks.

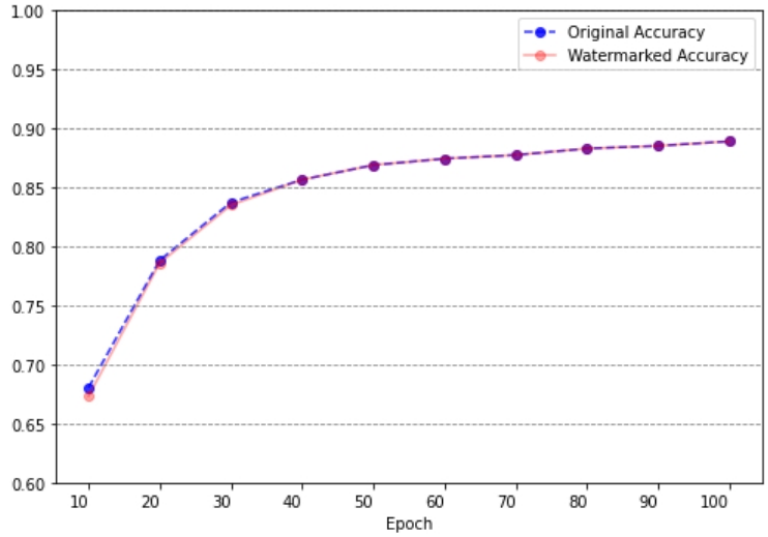

Figure 11. The Impact of Watermarking on the Classification Accuracy of Medical Images.

4.5. Security Analysis

We assume a security execution environment for medical image generation at the source side like Computed tomography (CT) and Magnetic resonance imaging (MRI). Thus, an adversary can not compromise medical imaging devices and then modify original images. We also assume a secure communication channel to deliver decoder to user side. Our approach integrates invisible and resilient watermarks into medical images at the data source. This ensures that data recipients can easily extract these watermarks and verify the authenticity of the images upon arrival. Since an attacker cannot influence the watermark embedding or extraction processes, they are unable to generate counterfeit medical images with valid watermarks. The security mechanisms inherent in the digital watermarking scheme effectively safeguard against forgery attacks within medical network systems.

5. Conclusions and Future Work

This study presents AGFI-GAN, an advanced model for secure and robust watermarking of medical images, designed to meet the unique challenges of maintaining both invisibility and robust against attacks in a digital healthcare environment. With the integration of an Attention-guided Module (AGM) and a Feature Integration Module (FIM), AGFI-GAN achieves significant improvements in watermark embedding without compromising image quality. The AGM effectively minimizes visual distortion by dynamically adapting the watermark's strength across various regions based on global feature analysis, reducing visibility in sensitive or smooth areas while allowing for stronger watermark embedding. This adaptive approach is further supported by the FIM, which enhances feature extraction by fusing shallow and deep layers, improving the model's robustness against image noise and attacks. The experimental evaluation validates AGFI-GAN's superiority over baseline models in terms of watermark imperceptibility, robustness , and classification accuracy on watermarked datasets. Results indicate that the framework preserves image fidelity essential for diagnostic tasks, confirming its compatibility with real-world medical image applications where accuracy is paramount.

 doi:10.20944/preprints202411.1711.v1

Future work could further explore optimizing the model for other image modalities and enhancing its resilience to emerging forms of digital attacks, thereby strengthening its applicability across diverse healthcare settings. Additionally, the AGFI-GAN model's computational demands currently limit its deployment on resource-constrained edge devices, commonly used in real-time medical imaging applications. Future research will focus on making the model lightweight and adaptable, possibly through model pruning, quantization, and other compression techniques that can reduce computational overhead while preserving watermark robustness and invisibility. This approach could enable deployment across a range of medical applications, including portable diagnostic tools and remote healthcare environments.

Author Contributions: Conceptualization, X.L., R.X and X.X.; methodology, X.L. and R.X.; software, X.L. and R.X.; validation, X.L., R.X and X.X.; formal analysis, X.L. and R.X.; investigation, X.L., R.X and X.X.; resources, X.L. and R.X.; data curation, X.L. and R.X.; writing—original draft preparation, X.L. and R.X.; writing—review and editing, X.L., R.X and X.X.; visualization, X.L. and R.X.; supervision, R.X.; project administration, R.X.; funding acquisition, R.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AEM	Attention-Enhanced Module
BN	Batch Normalization
DW	Digital Image Watermarking
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
AGFI-GAN	Feature-Integrated and Attention-Enhanced Model Based on GAN
FIM	Feature-Integrated Module
GAN	Generative Adversarial Network
PSNR	Peak Signal-to-Noise Ratio
ReLU	Rectified Linear Unit
SIFT	Scale-Invariant Feature Transform
SSIM	Structural Similarity Index Metric

References

1. Anand, A.; Singh, A.K. Watermarking techniques for medical data authentication: a survey. *Multimedia Tools and Applications* **2021**, *80*, 30165–30197.
2. Aherrahrou, N.; Tairi, H. PDE based scheme for multi-modal medical image watermarking. *Biomedical engineering online* **2015**, *14*, 1–19.
3. Singh, A.K.; Anand, A.; Lv, Z.; Ko, H.; Mohan, A. A survey on healthcare data: a security perspective. *ACM Transactions on Multimidia Computing Communications and Applications* **2021**, *17*, 1–26.
4. Alshanbari, H.S. Medical image watermarking for ownership & tamper detection. *Multimedia tools and applications* **2021**, *80*, 16549–16564.
5. Nosowsky, R.; Giordano, T.J. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule: implications for clinical research. *Annu. Rev. Med.* **2006**, *57*, 575–590.
6. Anand, A.; Singh, A.K. Hybrid nature-inspired optimization and encryption-based watermarking for e-healthcare. *IEEE Transactions on computational social systems* **2022**, *10*, 2033–2040.
7. Anand, A.; Singh, A.K. An improved DWT-SVD domain watermarking for medical information security. *Computer Communications* **2020**, *152*, 72–80.
8. Aparna, P.; Kishore, P.V.V. A blind medical image watermarking for secure e-healthcare application using crypto-watermarking system. *Journal of Intelligent Systems* **2019**, *29*, 1558–1575.
9. Liu, X.; Xu, R.; Chen, Y. A Decentralized Digital Watermarking Framework for Secure and Auditable Video Data in Smart Vehicular Networks **2024**.

10. Aparna, P.; Kishore, P.V.V. An efficient medical image watermarking technique in E-healthcare application using hybridization of compression and cryptography algorithm. *Journal of Intelligent Systems* **2018**, *27*, 115–133.
11. Ko, H.J.; Huang, C.T.; Horng, G.; Shiuh-Jeng, W. Robust and blind image watermarking in DCT domain using inter-block coefficient correlation. *Information Sciences* **2020**, *517*, 128–147.
12. Liu, J.; Huang, J.; Luo, Y.; Cao, L.; Yang, S.; Wei, D.; Zhou, R. An optimized image watermarking method based on HD and SVD in DWT domain. *IEEE Access* **2019**, *7*, 80849–80860.
13. Jin, T.; Zhang, W. A novel interpolated DFT synchrophasor estimation algorithm with an optimized combined cosine self-convolution window. *IEEE Transactions on Instrumentation and Measurement* **2020**, *70*, 1–10.
14. Liu, X.; Liu, Z.; Chatterjee, S.; Portfleet, M.; Sun, Y. Understanding human behaviors and injury factors in underground mines using data analytics. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021, pp. 2459–2462.
15. Liu, X.; Zhao, C. AGFA-Net: Attention-Guided and Feature-Aggregated Network for Coronary Artery Segmentation using Computed Tomography Angiography. *arXiv preprint arXiv:2406.08724* **2024**.
16. Ahmadi, M.; Norouzi, A.; Karimi, N.; Samavi, S.; Emami, A. ReDMark: Framework for residual diffusion watermarking based on deep networks. *Expert Systems with Applications* **2020**, *146*, 113157.
17. Guo, X.; Yang, W.; Zhang, L.; Shi, Y.; Li, J.; Sun, J.; Wan, W. Deep image watermarking with loss-driven modification. *Multimedia Tools and Applications* **2024**, *83*, 37665–37685.
18. Geng, H.; Zhou, M. Novel post-photographic technique based on deep convolutional neural network and blockchain technology. *The Journal of Supercomputing* **2024**, *80*, 6119–6139.
19. Fan, Y.; Li, J.; Bhatti, U.A.; Shao, C.; Gong, C.; Cheng, J.; Chen, Y. A multi-watermarking algorithm for medical images using inception v3 and dct. *CMC-Computers Materials & Continua* **2023**, *74*, 1279–1302.
20. Zhang, W.; Li, J.; Bhatti, U.A.; Liu, J.; Zheng, J.; Chen, Y.W. Robust multi-watermarking algorithm for medical images based on GoogLeNet and Henon map. *Comput. Mater. Contin* **2023**, *75*, 565–586.
21. Nawaz, S.A.; Li, J.; Shoukat, M.U.; Bhatti, U.A.; Raza, M.A. Hybrid medical image zero watermarking via discrete wavelet transform-ResNet101 and discrete cosine transform. *Computers and Electrical Engineering* **2023**, *112*, 108985.
22. Nawaz, S.A.; Li, J.; Bhatti, U.A.; Shoukat, M.U.; Li, D.; Raza, M.A. Hybrid watermarking algorithm for medical images based on digital transformation and MobileNetV2. *Information Sciences* **2024**, *653*, 119810.
23. Saxena, D.; Cao, J. Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)* **2021**, *54*, 1–42.
24. Zhu, J.; Kaplan, R.; Johnson, J.; Fei-Fei, L. Hidden: Hiding data with deep networks. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 657–672.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
26. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
27. Zhang, K.; Guo, Y.; Wang, X.; Yuan, J.; Ding, Q. Multiple feature reweight densenet for image classification. *IEEE access* **2019**, *7*, 9872–9880.
28. Adegun, A.A.; Viriri, S. FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images. *IEEE Access* **2020**, *8*, 150377–150396.
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.
30. Yu, C. Attention based data hiding with generative adversarial networks. Proceedings of the AAAI conference on artificial intelligence, 2020, Vol. 34, pp. 1120–1128.
31. Zhang, K.A.; Xu, L.; Cuesta-Infante, A.; Veeramachaneni, K. Robust invisible video watermarking with attention. *arXiv preprint arXiv:1909.01285* **2019**.
32. Zhang, H.; Wang, H.; Cao, Y.; Shen, C.; Li, Y. Robust data hiding using inverse gradient attention. *arXiv preprint arXiv:2011.10850* **2020**.
33. Singh, A.K.; Kumar, B.; Singh, G.; Mohan, A. *Medical image watermarking*; Springer, 2017.
34. Darabi, P.K. Brain Tumor Detection. [Online]. Available: https://www.kaggle.com/code/pkdarabi/brain-tumor-detection-by-cnn-pytorch/input, 2022.

 doi:10.20944/preprints202411.1711.v1

35. Tenebris97. liver-segmentation. [Online]. Available: https://www.kaggle.com/code/tenebris97/liver-segmentation-resnet-50/input, 2021.
36. Darabi, P.K. Diagnosis Of Pneumonia. [Online]. Available: https://www.kaggle.com/code/pkdarabi/diagnosis-of-pneumonia-by-cnn-classifier/input, 2024.
37. ahmadpk. Coronary Artery. [Online]. Available: https://www.kaggle.com/code/ahmadpk/heart-disease-classification/input, 2023.
38. Wang, B. Hand MRI. [Online]. Available: https://viterbi-web.usc.edu/~jbarbic/hand-mri-dataset/download.php, 2023.
39. Amine, M.M.E. Spine Segmentation. [Online]. Available: https://www.kaggle.com/datasets/pycadmk/spine-segmentation-from-ct-scans, 2024.
40. Bellavia, F.; Colombo, C. Is there anything new to say about SIFT matching? *International journal of computer vision* **2020**, *128*, 1847–1866.
41. Pele, O.; Werman, M. A linear time histogram metric for improved sift matching. Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III 10. Springer, 2008, pp. 495–508.
42. Liu, Y.; Guo, M.; Zhang, J.; Zhu, Y.; Xie, X. A novel two-stage separable deep learning framework for practical blind watermarking. Proceedings of the 27th ACM International conference on multimedia, 2019, pp. 1509–1517.
43. Luo, X.; Zhan, R.; Chang, H.; Yang, F.; Milanfar, P. Distortion agnostic deep watermarking. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 13548–13557.
44. Jia, Z.; Fang, H.; Zhang, W. Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression. Proceedings of the 29th ACM international conference on multimedia, 2021, pp. 41–49.
45. Anand, A.; Bedi, J.; Rida, I. MIWET: Medical image watermarking using encryption and fusion technique. *Computers and Electrical Engineering* **2024**, *115*, 109114.