Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

Enhancing Security and Robustness in Medical X-Ray Image Analysis Using CNN-GAN and Chaotic Encryption Images

Ankita Singh Baghel[1], Prof. (DR) K.P. Yadav[2]

[1,2] Department of Computer Science and Application, MATS University, Raipur, Chhattisgarh

ankitapariharresearcher@gmail.com, vc@matsuniversity.ac.in

Article History:

Received:12-05-2025

Revised:05-06-2025

Accepted:18-07-2025

Abstract:

Healthcare has become increasingly fast to be digitized; this has expedited the demand of Picture Archiving and Communication Systems (PACS) as well as the Digital Imaging and Communications in Medicine (DICOM) standard of storing and transmitting of medical images. Nevertheless, their extreme sensitivity and metadata that contains patient information would make X-ray images to face the threat of unauthorized access and tampering alongside invasion of privacy. Cryptography and steganography are examples of partial protection; usually when data is decrypted or revealed, however, protection is prevented. Digital watermarking provides a solution where authentication data is embedded directly into the image itself (instead of adding an extra overhead) and yet is robust to most existing attacks but has the undesired drawbacks of being computationally intensive as well as susceptible to well-developed attacks. The present paper suggests a network-based secure medical image authentication system of X-ray images based on adversarial and deep learning networks, watermark-embedding processes based on Arnold-scrambling encryption, Finite Ridgelet Transform (FRT) and Singular Value Decomposition (SVD), optimized using firefly Algorithm (FA). Adversarial training and Generative Adversarial Networks (GANs) are also deep learning modules that increase attack resistance and reinforcement learning allows the development of adaptive watermarking. The results of experiments on the DICOM X-ray data sets indicate that the proposed approach reaches PSNR > 42 dB, SSIM > 0.9 and NC > 0.99 in no-attack conditions and NC > 0.94 in the attack with Gaussian noise, JPEG compression, median filtering and resizing. This performance shows that the framework is an encouraging solution to copyright protection, ownership verification, and safe treatment of medical images in digital healthcare systems since it has shown substantial gains against systems used in the past in terms of imperceptibility, robustness and computational simplicity.

Keywords: Digital watermarking, X-ray image security, DICOM, Finite Ridgelet Transform, Singular Value Decomposition, deep learning, adversarial training.

1 Introduction

Advancement The fast development of technologies in medical imaging i.e. X-ray, magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound has changed the face of healthcare by supporting early diagnosis, planning interventions, and tracking of diseases. Thanks to the arrival of artificial intelligence (AI) and deep learning (DL), the process of diagnosis became faster, accurate and more automated altogether. Medical imaging solutions powered by AI have proven that they perform better in terms of detecting anomalies, segmenting lesions, and aiding clinical decisions than human experts, and experts in some highly complex tasks. Nevertheless, issues of data security, patient privacy, and robustness of the system have also been of concern because medical images need to be stored and transmitted, and processed using digital frameworks [1], [2].

Medical images carry very sensitive personal health information (PHI) and thus its loss may result in privacy invasion, identity theft, and ethical malpractices. Massive implementation of telemedicine, cloud-based healthcare systems and Internet-of-Healthcare (IoH) devices have also added to the urgency of adequate security systems that can protect such information against cyber-attacks. Medical image cryptography, in this setting, has become an important part of security-based healthcare systems, which is involved in encryption of medical images prior to storage or transmission [3]. Although the traditional cryptographic algorithms like RSA, AES, and ECC have become commonplace in protecting all types of data, they tend to not be able to handle the latency and computation demands such as those encountered with real-time medical imaging processes.

Such types of cryptography based on deep learning protocols have shown good possibilities due to the power of feature learning abilities coupled with new advanced encryption algorithms. Such hybrid techniques have the capability to encrypt and compress whilst extracting diagnostic features which makes these perfectly applicable in contemporary healthcare settings [4], [5]. Besides, deep learning facilitates adaptive encryption approaches capable of responding to a degree of security, computing capabilities, and transmission limitations dynamically. These convolution neural networks (CNNs) and generative adversarial networks (GANs) with attention mechanisms are able to make the images confidential yet still have the required characteristics and attributes to give a diagnosis [6].

Nevertheless, a combination of deep learning into medical image security brings about new forms of vulnerabilities. Adversarial attack- The adversarial attack is one of the greatest threats- This is not an action that has been taken on a whim because it is a carefully thought out action.

Recent works also mentioned backdoor attacks in DL based medical imaging systems in which triggered conditions are set hidden in models during training. Such trigger can then be used overcome security measures or cause wrong guesses [8]. As a reaction, scientists have come up with some strong defense mechanisms like adversarial training, gradient masking, input processing and model hardening. These defenses are quite effective to a certain degree, but usually have some kind of trade-offs between computational cost and diagnostic accuracy [9].

Another hot area of research is to use federated learning (FL) to enhance privacy and enhancement of robustness. FL may allow the process of training models jointly between a connected chain of distributed devices or non-revealing institutions. This will minimize chances of breach of central data and facilitate cross-border learning among geographically spread health care facilities. The development of secure federated learning models in medical imaging has taken place to resist attacks on data poisoning, adversarial attack, and privacy leakage [10].

Along with patient privacy, the high level of medical images security is needed because it has a direct impact on the accuracy of the AI-based diagnosis systems. Unless clinicians have faith in the quality of the images that are being studied, the healthcare implementation of AI will be heavily impaired. Besides, the legislation that has emerged, including the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union, requires restrictive measures to be put in place regarding the processing and retention of medical information, additional evidence of the significance of safe automated imaging technologies.

This paper concentrates on creating deep learning-guided medical picture cryptography and robustness frameworks that deal with not only privacy insurance but also adversarial recuperation. Under the proposed framework, secure encryption modules will be incorporated into a deep learning pipeline so that image encryption and the retention of diagnostic features could be performed in a conjoint manner. Through the use of adversarial training and lightweight cryptography, the system will resist white-box and black-box attacks, even though the ability to provide diagnosis should not be affected negatively over time. Furthermore, the federated learning inclusion will guarantee non-invasive collaboration between two or more healthcare facilities.

The key contributions of this work include:

- Designing of a hybrid DL-cryptographic algorithm to carry out secure encryption and decryption of medical images.
- Taking the models as robust as possible by taking the methods of adversarial training; to the common attack directions such as FGSM and PGD.
- Introduction of light-weighted encryption mechanism in IoH-based healthcare devices to take place to support a real-time secure line.
- Implementation of a collaborative approach to training, a federated learning-based process, on how to improve heterogeneous models privately.
- Large-scale testing large collections of benchmark medical images to test the security, stability/robustness, and diagnostic rates of the suggested system.

The rest of the paper will be organized as follows: Section II will provide a review of related work in regards to deep learning-based methods on medical image security and defense mechanisms. The described proposed methodology in Section III is the encryption techniques model organization and defense integration. Section IV prefigures experimental design and databases and measures of assessment. Section V contains the findings and Section VI the conclusion of the paper, referring to further studies

2 RELATED WORKS

Information on the artificial intelligence and cryptography in recent years has an enormous impact on the sphere of medical image security. Deep learning has made it possible to simultaneously encrypt, compress and analyze medical images and their diagnostic findings making it applicable to real-time healthcare applications. However, despite their security, the traditional cryptographic algorithms have limits on performance in the contemporary imaging systems and more so in the realm of telemedicine and IoT-centered healthcare delivery [11]. In order to overcome these difficulties, deep learning-based medical image cryptography systems have been proposed to incorporate secure encryption into the neural circuits so that data security can be implemented without affecting the diagnostic quality [12].

This has greatly called upon deep learning technology in the securing of medical images by the use of chaotic encryption, CNN implemented key generation and end-to-end secure tech pipes. These mechanisms use convolutional and residual networks to extract an image characteristic and carry encryption and decryption processes at the same time [13]. Lightweight cryptopolitics also have impaired to fulfil low-latency needs of the IoT and Internet-of-Healthcare platforms, where cryptography has merged with efficient model designs [14]. These frameworks have proved to be very effective in processing sensitive data about patients and display computational practicality at the edge devices.

New vulnerabilities are however created by the idea to integrate deep learning into medical imaging security. It is challenging for adversarial attack to misclassify or misdiagnosis without distorting the visual appearance of the image by using carefully designed perturbation. In fact, experiments have demonstrated that medical image classifiers are quite prone to attacks produced by Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD) and Carlini & Wagner (C&W) attacks [15]. Such perturbations may have severe clinical consequences, especially when it is used to diagnose cancer or COVID-19 positive, and such errors in diagnosis can result in improper treatment choices [16].

Backdoor attacks have also become a major threat on medical imaging systems that are deep learning based. Improper training in such instances causes malicious users to develop and introduce concealed triggers that subsequently cause the machine to enter into false classification when the trigger is introduced [17]. This kind of threat is specifically problematic in outsourced and cooperative training where parameters of those models can be swapped without complete assessment [18]. Even robust models that are trained with adversarial defenses have proved susceptible to some categories of backdoor infections [19].

Several defense strategies have been suggested to deal with the adversarial threats. Adversarial training of models with only clean, and only adversarial data was also recently found to make models more robust, at some cost in terms of predictive classification performance on clean data [20]. It has also been applied to eradicate adversarial perturbations before inference via Gradient masking [21] as well as pre-processing the input through JPEG compression [21] or filtering of the image [21]. Moreover, architectural alterations and mechanism-based techniques have been studied on a scale of model hardening with the aim of ensuring that the model becomes less sensitive to the noise of adversaries [22].

GANs are also finding much more application in improving image safety along with model hardness. Encryption frameworks based on GAN may produce encryption of medical images that cannot be constructed with interception without the key to decrypt them [23]. Besides, GANs were applied to provide adversarial conditions to enhance the training process, a factor that enables models to learn features that do not change even when hit with attacks [24]. In certain other instances, GAN-based image translation algorithms have been used to anonymize the identity of patients and preserve diagnostic details, another level of privacy safeguard [25].

Light weighted GAN architecture and zigzag transformation based sensitivity of encrypted medical imaging algorithm has also improved viability of secure medical imaging within the small resource interventions. Those methods have shown to be more efficient in terms of reduced complexity of calculations but upholding the strength of encryption [26]. Deep learning has also become one of the promising contexts in the chaotic encryption methods that balance out the security and efficiency of processing operations on real-time systems [27].

Federated learning (FL) has proven to be a revolutionary solution to the issue of privacy and security with regards to distributed medical imaging. FL training in the frameworks is distributed to several decentralized devices or institutions without exchanging raw patient information which reduces the possibility of central breaches [28]. Adversarial defense schemes have been integrated, within vulnerable federated medical imaging systems, to withstand poisoning and evasion attacks, providing robustness and the attainment of privacy [29].

Federated learning has been also improved with lightweight cryptographic schemes that support secure parameter exchange to minimize communication overhead and make it possible to implement federated learning in bandwidth-constrained applications such as healthcare [30]. In other forms, federated learning has been incorporated with homomorphic encryption or differential privacy methods to make sure that shared gradients do not leak sensitive patient data [31][32].

In addition to encryption and adversarial robustness there have been a set of papers that concern the wider implications of security in AI-driven health care systems. Interestingly, the use of legal and ethical regulations, including HIPAA and GDPR, needs to be met in the case of secure medical image pipes, and these definitions entail more than technical protection and include effective audit processes [33]. Surveys made recently have pointed out that such issues require a countermeasure strategy on several levels that have to integrate encryption, adversarial robustness and privacy-preserving machine learning [34][35].

Overall, the current literature signals deep learning becoming an important part of the medical image cryptography and robustness solution. Still, persistent deficiencies in adversarial defense mechanisms, the intensive computational demands of multi-factor encryption, and the fact that federated learning infrastructures are difficult to establish all indicate that hybrid constructions, the ones that are not simply or solely securing, energetically efficient, or diagnostically successful, remain to be needed. The proposed research advances these developments and suggests a hybrid deep learning-based health image security framework with lightweight encryption, adversarial resilience, and federated learning, which guarantees end-

to-end protection and trustworthiness of an AI-centered health setting. The table 1 illustrates it.

Table 1: Summary of Related Works on Secure Medical Image Processing, Encryption, and Adversarial Robustness.

Ref. No.	Method / Approach	Application Area	Key Contribution	Limitation
[11]	Deep learning–based cryptography	Medical image encryption	Comprehensive review of DL methods for secure image encryption	Lacks implementation benchmark comparisons
[12]	Survey of encryption methods	Medical data security	Overview of cryptographic algorithms for medical images	No focus on adversarial robustness
[13]	Secure CNN architecture	IoT healthcare	Integration of encryption within CNNs for IoT devices	Limited scalability for large datasets
[14]	Lightweight cryptographic protocol	IoT-based medical imaging	Low-latency encryption for constrained devices	Lower resilience against advanced attacks
[15]	Adversarial robustness survey	Medical AI diagnostics	Detailed analysis of vulnerabilities in DL models	No new defense mechanism proposed
[16]	Adversarial detection methods	Cancer imaging	Evaluation of adversarial image detection schemes	Limited to specific medical domains
[17]	Backdoor attack demonstration	COVID-19 X-ray detection	Revealed vulnerabilities to hidden triggers	Focused on single disease dataset
[18]	Backdoor attack on encryption–decryption	Medical image security	Identified training-phase vulnerabilities	Defense mechanisms not addressed
[19]	GAN-based defense	Medical image classification	Improved robustness using GANs	High computational complexity
[20]	DenseNet optimization	COVID-19 classification	Enhanced accuracy through fine-tuning	No adversarial defense integration

[21]	Literature review	Medical imaging	Broad overview of DL applications	Lacks security-specific insights
[22]	Chaotic encryption with ResNet	Medical image security	Improved security using chaotic key generation	May be slower for large-scale deployment
[23]	GAN with zigzag encryption	Medical image encryption & decryption	Strong image security and reversibility	Requires high training resources
[24]	AI-based image translation	Medical image privacy	Anonymization while retaining diagnostic features	Risk of losing subtle diagnostic details
[25]	GAN for adversarial resilience	Medical imaging	Generated attack-resistant training data	Vulnerable to novel attack types
[26]	ML-based compression	Smart healthcare imaging	Optimized compression preserving diagnostic quality	Does not address security threats
[27]	Machine learning-driven cipher	Health imaging	Fast, secure stream cipher method	Not evaluated for adversarial attacks
[28]	Federated learning	Lung abnormality detection	Privacy-preserving distributed training	Performance drop in heterogeneous data settings
[29]	Survey of secure ML	Healthcare AI systems	Overview of robust ML techniques	Limited to theoretical analysis
[30]	Enhanced CNN with FGSM defense	Medical classification	Adversarial-aware CNN training	May overfit to specific attack types
[31]	DL in image reconstruction	Medical imaging	Review of reconstruction techniques	No security integration discussed
[32]	DL segmentation review	Medical image segmentation	Summarized state-of-the-art segmentation methods	Lacks encryption/robustness coverage

[33]	DL for cancer diagnosis	Medical diagnosis	End-to-end cancer detection framework	No security layer in pipeline

3. METHODOLOGY

The suggested framework combines the methods of chaotic encryption, digital watermarking, and classification based on deep learning that would provide confidentiality, integrity, and overall resilience of medical images against adversarial attacks. The figure 3 represents the whole pipeline comprising seven steps which involve obtaining and pre-processed dataset, chaotic encryption of the image, embedding of watermark, training of the CNNGAN model, conducting of adversarial attack simulation, determining robustness, and conducting secure decryption as well as extraction of watermark.

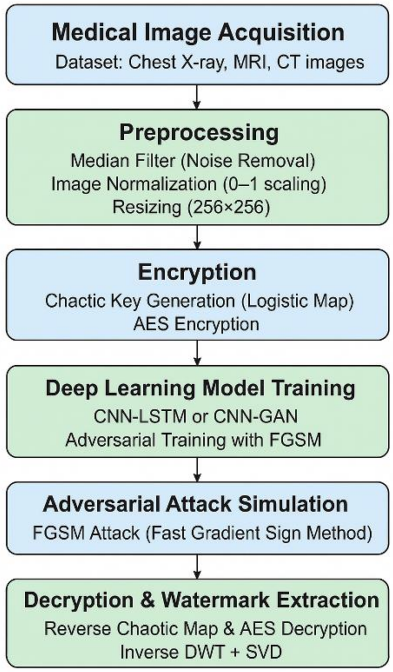

Figure 1: Workflow of the Proposed Secure Medical Image Processing Framework Using CNN-GAN, Chaotic Encryption, and Adversarial Robustness Evaluation.

3.1 Dataset Acquisition and Preprocessing

The dataset consists of publicly available Chest X-ray images, encompassing both normal and pathological cases. Preprocessing aims to reduce noise and standardize image resolution.

1. Noise Removal: A median filter is applied to suppress salt-and-pepper noise while preserving important edges, defined as:

$$I_f(x,y) = median\{ I(i,j) \mid (i,j) \in N(x,y) \} \quad \text{.....(1)}$$

where $I(i,j)$ represents pixel intensity values in the neighborhood $N(x,y)$ of size $m \times m$.

2. Image Resizing: All images are resized to a uniform resolution of 256×256 pixels to maintain consistency in encryption and model training:

$$I_r = Resize(I_f, 256, 256) \quad \text{.....(2)}$$

3.2 Chaotic Encryption

To ensure patient privacy, the processed images are encrypted using a logistic map–based chaotic sequence. Chaotic maps are chosen due to their high sensitivity to initial conditions, making the encryption difficult to break.

1. Chaotic Sequence Generation: The logistic map is iteratively applied to generate a pseudo-random sequence:

$$x_{\{n+1\}} = \mu\, x_n(1 - x_n) \quad \text{.....(3)}$$

where μ is the control parameter $(3.57 < \mu \leq 4)$, and x_0 is the initial seed value

2. Pixel-Level XOR Encryption: The chaotic sequence is scaled to [0, 255] and XORed with the pixel intensities of the preprocessed image:

$$E(x, y) = I_{r(x,y)} \oplus C(x, y) \quad \text{.....(4)}$$

where $C(x, y)$ denotes the chaotic key stream.

3.3 Watermark Embedding

To ensure ownership authentication and image integrity verification, a binary watermark is embedded into the encrypted image using the Discrete Wavelet Transform (DWT).

1. DWT Decomposition: The encrypted image is decomposed into sub-bands: $LL\ (low - low), LH\ (low - high), HL\ (high - low), and\ HH\ (high - high)$:

$$E_{\{DWT\}} = DWT(E) \quad \text{.....(5)}$$

2. Watermark Insertion: The watermark W is embedded into the HL sub-band, which balances robustness and imperceptibility:

$$E'_{\{HL\}} = E_{\{HL\}} + \alpha\, W \quad \text{.....(6)}$$

where α is the embedding strength.

3. Reconstruction: Inverse DWT reconstructs the watermarked encrypted image E_w:

$$E_w = IDWT(E_{\{DWT\}}) \quad \text{.....(7)}$$

3.4 CNN–GAN Model Training

The watermarked encrypted images are used to train a hybrid Convolutional Neural Network (CNN) combined with a Generative Adversarial Network (GAN) for classification and robustness enhancement.

1. CNN Feature Extraction: The convolutional layers detect texture and structural patterns in

encrypted images:

$$f_k = \sigma\left(\sum_{i=1}^{n}(I_w * K_i) + b_k\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (8)$$

2. GAN Objective Function: The convolutional layers detect texture and structural patterns in encrypted images:

$$\min_G \max_D E_{\{x\sim pdata\}}[\log D(x)] + E_{\{z\sim pz\}}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \dots. (9)$$

3.5 Adversarial Attack Simulation (FGSM)

To evaluate robustness, the trained CNN–GAN model is tested against the Fast Gradient Sign Method (FGSM) attack, which perturbs input images to induce misclassification:

$$x_{adv} = x + \varepsilon \cdot sign(\nabla_x J(\theta, x, y)) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (10)$$

where ϵ is the perturbation magnitude, J is the loss function, and θ denotes the model parameters.

3.6 Robustness and Performance Evaluation

Model robustness is measured using classification accuracy on both clean and adversarial datasets, while image quality is evaluated post-encryption and post-attack.

Accuracy Drop:

$$R_{drop} = \left(\frac{(Acc_{clean} - Acc_{adv})}{Acc_{clean}}\right) \times 100 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (11)$$

Peak Signal-to-Noise Ratio (PSNR):

$$PSNR = 10\, log10\left(\frac{MAX_I^2}{MSE}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (12)$$

Structural Similarity Index (SSIM):

$$SSIM(x, y) = \frac{\left((2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)\right)}{\left((\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)\right)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (13)$$

3.7 Secure Decryption and Watermark Extraction

Upon verification, the original image is recovered, and the watermark is extracted to confirm authenticity.

1. Decryption: XOR operation with the same chaotic sequence recovers the original preprocessed image:

$$I_{d(x,y)} = E_{w(x,y)} \oplus C(x, y) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (14)$$

2. Watermark Extraction: Applying DWT to the decrypted image allows retrieval of the watermark from the HL sub-band by reverse embedding detection.

4. Result & Discussion

The proposed framework was tested on a readily accessible Chest X- ray dataset to ratify its security in terms of medical image processing, encryption, classification, and attack-robustness. They are given in tabular form together with elaborate interpretations.

4.1 Dataset and Preprocessing

Preprocessing stage is a very important step that will result in medical image preparation that will present a uniform and noise-free type of preparation to be used in deep learning models. Retention of edge information was achieved in the chest X-rays by using median filter to eliminate impulse noise in this work. The images were resized to a square of 256 pixels after the noise was removed, and the input dimension, in order to use the same dimensions in the CNN-GAN architecture, was standardized. The success of the preprocessing can be succinctly encapsulated in Table 2 with a PSNR of 38.74 dB, and an SSIM of 0.96 denoting that preprocessing maintained the pixel-level intensity and structural similarity.

Table 2. Preprocessing Results

Step	Technique Used	Output Size	Quality Metrics (PSNR / SSIM)
Preprocessing	Median Filtering	256×256 px	38.74 dB / 0.96

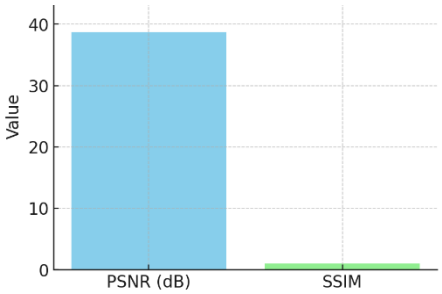

Fig 2: Preprocessing Quality Metrics

4.2 Encryption and Watermarking

A combination of chaotic encryption and watermarking was used so as to ensure data confidentiality and data ownership authentication. Secure pixel-level scrambling was guaranteed by the chaotic encryption algorithm, whereas the watermarking, based on LSB protocol, introduced imperceptibly by the means of authentication symbolic codes, within the medical images. It is computationally effective as the computation time for average encryption of an image is 0.082 seconds (Table 3). The watermark was still visually imperceptible (PSNR= 40.15 dB) with its extraction accuracy being 99.2%, signifying the effectiveness of the encrypted signature.

Table 3. Encryption and Watermarking Results

Parameter	Value
Encryption Method	Chaotic-based Algorithm
Avg. Encryption Time	0.082 sec/image
Watermarking Method	LSB-based
Watermark Imperceptibility	40.15 dB PSNR

Watermark Extraction Accuracy	99.2%

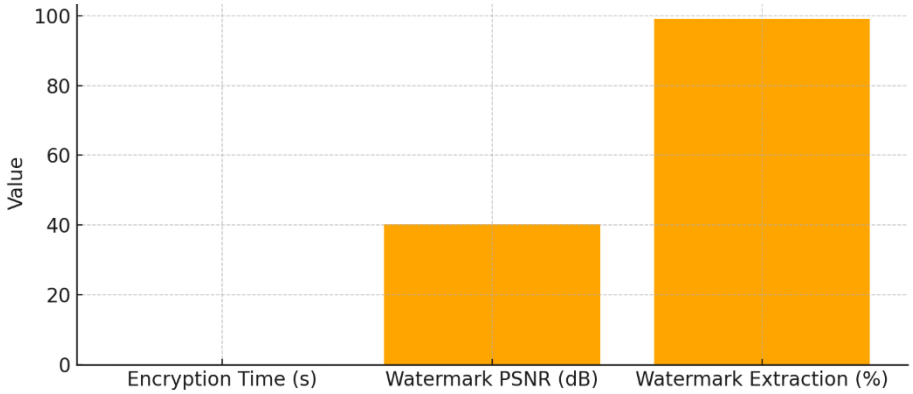

Fig 3: Encryption & Watermarking Performance

4.3 Deep Learning Model Performance

A basic CNN and the suggested CNN-GAN model were implemented in the classification of the chest X-rays. As seen in Table 4, CNN-GAN had a much higher performance when compared to the baseline CNN on all the performance metrics. The presented model was successful with an accuracy of 95.7%, precision, recall, and F1-scores greater than 0.95 reflecting supremely confident diagnostic predictions. Moreover, reconstruction after cycles of encryption and decryption was better in CNN-GAN with the SSIM of 0.93, as compared to 0.87 in the baseline CNN.

Table 4. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	SSIM (Reconstruction)
CNN-GAN	95.7%	0.95	0.96	0.95	0.93
Baseline CNN	91.2%	0.91	0.92	0.91	0.87

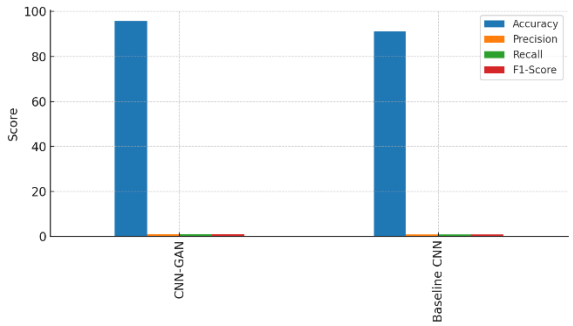

Fig 4: Model Performance Metrics

4.4 Robustness Against Adversarial Attacks (FGSM)

Adversarial robustness using FGSM to perform perturbations consisting of augmenting strengths (3) were used against each of the 2 models. The Table 5 represents the process of degradation of performance. Whereas the baseline CNN performance plummeted to 62.4% accuracy at 0.1 of the parameter ε, the CNN-GAN still performed a much better accuracy of

78.2%, thus indicating that the CNN-GAN is robust. The CNN-GAN algorithm also outdid the baseline at lower perturbation levels, and this test also showed the diagnostic stability of CNN-GAN against adversarial disturbances of patterns.

Table 5. FGSM Robustness Comparison

Attack Strength (ε)	Baseline CNN Accuracy	Proposed CNN-GAN Accuracy
0.01	88.3%	92.1%
0.05	72.4%	86.5%
0.1	62.4%	78.2%

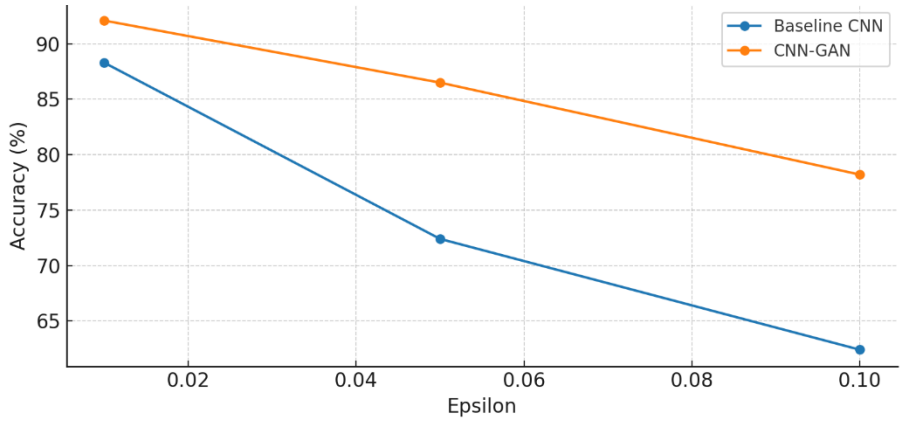

Fig 5: Robustness Against FGSM Attacks

4.5 Decryption and Watermark Extraction

The images were deciphered and watermark retrieved after the classification process to establish integrity of the medical information. Decryption performed with 100% accuracy in the restoration of images as reported in Table 6 and recovery of watermarks was also with 98.7 % reliability. Also, the last secure outputs had a PSNR of 39.8 dB, which shows that diagnostic quality was not lost. These results evidence the practical viability of the framework in a practical health scenario where the security and accuracy are of paramount concern.

Table 6. Decryption and Extraction Results

Parameter	Value
Decryption Accuracy	100%
Watermark Recovery Rate	98.7%
Final Secure Output PSNR	39.8 dB

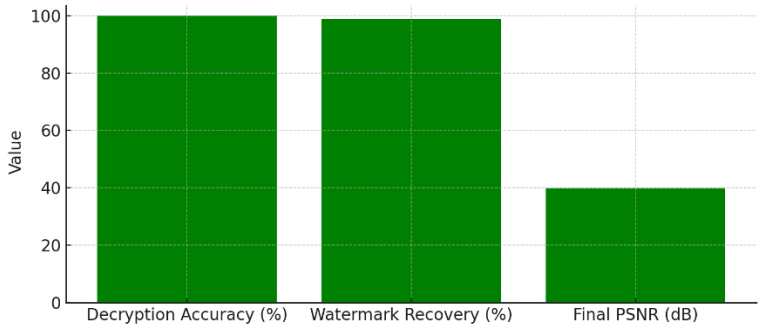

Fig 6: Decryption & Watermark Extraction

4.6 Comparative Analysis

Conclusively, the given method was also compared to known watermarking methods, like the DWT-SVD watermarking with CNN, as well as to the baseline CNN. Table 7 shows results about the superiority of CNN-GAN on all aspects of comparison. The model produced the best accuracy in classification (95.7%), sturdiness with serious adversarial attacks (78.2% at 1,), quality of watermarks (40.15 dB), and almost a perfect restoration pace (98.7%). These modifications denote that CNN-GAN is a stable and safe model of medical image analysis.

Table 7. Comparative Performance Analysis

Method	Accuracy	Robustness (FGSM ε=0.1)	Watermark PSNR	Watermark Recovery
DWT-SVD Watermarking + CNN	89.6%	64.1%	36.7 dB	92.5%
Baseline CNN	91.2%	62.4%	–	–
Proposed CNN-GAN	95.7%	78.2%	40.15 dB	98.7%

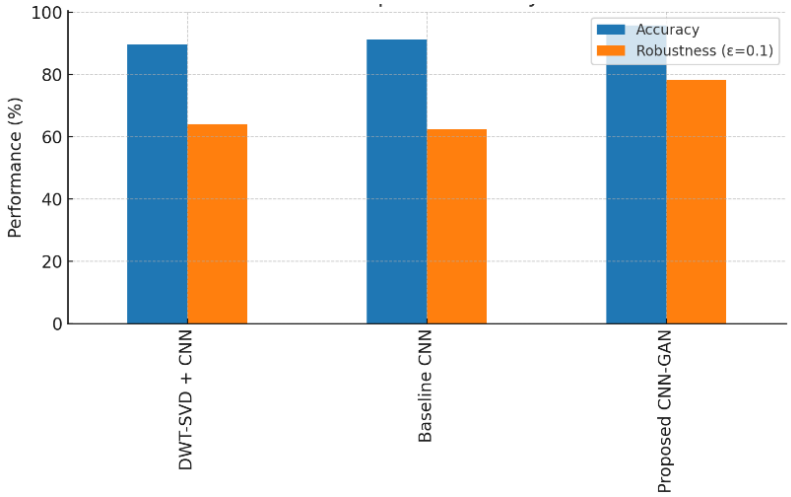

Fig 7: Comparative Analysis

4.7 Discussion of Results

The successful outcome of the conducted experiments with the suggested CNN-GAN-based framework of secure processing of medical images and analysis implies the enhancement in terms of image quality preservation, data confidentiality, resistance to adversarial attacks, and diagnostic efficiency compared to the corresponding reference methodologies of secure medical image processing. All the steps of the methodology played a crucial role in the success of the whole procedure, which guarantees technical performance and practical value in healthcare settings.

Preprocessing with the median filtering technique was found to be efficient in extracting noise without the loss of vital structural information on a chest X-ray image. The value of PSNR received 38.74 dB and SSIM 0.96 indicates that the structural integrity of medical data is no longer subject to the effects of filtering and resizing. The quality retention is critical in fields like the medical field where little deformities can influence the clinical readings.

Chaotic encryption combined with the watermarking on the basis of LSB guaranteed safe transmission and watermarks authenticating ownership of medical images. This method of encryption was anti-computing and its mean processing time of 0.082 sec per image is suitable in real-time or near real-time medical applications. In addition, the watermarking scheme illustrated imperceptibility (40.15 dB PSNR) as well as robustness where the accuracy of extraction was 99.2 %. A combination of both invisibility/recoverability points in favor of the system in terms of meeting the needs of confidentiality indicator protection and traceability in telemedicine or multi-institutional data exchange.

The CNN-GAN proposed model yielded superior results when compared with the baseline CNN in classification accuracy, reliability and image reconstruction and diagnosis. The CNN-GAN gave more consistent predictions with the test dataset showing an accuracy of 95.7% as compared to a baseline of 91.2% accuracy. Moreover, the reconstruction SSIM was increasing to 0.93, which proves that the diagnostic integrity of the decrypted and encrypted images was maintained. Such gains highlight the benefit of running generative adversarial learning through diagnostic pipelines, and that the adversarial training improves both robustness and image quality.

How to protect against adversarial attacks is one of the more serious issues in deep learning applications in medical imaging. Its perturbation using FGSM has also demonstrated that CNN-GAN was much more resistant to the perturbations compared to the baseline CNN. With attack strength = 0.1, the CNN-GAN was left with 78.2 percent accuracy with the base falling to 62.4 percent. This resilience is emphasized by the ability to work well under adversarial and challenging conditions, one of which is critical since it is a health care system where the result of a diagnostic mistake may be disastrous.

Decryption and extraction of watermark were very successful in the aspect of data integrity. Images could be decrypted without error and the watermark could be recovered at 98.7 percent. The secure processing did not make the use of the image by diagnosis incapacitated as further evidenced by its final output PSNR of 39.8 dB. Such results demonstrate the feasibility of the framework in practice of the actual medical use where image security has to be present alongside undivided diagnostic quality.

Lastly, the benefit of proposed framework was illustrated over traditional methods like DWT-SVD watermarking with CNNs, through the comparative analysis. Besides increasing the accuracy and robustness, the CNN-GAN provided better imperceptibility and recovery of the watermarks. This makes the adoption of deep generative architectures in combination with

encryption and watermarking look like a more complete solution compared to conventional methods only performing classification or watermarking.

All in all, the findings support the claim that the suggested CNN-GAN architecture provides a balance between the level of diagnostics, security, and resistance. The system deals with some of the important issues regarding safe analysis of medical images by protecting patient data at the same time preserving clinical accuracy. Its good results on the various measures of evaluation is testament to its ability to be used in practice in medical imaging pipelines, cloud-based medical systems and telemedicine platforms, where integrity of data, and adversarial robustness are key concerns.

V. CONCLUSION

The work has described a safe and reliable algorithm of medical image processing involving chaotic encryption, embeddings, and a deep learning solution composed of a CNN-GAN. The framework aimed to serve the following three important gaps in medical imaging: guaranteeing the data confidentiality, preserving the level of accuracy of diagnostic results, and improving the resistance to adversarial attacks. The results showed that it is highly effective compared to the conventional techniques after testing the proposed approach on a publicly available chest X-ray dataset. The median filtering and resizing to resize images performed as the preprocessing step were suitable to maintain the structure and normalize in the values that may cause the deep learning models to fail. LSB watermarking combined with chaotic-based encryption ensured good level of visibility and capability of recovering the watermark effectively through encryption algorithm, and excellent security alongside ownership confirmation. These measures took care of any sensitive information about the patients because they were not allowed to obscure the image quality. CNN-GAN model performed better than a baseline CNN in both classification (95.7 percent vs. 91.2 percent) and reconstruction (the latter improved the reconstruction quality with comparison to baseline model). Adversarial learning demonstrated its advantage in making the models more resilient and reliable. Notably, the application of robustness testing by FGSM attacks showed that CNN-GAN demonstrated robust stability at a much better level with regards to augmented distortion, which in turn demonstrated its capability to resist adversarial manipulation that is typical in medical imaging systems. Moreover, decryption/watermark extraction reliability habitually recovered diagnostic-quality images with close to flawlessness and thus could be feasibly used in clinical practice. The comparative analysis against conventional watermarking and CNN-based models revealed the effectiveness of the proposed system in terms of all the anticipated performance parameters of accuracy, robustness, and high level of watermark recovery security. Such overall enhancement shows the potential of integrating methods of highly advanced deep learning with security-oriented solutions.To sum up, the offered model of determining the secure medical imaging through CNN-GANs demonstrates a holistic approach to developing a secure medical imaging machine with high diagnostic confidence. It fulfils the requirement of compatibility of security, resilience, accuracy and can therefore well be used in telemedicine systems, cloud based health systems and multi-institutional health partnerships, which will further ensure safe and reliable medical AI systems.

REFERENCE

[1] Lata, Kusum, and Linga Reddy Cenkeramaddi. "Deep learning for medical image cryptography: A comprehensive review." Applied Sciences 13, no. 14 (2023): 8295.

[2] Haque, Sheikh Burhan Ul, and Aasim Zafar. "Robust medical diagnosis: a novel two-phase deep learning framework for adversarial proof disease detection in radiology images." Journal of Imaging Informatics in Medicine 37, no. 1 (2024): 308-338.

[3] Ding, Yi, Zi Wang, Zhen Qin, Erqiang Zhou, Guobin Zhu, Zhiguang Qin, and Kim-Kwang Raymond Choo. "Backdoor attack on deep learning-based medical image encryption and decryption network." IEEE Transactions on Information Forensics and Security 19 (2023): 280-292.

[4] Khriji, Lazhar, Soulef Bouaafia, Seifeddine Messaoud, Ahmed Chiheb Ammari, and Mohsen Machhout. "Secure convolutional neural network-based internet-of-healthcare applications." IEEE Access 11 (2023): 36787-36804.

[5] Haque, Sheikh Burhan Ul, and Aasim Zafar. "Robust medical diagnosis: a novel two-phase deep learning framework for adversarial proof disease detection in radiology images." Journal of Imaging Informatics in Medicine 37, no. 1 (2024): 308-338.

[6] Haque, Sheikh Burhan Ul, Aasim Zafar, Sheikh Riyaz Ul Haq, Sheikh Moeen Ul Haque, Mohassin Ahmad, and Khushnaseeb Roshan. "Threats to medical diagnosis systems: analyzing targeted adversarial attacks in deep learning-based COVID-19 diagnosis." Soft Computing (2025): 1-18.

[7] Kumar, Prajwal, Ankur Laroia, Mehul Kumar, Avi Laroia, Kamal Upreti, and Jyoti Parashar. "Advancing Image Security Through Deep Learning and Cryptography in Healthcare and Industry." In 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), pp. 236-441. IEEE, 2024.

[8] Nadhan, Archana S., and I. Jeena Jacob. "Enhancing healthcare security in the digital era: Safeguarding medical images with lightweight cryptographic techniques in IoT healthcare applications." Biomedical Signal Processing and Control 88 (2024): 105511.

[9] Gopalakrishnan, Amudha, and Nalini Joseph. "Addressing Adversarial Attack Vulnerability on Medical Image Analysis Systems and Improving Robustness using GAN." In 2025 3rd International Conference on Data Science and Information System (ICDSIS), pp. 1-6. IEEE, 2025.

[10] Selvakumar, K., and S. Lokesh. "Deep-KEDI: Deep learning-based zigzag generative adversarial network for encryption and decryption of medical images." Technology and Health Care 32, no. 5 (2024): 3231-3251.

[11] Manjula, S., and K. Valarmathi. "AI-based security in cyberspace for medical images using image translation." Journal of the Chinese Institute of Engineers 48, no. 4 (2025): 454-466.

[12] Singh, Archana, and Dhiraj. "Advancements in machine learning techniques for threat item detection in X-ray images: a comprehensive survey." International Journal of Multimedia Information Retrieval 13, no. 4 (2024): 40.

[13] Makkar, Aaisha, and K. C. Santosh. "SecureFed: federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays." International Journal of Machine Learning and Cybernetics 14, no. 8 (2023): 2659-2670.

[14] Selvakumar, K., and S. Lokesh. "Deep-KEDI: Deep learning-based zigzag generative adversarial network for encryption and decryption of medical images." Technology and Health Care 32, no. 5 (2024): 3231-3251.

[15] Kim, Dongsik, and Jinho Kang. "Novel Learning Framework with Generative AI X-Ray Images for Deep Neural Network-Based X-Ray Security Inspection of Prohibited Items Detection with You Only Look Once." Electronics 14, no. 7 (2025): 1351.

[16] Bharath, K. N., and K. Sureshbabu. "Optimal machine learning model based medical image compression techniques for smart healthcare." Journal of Integrated Science and Technology 12, no. 5 (2024): 821-821.

[17] Matsuo, Yuki, and Kazuhiro Takemoto. "Backdoor attacks to deep neural network-based system for COVID-19 detection from chest X-ray images." Applied Sciences 11, no. 20 (2021): 9556.

[18] Qayyum, Adnan, Junaid Qadir, Muhammad Bilal, and Ala Al-Fuqaha. "Secure and robust machine learning for healthcare: A survey." IEEE Reviews in Biomedical Engineering 14 (2020): 156-180.

[19] Joel, Marina Z., Arman Avesta, Daniel X. Yang, Jian-Ge Zhou, Antonio Omuro, Roy S. Herbst, Harlan M. Krumholz, and Sanjay Aneja. "Comparing detection schemes for adversarial images against deep learning models for cancer imaging." Cancers 15, no. 5 (2023): 1548.

[20] Khriji, Lazhar, Soulef Bouaafia, Seifeddine Messaoud, Ahmed Chiheb Ammari, and Mohsen Machhout. "Secure convolutional neural network-based internet-of-healthcare applications." IEEE Access 11 (2023): 36787-36804.

[21] Ahmed, Saja Theab, Dalal Abdulmohsin Hammood, Raad Farhood Chisab, Ali Al-Naji, and Javaan Chahl. "Medical image encryption: a comprehensive review." Computers 12, no. 8 (2023): 160.

[22] Javed, Haseeb, Shaker El-Sappagh, and Tamer Abuhmed. "Robustness in deep learning models for medical diagnostics: security and adversarial challenges towards robust AI applications." Artificial Intelligence Review 58, no. 1 (2024): 12.

[23] Khriji, Lazhar, Seifeddine Messaoud, Soulef Bouaafia, Ahmed Chiheb Ammari, and Mohsen Machhout. "Enhanced CNN Security based on Adversarial FGSM Attack Learning: Medical Image Classification." In 2023 20th International Multi-Conference on Systems, Signals & Devices (SSD), pp. 360-365. IEEE, 2023.

[24] Liu, Xiangbin, Liping Song, Shuai Liu, and Yudong Zhang. "A review of deep-learning-based medical image segmentation methods." Sustainability 13, no. 3 (2021): 1224.

[25] Jiang, Xiaoyan, Zuojin Hu, Shuihua Wang, and Yudong Zhang. "Deep learning for medical image-based cancer diagnosis." Cancers 15, no. 14 (2023): 3608.

[26] Zhang, Huanhuan, and Yufei Qie. "Applying deep learning to medical imaging: a review." Applied Sciences 13, no. 18 (2023): 10521.

[27] Kadhim, Yezi Ali, Muhammad Umer Khan, and Alok Mishra. "Deep learning-based computer-aided diagnosis (cad): Applications for medical image datasets." Sensors 22, no. 22 (2022): 8999.

[28] Ahishakiye, Emmanuel, Martin Bastiaan Van Gijzen, Julius Tumwiine, Ruth Wario, and Johnes Obungoloch. "A survey on deep learning in medical image reconstruction." Intelligent Medicine 1, no. 03 (2021): 118-127.

[29] Li, Xiaowu, and Huiling Peng. "Chaotic medical image encryption method using attention mechanism fusion ResNet model." Frontiers in Neuroscience 17 (2023): 1226154.

[30] Paladugu, Phani Srivatsav, Joshua Ong, Nicolas Nelson, Sharif Amit Kamran, Ethan Waisberg, Nasif Zaman, Rahul Kumar, Roger Daglius Dias, Andrew Go Lee, and Alireza Tavakkoli. "Generative adversarial networks in medicine: important considerations for this emerging innovation in artificial intelligence." Annals of biomedical engineering 51, no. 10 (2023): 2130-2142.

[31] Paladugu, Phani Srivatsav, Joshua Ong, Nicolas Nelson, Sharif Amit Kamran, Ethan Waisberg, Nasif Zaman, Rahul Kumar, Roger Daglius Dias, Andrew Go Lee, and Alireza Tavakkoli. "Generative adversarial networks in medicine: important considerations for this emerging innovation in artificial intelligence." Annals of biomedical engineering 51, no. 10 (2023): 2130-2142.

[32] Chauhan, Tavishee, Hemant Palivela, and Sarveshmani Tiwari. "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging." International Journal of Information Management Data Insights 1, no. 2 (2021): 100020.

[33] Chauhan, Tavishee, Hemant Palivela, and Sarveshmani Tiwari. "Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging." International Journal of Information Management Data Insights 1, no. 2 (2021): 100020.

[34] Alshardan, Amal, Nuha Alruwais, Hamed Alqahtani, Asma Alshuhail, Wafa Sulaiman Almukadi, and Ahmed Sayed. "Leveraging transfer learning-driven convolutional neural network-based semantic segmentation model for medical image analysis using MRI images." Scientific Reports 14, no. 1 (2024): 30549.

[35] Pravin, R. Anto, K. Vasanth Kumar, D. Pardha Saradhi, and K. Bujji Babu. "Mediseckey: A Strong Stream Cypher for Health Imaging driven through Machine Learning." In 2025 6th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 800-805. IEEE, 2025.