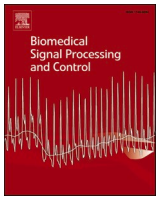
Biomedical Signal Processing and Control 92 (2024) 105967

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



A reversible fragile watermarking technique using fourier transform and Fibonacci Q-matrix for medical image authentication

Riadh Bouarroudj [a,*], Feryel Souami [a,b], Fatma Zohra Bellala [a], Nabil Zerrouki [c]

[a] LRIA, Department of Computer Science, University of Science and Technology Houari Boumediene (USTHB), B.P. 32, El Alia 16111, Bab Ezzouar, Algiers, Algeria
[b] Department of Computer Science, University of Algiers 1, Algiers 16000, Algeria
[c] Center for development of advanced technologies (CDTA), Baba Hassen, Algiers, Algeria

ARTICLE INFO

Keywords:
Fragile watermarking
Reversible data-hiding
Medical image authentication
Self-embedding
Watermark encryption
Fibonacci Q-matrix

ABSTRACT

Image authentication techniques are crucial for a multitude of multimedia applications. When images are transmitted through non-secure channels like the internet, they are vulnerable to unauthorized alterations and manipulations. Therefore, it is essential in sensitive fields, like the medical industry, to employ image authentication techniques to guarantee the integrity and authenticity of transmitted images. In this paper, to address these security issues and to ensure the authenticity of medical images, we present a reversible fragile watermarking model where the watermark is first generated from the cover image using DCT and then encrypted by the Fibonacci Q-matrix technique, enhancing the model's security. Subsequently, DFT is performed on the host image and the obtained sub-band is segmented into 2×2 blocks, the watermark is then embedded within the frequency coefficients using a new embedding technique. Experimental results indicates that the presented model produces a great watermarked image quality, with a PSNR superior to 117 dB, while maintaining an acceptable embedding capacity of 0.25 BPP and high sensitivity to various attacks, as it was tested against 16 different attacks and could detect tampering in all of them. Furthermore, the presented scheme offers nearly perfect reversibility with a PSNR of 324 dB, which is very useful in the medical field where we need images with as little distortion as possible.

1. Introduction

With the rapid development of the internet and multimedia applications, the transfer of digital images has become much faster and easier, enabling the emergence of new services such as telemedicine. However, the transmission of confidential images through insecure environments like the internet exposes them to potential manipulations and attacks. In these situations, many security issues can be encountered, one can cite copyright protection, integrity, and authentication. These security issues are critical in sensitive domains like the medical field, where even a minor alteration in the medical image can have a huge impact on decision-making. To solve these security issues, data-hiding techniques like digital watermarking have been developed to guarantee the safety of transmitted data.

Digital watermarking is a data-hiding technique that involves hiding an information known as the watermark (generally an image), inside another image called the cover image (or host image) without causing any distortion to the cover image. The obtained image, known as the watermarked image, is then sent to the receiver who can extract the embedded watermark to verify if the watermarked image is altered or not. Watermarking algorithms can be classified into robust, fragile, and semi-fragile approaches. Robust watermarking methods [1–6] are employed to assure copyright protection due to their resistance to attacks; the extracted watermark is successfully extracted despite the encountered manipulations. For instance, Gangadhar et al. [1] introduced a robust watermarking technique to ensure the security of medical images. First, the Improved Discrete Wavelet Transform (IDWT) is performed on the cover image, the low-frequency sub-band (LL) is then selected and decomposed into four equal size blocks. Subsequently, the suitable regions for embedding are selected utilizing entropy. Finally, Singular Value Decomposition (SVD) is performed on each block, and the watermark is embedded inside the first column of the U matrix. In contrast, fragile watermarking techniques [7–18] are employed to assure integrity and authentication since any manipulation of the watermarked image will cause the loss of the embedded watermark, as illustrated by the scheme introduced by Nguyen et al. [7] for image

* Corresponding author.
E-mail address: rbouarroudj@usthb.dz (R. Bouarroudj).

https://doi.org/10.1016/j.bspc.2024.105967
Received 1 November 2023; Received in revised form 2 January 2024; Accepted 28 January 2024
Available online 12 February 2024

authentication and tamper localization. The cover image is first segmented into 8 × 8 blocks, and a first-level DWT is performed on each block. Afterward, the three sub-bands LL, LH, and HL go through a second-level DWT, and one watermark bit is inserted into each second-level low-frequency sub-band (LL2, LHLL, HLLL). Semi-fragile watermarking techniques [19–21] are similar to fragile watermarking ones since they are also sensitive to malicious attacks but they tolerate certain non-malicious manipulations like JPEG compression. In this context, Tuncer et al [20] proposed a semi-fragile watermarking technique where Integer Wavelet Transform (IWT) is first performed on the cover image, and the HH sub-band is selected and segmented into 4 × 4 blocks. For each block, the Manhattan distance is then computed and the element with the minimum absolute distance is selected using the Center Symmetric Local Binary Pattern (CSLBP) to insert the watermark. This latest was generated using logistic map based PRNG and Advanced Encryption Standard (AES).

Watermarking techniques can be also classified according to the domain in which the watermark is embedded. Spatial domain techniques [8,15,18] directly embed the watermark within the cover image pixels using embedding techniques like the Least Significant Bit (LSB) substitution, as illustrated by the model proposed by Gul et al [8] where the cover image is segmented into 32 × 32 blocks. Afterward, SHA-256 hash technique is performed on the blocks and the result is embedded into the element's LSBs. Frequency domain techniques [5,6,12,16], also known as transform domain techniques, apply a frequency transformation to the original image to obtain a frequency sub-band, within which the watermark is inserted. Subsequently, an inverse frequency transformation is performed to obtain the watermarked image. Among the most frequently used transformation, we can mention, Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), and Discrete Wavelet Transform (DWT) which was used in the scheme proposed by Nguyen [17] to ensure the images' authenticity. The host image is first segmented into 8 × 8 blocks and DWT is performed on each one of them. Afterward, the LL sub-band undergoes an SVD transform to obtain U and V matrices, and the watermark is hidden into one of the two matrices using 1D DCT.

Furthermore, watermarking approaches are categorized as reversible or non-reversible approaches. In Reversible methods [10,12,15], the cover image can be reconstructed from the watermarked version if it is unaltered like the model presented by Azizoglu et al. [11] where the cover image is first segmented into 4 × 4 blocks and the blocks are scrambled using the Knuth shuffle algorithm. Afterward, DWT is performed on each block and the average values of the LL sub-bands are passed to the MD5 hash function to obtain a 128-bit code. Only the first 9 bits are selected and inserted in the LH, HL, and HH sub-bands with a capacity of 3 bits inside each sub-band. Non-reversible techniques [3,4,21], on the other hand, can't recover the original image from the watermarked version. Watermarking methods can be also divided into blind and non-blind techniques. In non-blind techniques [13,14], the receiver requires the cover image during the extraction process as illustrated by the model presented by Nejati et al. [13] for the authentication of grayscale images. Fourier transform is first employed on the cover image, and QR factorization is then performed on both the obtained sub-band and the watermark. Subsequently, the R matrix of the watermark is incorporated into the R matrix of the cover image following an embedding strength alpha. Whereas, blind techniques [2,12,19] don't need the original image during extraction.

In situations where the watermarked image is attacked, certain techniques can localize the tampered areas and even recover the original pixel values [8,9,19] like the scheme proposed by Su et al. [9]. First, the cover image is segmented into 2 × 2 blocks, the mean of each block is then computed and the 6 MSBs of the result are selected. Once all the blocks are processed, the result undergoes an Arnold transformation to obtain the recovery code, which is inserted inside the cover image using Turtle Shall-based Data Hiding (TSDH). Finally, the authentication code is also computed by dividing the image into 4 × 4 blocks, Principal Component Analysis (PCA) is then applied to each block, and the result is scrambled using the Arnold transformation. The obtained code is then embedded inside the image to obtain the watermarked version.

Based on recent reviews of fragile watermarking methods [22,23,24] and our thorough research, we identified some research gaps that motivated our study. Despite previous extensive research in this field, it remains uncommon to find fragile watermarking models that yield a high watermarked image quality (with a PSNR > 100db) while keeping an acceptable payload attaining 0.25 BPP. To overcome these limitations we propose a reversible fragile watermarking algorithm for medical image authentication. The main goal of this approach is to achieve a great watermarked image quality while maintaining an acceptable payload and a high sensitivity to attacks. The proposed model first generates the watermark from the cover image using DCT and then applies the Fibonacci Q-matrix technique to scramble the watermark, enhancing its security. Subsequently, DFT is performed on the host image, and the obtained sub-band is segmented into 2 × 2 blocks. The scrambled watermark is then embedded into each block with a capacity of 0.25 BPP. Finally, an inverse DFT is employed to obtain the watermarked image. Experimental results illustrate that the presented scheme is sensitive to various attacks whether they are malicious or not, while maintaining a great watermarked image quality and an acceptable embedding capacity. Furthermore, the proposed method provides nearly perfect reversibility, which is helpful in sensitive domains like the medical field. The main contributions of the presented approach are summarized as follows:

- Providing a great watermarked image quality while keeping an acceptable payload of 0.25 BPP and a high sensitivity against various attacks.
- Achievement of nearly perfect reversibility with restored images reaching a PSNR of 327 dB, which is very useful in sensitive domains like the medical field.
- The use of a novel embedding technique involving adding/subtracting a constant value to/from the element with the absolute maximum value in each block if the watermark bit is equal to 1; otherwise, no modification is made. A vector "Max" containing the absolute maximum value of each block is also constructed during the embedding process and used as a key during the extraction, enhancing the model's security.
- The possibility of embedding watermarks containing both positive and negative pixel values into 8-bit, 12-bit, and 16-bit images of various types and formats, including those that may contain negative pixel values.

The paper's remaining sections are arranged as follows. Preliminaries and methods employed in the model are introduced in Section 2. The proposed watermarking model is presented in Section 3. Experimental results are discussed in Section 4, and Section 5 concludes the paper.

2. Background

2.1. Discrete fourier transform

Discrete Fourier transform (DFT) is defined in the discrete signal domain and can express discrete information. Therefore, 2D-DFT is commonly employed in digital image processing to transform an image from the spatial domain to the frequency domain [2,13]. The definition of 2D-DFT for an image with dimensions M × N is expressed by the subsequent equation:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp\left[-j2\pi \left(\frac{xu}{M} + \frac{yv}{N} \right) \right] \quad (1)$$

Where f(x,y) denotes the pixel value at coordinate (x, y) in the spatial domain, F(u,v) denotes the coefficient value at coordinate (u, v) in the frequency domain, and "j" is the imaginary unit.

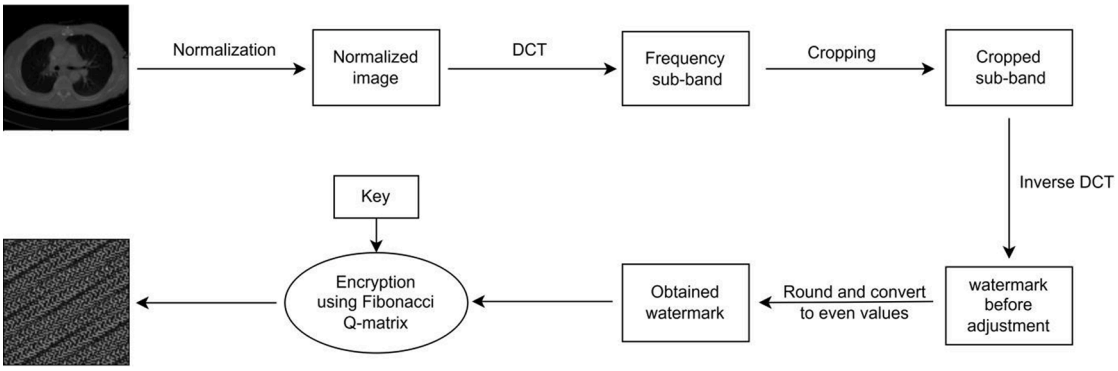


Fig. 1. Block diagram of the watermark generation.

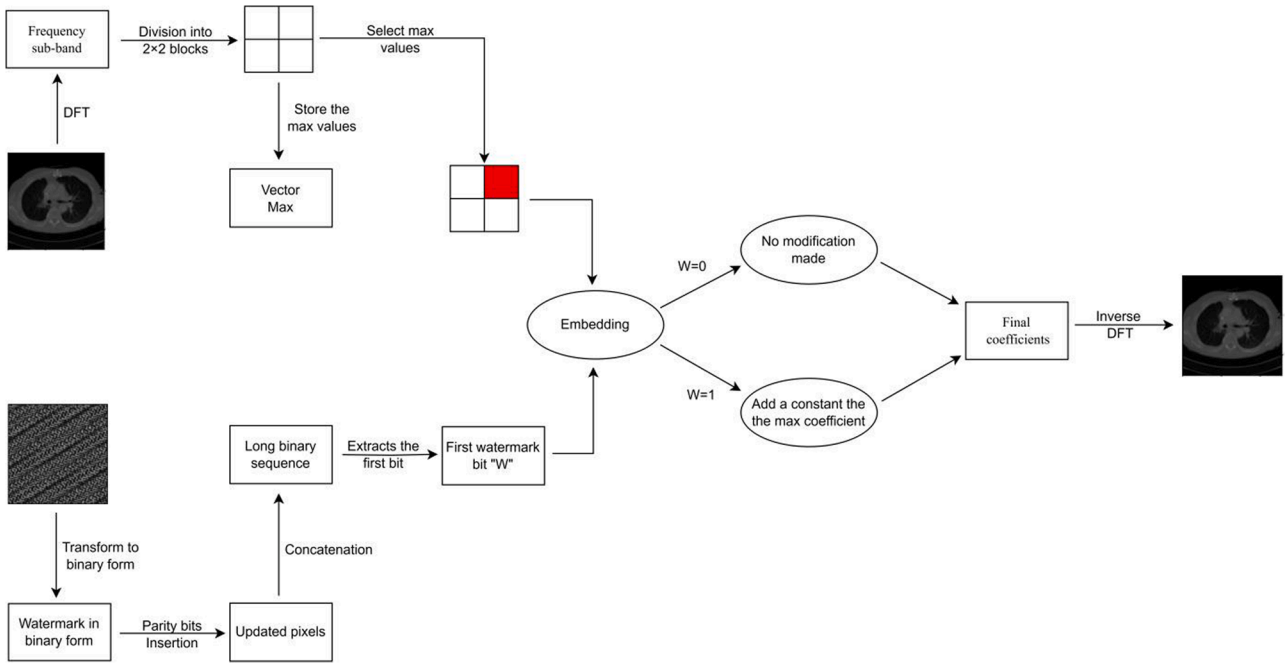


Fig. 2. Block diagram of the embedding process.

The inverse DFT (IDFT) is the process of transforming the frequency domain representation back to the spatial domain. It can be defined as follows:

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)\exp\left[j2\pi\left(\frac{xu}{M}+\frac{yv}{N}\right)\right] \quad (2)$$

2.2. Discrete cosine transform

The Discrete Cosine Transform (DCT) is a mathematical transformation commonly used in image processing technologies like JPEG compression. The 2D DCT is utilized to transform an image of size M × N from the spatial domain to the frequency domain using the following equation [4,25]:

$$F(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{(2x+1)u\pi}{2M}\right]*\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (3)$$

Where f(x,y) denotes the pixel value at coordinate (x, y) in the spatial domain, F(u,v) denotes the coefficient value at coordinate (u, v) in the frequency domain. α(u) and α(v) are calculated as follows [26]:

$$\alpha(u) = \begin{cases} 1/\sqrt{n}, u = 0 \\ 2/\sqrt{n}, u > 0 \end{cases}$$

The 2D Inverse DCT (2D-IDCT) is the inverse process of 2D-DCT, used to transform the frequency domain representation back to the spatial domain. Its formula is defined by [25]:

$$f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)F(u,v)\cos\left[\frac{(2x+1)u\pi}{2M}\right]*\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (4)$$

2.3. Fibonacci Q-matrix

The Fibonacci sequence, also known as the golden section sequence, was introduced by the mathematician Leonardo Fibonacci using rabbit reproduction as an example, hence the term ‘‘Rabbit sequence’’ [27]. The Fibonacci sequence is expressed by:

$$f_n = f_{n-1} + f_{n-2} (n > 1, n \in N^*) \quad (5)$$

Where $f_1 = 1$ and $f_2 = 1$. The one-dimensional Fibonacci sequence can be extended to a two-dimensional Fibonacci Q-matrix which is given by [28]:

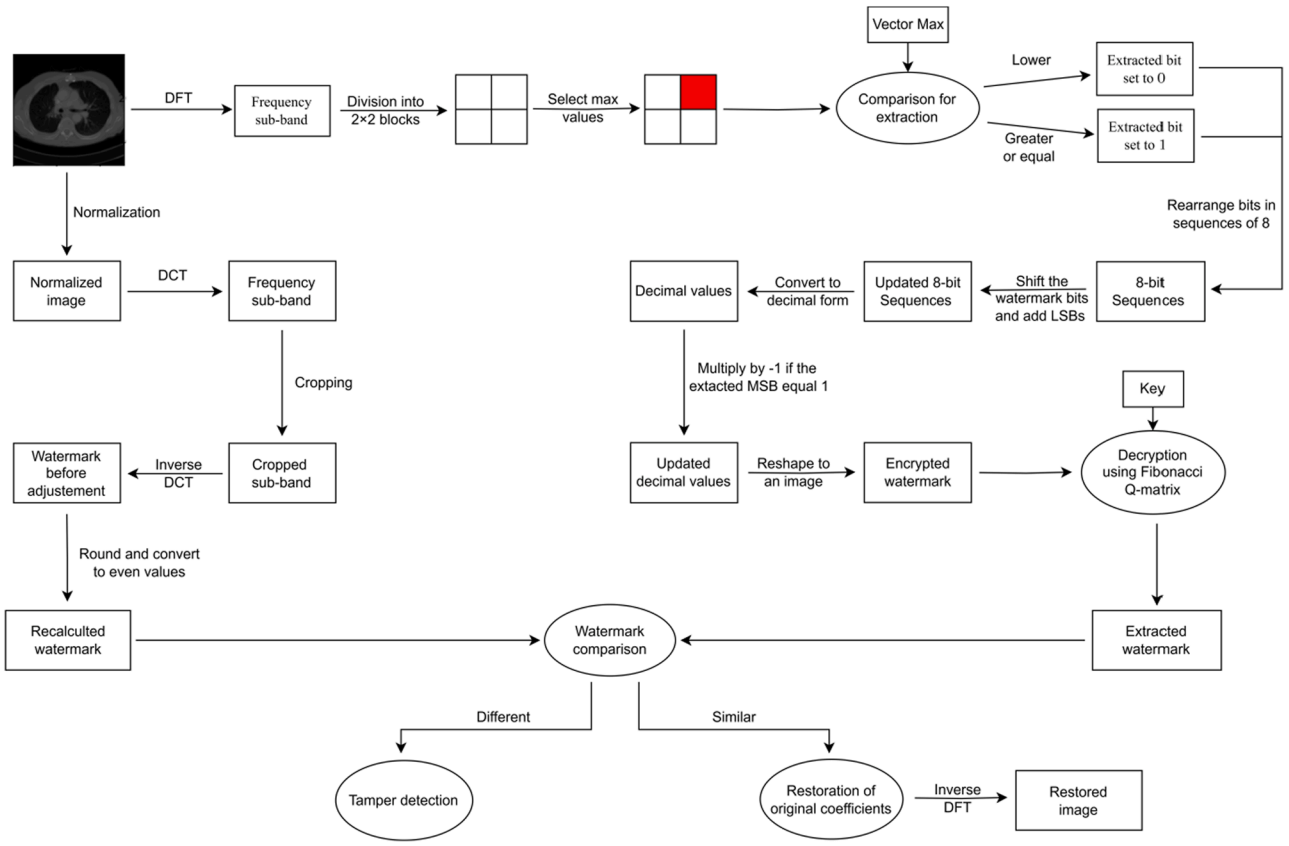


Fig. 3. Block diagram of the extraction process.

Fig. 4. Example of cover images from the different sets.

Fig. 5. Example of watermarked images from the different sets.

Fig. 6. Extracted watermarks from watermarked images of Fig. 5.

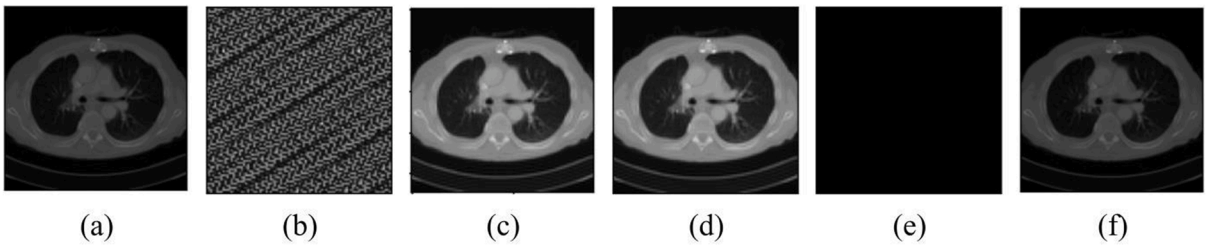


Fig. 7. Example of the extraction process on a watermarked image: (a) watermarked image, (b) extracted watermark, (c) extracted watermark after decryption, (d) recalculated watermark, (e) tamper detection image, (f) restored image.

Table 1
Perceptual quality of the different watermarked images.

Sets	PSNR	MSE	SSIM	Watermark BER
Set 1	140.14	7.26×10^{-8}	$1–2 \times 10^{-9}$	0.01 %
Set 2	139.82	1.8×10^{-7}	$1–2 \times 10^{-9}$	0.01 %
Set 3	137.73	2.9×10^{-7}	$1–3 \times 10^{-9}$	0.02 %
Set 4	117.17	1.33×10^{-7}	$1–5 \times 10^{-9}$	0.001 %
Set 5	119.94	8.33×10^{-8}	$1–4 \times 10^{-9}$	0.001 %

Table 2
The detailed PSNR values of the different watermarked images.

Sets	Minimum	Average	Maximum
Set 1	137.33	140.14	144.47
Set 2	136.53	139.82	142.09
Set 3	135.79	137.73	140.78
Set 4	112.84	117.17	122.31
Set 5	114.64	119.94	126.67

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{6}$$

The n^{th} power of the Fibonacci Q-matrix is calculated as follows [27,28]:

$$Q^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} \tag{7}$$

The inverse matrix of the n^{th} power of the Fibonacci Q-matrix is defined as follows [27,28]:

Table 3
Perceptual quality of the different restored images.

Sets	PSNR	MSE	SSIM
Set 1	327.49	4.93×10^{-26}	$1–3 \times 10^{-14}$
Set 2	324.18	6.85×10^{-26}	$1–5 \times 10^{-14}$
Set 3	325.15	6.05×10^{-26}	1
Set 4	321.05	6.15×10^{-28}	$1–6 \times 10^{-15}$
Set 5	323.1	3.91×10^{-28}	$1–3 \times 10^{-15}$

$$Q^{-n} = \begin{bmatrix} F_{n-1} & -F_n \\ -F_n & F_{n+1} \end{bmatrix} \tag{8}$$

3. Proposed method

In this section, we describe the proposed fragile watermarking model. The presented model involves three main processes: watermark generation, embedding, and extraction for tamper detection.

3.1. Watermark generation

Generating a watermark from the cover image can be more complex compared to using an independent watermark. Nevertheless, the utilization of a self-generated watermark enhances security and makes the watermarking process more blind, eliminating the necessity of storing the watermark on the receiver's side. In our approach, we opted for self-generated watermark derived from the cover image, which is a grayscale image with a bit depth of 8, 12, or 16 bits, and can contain both positive and negative pixel values. The generation process is shown in Fig. 1 and consists of the following steps:

Step 1: If the cover image has a bit depth of 12 or 16 bits and contains only positive pixel values, it undergoes normalization to obtain an image with pixel values in the range of [0,255]. Otherwise, if the cover image has a bit depth superior to 8 and contains both positive and negative pixel values, it gets normalized to the range of [-255,255].

Step 2: Apply DCT to the cover image to obtain a frequency sub-band.

Step 3: Since the low-frequency coefficients (which contain the most important image information) are located on the top-left corner of the sub-band, only the first x lines and x columns are retained, where x is the size of the watermark.

Step 4: Apply the inverse DCT to the cropped sub-band to obtain the watermark.

Step 5: Round the pixel values of the watermark to obtain integer values.

Step 6: Convert the pixel values to even values by adding 1 to negative values and subtracting 1 from positive values. The motivation behind this step will be explained in the embedding process.

Step 7: Scramble the watermark using the Fibonacci Q-matrix method with a secret key "K" to obtain the final watermark, enhancing

Table 4
Comparison of the presented model with other schemes.

Models	Number of samples	Self-embedding	Watermark encryption	Block size	Tamper Localization	Reversibility	Capacity (BPP)	PSNR (dB)
Melendez et al [16]	135	Yes	Yes	32×32	Approximate	Yes	0.4	39.18
Zhang et al [15]	10	No	Yes	2×2	No	Yes	2.25	42.27
Tamal et al [12]	6	Yes	Yes	5×5	Only on ROI	Yes	Not mentioned	51.9
Azizoglu et al [11]	5	Yes	No	4×4	Yes	Yes	0.5625	64.54
Nguyen et al [7]	4	No	No	8×8	Yes	Yes	0.0469	83.54
Nguyen [17]	5	No	No	8×8	Yes	No	0.0156	84
Proposed model	400	Yes	Yes	2×2	No	Yes	0.25	117.17–140.14

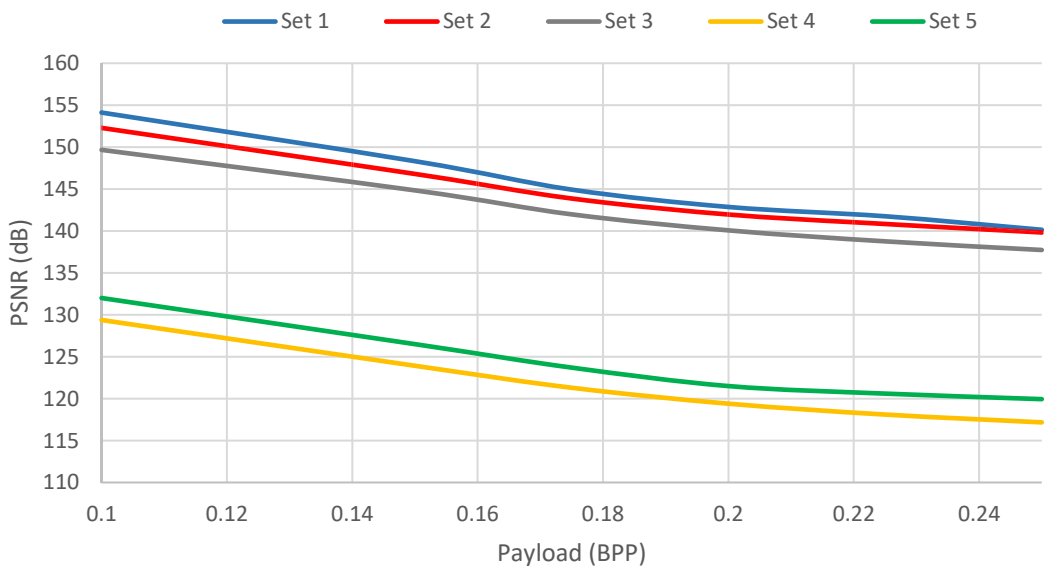


Fig. 8. Effect of embedding capacity on watermarked image quality.

Attacks	Attacked image	Recalculated watermark	Extracted watermark after decryption	Tamper detection image
Rotation (angle=30°)				
Flip direction				
Scaling (size 1024×1024)				
Translation (x = -20, y = -100)				

Fig. 9. Model's performance against geometrical attacks.

the model's security.

This process is required for both the embedding process and the extraction process, as the method is a self-embedding technique.

3.2. Embedding process

During the embedding phase, the watermark is incorporated into the cover image, as illustrated in Fig. 2. The process is carried out through the following steps:

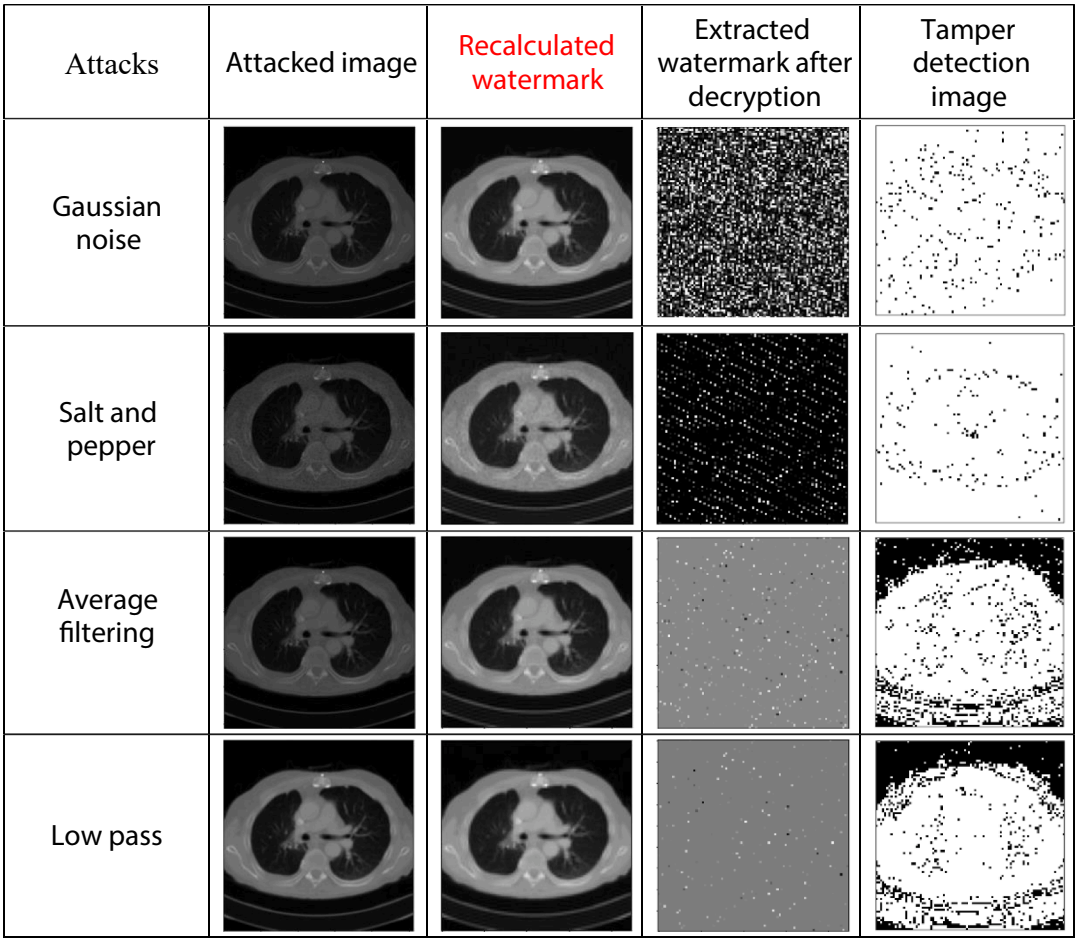


Fig. 10. Model's performance against noising and filtering attacks.

Step 1: The watermark undergoes treatment to handle negative pixel values. For each element, a parity bit "P" is set to 1 if the watermark pixel value is negative; otherwise, it is set to 0.

Step 2: The absolute watermark pixel values are converted to their binary representation. Then, the LSB of each pixel is deleted, and all the bits are shifted one position to the right to make way for the new MSB, where the parity bit "P" is inserted to determine whether the pixel value is positive or negative. This is the reason why the watermark is converted to even values during the generation process, as the LSB is lost during embedding.

Step 3: All the pixel values in binary form are concatenated together from the first line to the last one, and from the first column to the last one, forming a long digit.

Step 4: Perform DFT on the cover image which has a bit depth of 8, 12, or 16 bits to obtain the frequency sub-band.

Step 5: Divide the obtained sub-band into 2×2 non-overlapping blocks.

Step 6: The absolute maximum value of each block is stored in a vector "Max" which serves as a key during the extraction process. Subsequently, add 1 to each element in the vector "Max" to handle the minimal data loss caused by DFT and inverse DFT, ensuring a better watermark extraction.

Step 7: The first bit "w" is extracted from the long binary digit, and embedded into the coefficient with the maximum absolute value in each block.

Step 8: If $w = 0$, no modification is made; otherwise, the coefficient is added or subtracted by 1.5 if it is positive or negative respectively. This value was determined experimentally to ensure that it is small enough to not decrease the watermarked image quality and large enough to resist the DFT and inverse DFT manipulations.

Step 9: Apply the inverse DFT to obtain the watermarked image.

3.3. *Extraction process*

In the extraction phase, the recipient retrieves the watermark from the watermarked image to determine if this latest has been altered or not based on the extracted watermark. If the image is unaltered, the original version can be restored due to the model's reversibility. The extraction process is illustrated in Fig. 3. and involves the following steps:

Step 1: Compute the recalculated watermark based on the watermarked image using the watermark generation procedure, but without scrambling this time.

Step 2: Apply DFT to the watermarked image.

Step 3: Divide the sub-band into 2×2 non-overlapping blocks.

Step 4: Select the absolute maximum value in each block and compare it to its corresponding element in the vector "Max". If this value is lower than the value in the vector "Max", the extracted watermark bit is set to 0; otherwise, it is set to 1.

Step 5: Repeat the process for all blocks, then rearrange the bits in sequences of eight (8 bits per 8 bits) in the same order of extraction (from first to last line and from first to last column).

Step 6: Extract the first bit (MSB) from each eight-bit sequence, shift the seven remaining bits one position to the left, and add a zero to the right to obtain a new eight-bit sequence.

Step 7: Convert each new eight-bit sequence to its decimal form, then multiply it by -1 if its corresponding MSB extracted earlier was 1, to recover the original negative values.

Step 8: Decrypt the extracted watermark using the secret key "K".

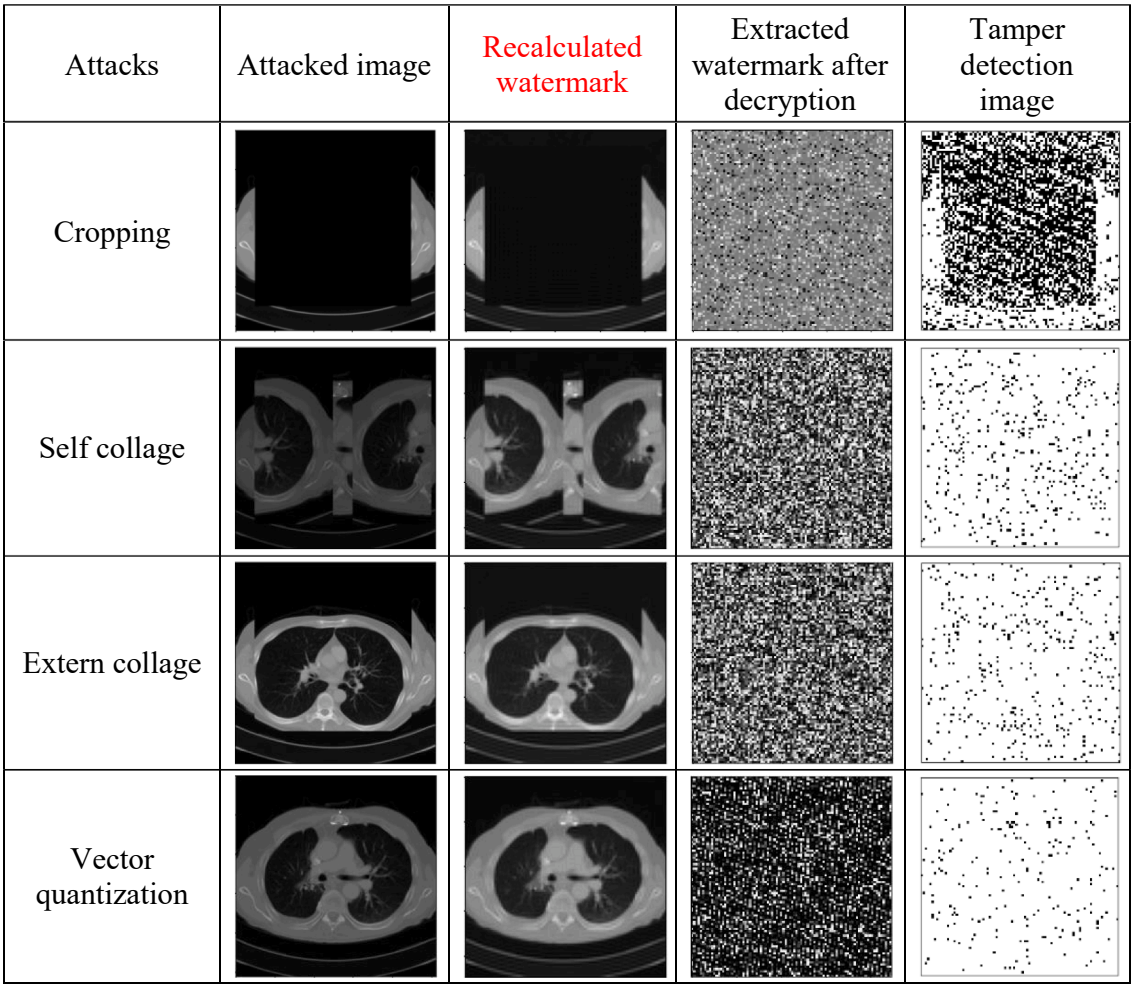


Fig. 11. Model's performance against well-known malicious attacks.

Step 9: Compare the extracted watermark with the recalculated watermark, if the two images are similar, that means the image is unaltered; otherwise, the image has been altered. The comparison is conducted pixel by pixel by transforming the pixel values to their binary form and then comparing them bit by bit since the embedding is done bit by bit. If more than one bit differs between the two pixel values, the watermark pixel is marked as tampered; otherwise, it is considered untampered.

Step 10: If the watermark is tampered with, it signifies that the image has been attacked and the authentication process ends; otherwise, it is possible to restore the original image by subtracting/adding 1.5 to the element with the absolute maximum value in each block, but only when its value is greater than or equal to the corresponding element in the vector "Max".

Step 11: Apply the inverse DFT to obtain the original image.

4. Experimental results

In this section, we present and discuss the experimental results. The presented scheme is implemented using Python 3.11.2 on a 64-bit Windows 10 operating system with 6 GB of RAM and Intel(R) Core (TM) i5-7200U CPU @ 2.50 GHz 2.70 GHz. To evaluate the performances of the proposed technique, it is tested on three different datasets [29,30,31] comprising 400 grayscale medical images with different bit depths (8, 12, 16), different types (CT scans, chest CT scans, MRI), and different formats (DICOM, TIFF, PNG, JPEG). The 400 images were divided into five sets presented as follows:

- Set 1: Comprising 67 DICOM files containing CT scan images with a bit depth of 12 (values ranging between 0 and 4095) [29].
- Set 2: Comprising 33 DICOM files containing CT scan images with a bit depth of 16 (values ranging between −4095 and 4095) [29].
- Set 3: Contains 100 CT scan images in TIFF format with a bit depth of 16 bits (values ranging between −4095 and 4095) [29].
- Set 4: Contains 100 chest CT scans in PNG format with a bit depth of 8 bits (values between 0 and 255) [30].
- Set 5: Contains 100 MRI images in JPEG format with a bit depth of 8 bits [31].

To ensure consistent results and prevent discrepancies solely due to image size differences, all images were re-sized to a uniform size of 512 × 512, as illustrated in Fig. 4 which shows some host images from the different sets used for watermarking.

4.1. Imperceptibility

Watermark invisibility and watermarked image quality are crucial for digital images, especially in telemedicine where the diagnoses are based on digital images. In this section, the quality of both watermarked and restored images are presented.

To assess the quality of watermarked and restored images, we used three image quality metrics: Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). Their formulas are outlined as follows:

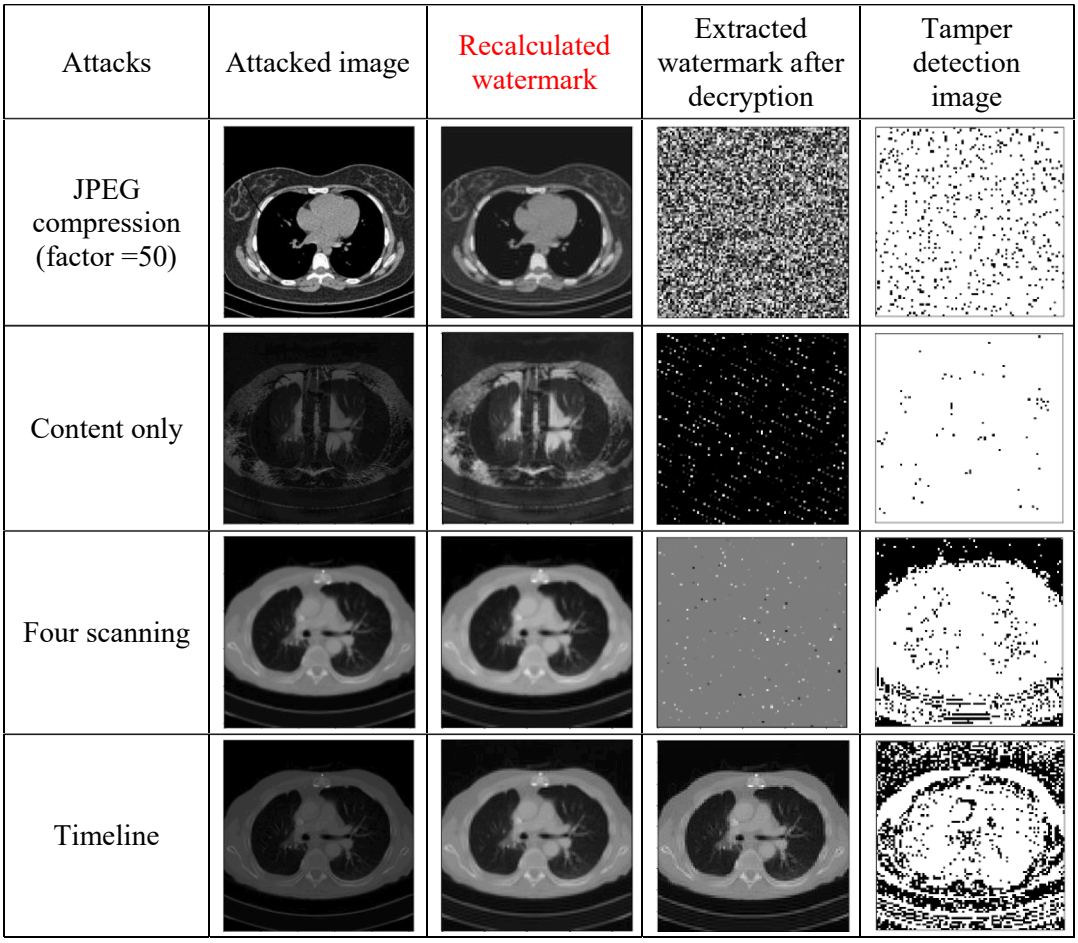


Fig. 12. Model's performance against other attacks.

$$MSE(I, IW) = \frac{1}{N^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(I_{ij} - IW_{ij}\right)^2 \tag{9}$$

$$PSNR(I, IW) = 20*\log_{10}\frac{Max}{\sqrt{MSE}} \tag{10}$$

$$SSIM(I, IW) = \frac{(2u_I u_{IW} + c1)(2cov + c2)}{(u_I^2 + i_{IW}^2 + c1)(o_I^2 + o_{IW}^2 + c2)} \tag{11}$$

Here: I represents the original image, IW the watermarked image, N the image size, o_I and o_{IW} variances of I and IW, cov the covariance of I and IW, c1 and c2 balancing constants, and Max the maximum possible pixel value in the image.

To evaluate also the accuracy of the extraction procedure, we use the Bit Error Rate (BER) metric, which computes the error between the extracted and recalculated watermark bits following the equation:

$$\text{BER} = \frac{\text{Number of incorrectly extracted bits}}{\text{Total number of bits}} \tag{12}$$

Watermarked images produced by our model for the different sets are depicted in Fig. 5, along with their corresponding extracted watermarks in Fig. 6. As observed, there is no noticeable difference to the human eye between the original images (Fig. 4) and their watermarked versions (Fig. 5), indicating minimal distortion introduced during the embedding process. A more in-depth illustration of the watermark extraction process from one of these watermarked images is depicted in Fig. 7. As we can see, the "tamper detection image", obtained after comparing the recalculated and extracted watermarks, is completely black which means that the watermarks are similar. Therefore, the watermarked image hasn't been altered, and the original image can be restored based on it.

Let's now access the quality of watermarked images by computing their PSNR, MSE, and SSIM. As observed in Table.1, the model achieves a great watermarked image quality with an average PSNR value ranging between 117.17 dB and 140.14 dB, an MSE close to zero, indicating minimal distortion introduced during watermark embedding, and an SSIM very close to 1, signifying a high level of structural similarity between the original and watermarked images. Additionally, the average watermark BER is close to zero, indicating a nearly perfect match between the extracted and recalculated watermarks. As mentioned in the extraction phase section, as we tolerate a one-bit difference between the extracted and recalculated watermark pixel values, the watermark bit error rate is completely eliminated, reaching a score of 0 % for the five sets used.

The PSNR disparity between the three first sets and the two remaining sets is due to the nature of the PSNR calculation which uses the maximum possible pixel value in the watermarked image. This method of computation results in higher PSNR values for images with larger bit depths, given that the maximum value of an 8-bit number is 255 while the maximum value for a 12-bit number is 4096. However, if we observe the MSE and SSIM results that don't rely on the maximum possible pixel value, we can notice that the model provides stable performances since the results are similar independently from the set used.

A detailed review of the obtained PSNR values for the five different sets is presented in Table 2. As observed, the model provides stable performances, as the minimum obtained PSNR values remain very high exceeding 112db for a PNG image. Furthermore, sets with the same maximum pixel value (sets 1, 2, 3 with max = 4095, and sets 4, 5 with

max = 255) provide similar PSNR values with an average difference not exceeding 3 dB.

Table 3 presents the quality of the restored images produced by our model. As we can see, the proposed method offers nearly perfect reversibility with an average PSNR value of 324 dB, an MSE near zero, and an SSIM very close to 1, illustrating the nearly perfect match between original and restored images. This reversibility is very useful in sensitive domains like the medical field where we need images with as little distortion as possible.

A comparison with recent fragile watermarking schemes that offer reversibility is presented in Table 4. As observed, the presented model produces higher watermarked image quality compared to the other models, with an average PSNR between 117.17 dB and 140.14 dB across 400 images, while maintaining an acceptable embedding capacity of 0.25 BPP. Furthermore, the presented scheme offers additional features compared to other techniques, like self-embedding and watermark encryption, which reinforces the security of the proposed technique.

It's worth noting that there is a strong inverse correlation between embedding capacity and watermarked image quality. Indeed, the more information is embedded inside an image, the more distortion is introduced during embedding which decreases the quality of the watermarked images. To illustrate this, we tested the proposed model with different embedding capacities ranging from 0.1 BPP to 0.25 BPP, on the five sets containing 400 images. As observed in Fig. 8, an increased embedding capacity leads to a degradation in watermarked image quality. For example, if we increase the embedding capacity from 0.1 BPP to 0.25 BPP in Set 3, the average PSNR of watermarked images drops from 149.67 dB to 137.73 dB.

As we aim to achieve high watermarked image quality while maintaining an acceptable embedding capacity, we opted for the maximum embedding capacity the model can reach, which is 0.25 BPP, especially since the model can produce high watermarked image quality superior to 117.17 dB even with a capacity of 0.25 BPP.

4.2. Performance against attacks

As previously mentioned, medical images play an important role in people's health, as even a minor alteration can significantly impact decision-making. Therefore, recognizing whether an image has been subjected to any attack or modification is a significant step for medical images.

To assess the performance of the presented technique against attacks, we subjected the watermarked image to 16 different attacks. The model is first tested against geometrical attacks (rotation, flip direction, scaling, and translation), and filtering attacks (Gaussian noise, salt and pepper, average filtering, and low pass) as shown in Fig. 9 and Fig. 10, respectively. The results clearly demonstrate the sensitivity of the presented scheme to these attacks, as the extracted watermark after decryption shows significant alterations. Consequently, the "tamper detection image" contains numerous white points, illustrating that the watermarks are not similar.

The wide alteration of the extracted watermark is due to the high sensitivity of the presented model to attacks, and also to the fact that any modifications on the extracted watermark before decryption can cause the decryption process to fail, leading to results that significantly differ from the original image, and enhancing the model's sensitivity to attacks.

To further evaluate the performances of the presented model, we test it against well-known malicious attacks including cropping [32], self collage [33], extern collage [33], and vector quantization [34,35] attacks, as shown in Fig. 11. The extracted watermark is widely affected by these attack illustrating the fragility of the model.

Finally, the presented technique is tested against other attacks that may not be detected by other models such as (i) four scanning attack [35,36]. that consists of resizing the image to a size N2 × M2 and then resizing it back to its original size involving some data loss, (ii) content only attack [35] that can cause problems for models that perform embedding in the spatial domain, since the attack replaces only the MSBs of the image pixels by another image MSBs while keeping the other bits unchanged, and (iii) timeline attack [37] which consists of adding the same small noise value to all the pixels over a period of time, which keep the high-frequency coefficients unchanged. This attack may cause problems for models performing embedding only in the high-frequency sub-band. The model is also tested against compression attacks like JPEG. As illustrated in Fig. 12, the presented scheme is sensitive to all these attacks as the extracted watermark after decryption shows significant alterations.

The watermarked image used during these tests is obtained after watermarking an image from the first set which contains DICOM files. The results against attacks are similar regardless of the set used, so we can choose another image from a different set without affecting the results. Except for JPEG compression which can't support a bit depth higher than 8. This is why we used another image from the fourth set for the JPEG compression attack.

Overall, experimental results affirm that the presented scheme produces a high watermarked image quality and provides nearly perfect reversibility. Additionally, the proposed method achieves satisfactory execution times of 1.31 s and 1.16 s for the embedding and extraction processes, respectively. This includes watermark generation, image reading and saving, watermarks comparison, and all the necessary processes in each procedure. The high sensitivity to diverse attacks further underscores the effectiveness and security of the presented approach, making it a promising solution for image authentication in sensitive domains like the medical industry.

5. Conclusion

This paper proposes a reversible fragile watermarking technique to ensure the authenticity of medical image. The watermark is first derived from the cover image using DCT, then encrypted utilizing Fibonacci Q matrix to obtain the final watermark. Afterward, DFT is performed on the cover image, and the obtained sub-band is segmented into 2 × 2 blocks. The watermark is then embedded into these blocks with a capacity of one bit per block. In the extraction phase, the recipient retrieves the watermark from the watermarked image using the secret key "Max" and decrypts the watermark using the second key "K". Based on the extracted watermark, the receiver determines if the image has been altered or not. If it is unaltered, the original image is reconstructed from the watermarked version due to the model's reversibility.

Experimental results affirm that the presented method produces a great watermarked image quality with an average PSNR between 117.17 dB and 140.14 dB while maintaining an acceptable embedding capacity of 0.25 BPP. Additionally, the model provides nearly perfect reversibility with a PSNR of 324 dB and a high sensitivity to various attacks, as it was tested against 16 different attacks and could detect tampering in all of them. In future works, we will focus on enhancing the proposed method by integrating a tamper localization process. This process is designed to accurately localize tampered areas when the watermarked image is subjected to attacks. Additionally, we will incorporate a recovery process intended to restore the localized tampered areas, resulting in a recovered image without any tampering despite the encountered attacks.

CRediT authorship contribution statement

Riadh Bouarroudj: Conceptualization, Investigation, Methodology, Software, Visualization, Writing – original draft. **Feryel Souami:** Project administration, Supervision. **Fatma Zohra Bellala:** Conceptualization, Validation, Writing – review & editing. **Nabil Zerrouki:** Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Y. Gangadhar, V.S. Giridhar Akula, P. Chenna Reddy, An evolutionary programming approach for securing medical images using watermarking scheme in invariant discrete wavelet transformation, Biomed. Signal Process. Control 43 (2018) 31–40, https://doi.org/10.1016/j.bspc.2018.02.007.

[2] H. Cao, F. Hu, Y. Sun, S. Chen, Q. Su, Robust and reversible color image watermarking based on DFT in the spatial domain, Optik 262 (2022) 169319, https://doi.org/10.1016/j.ijleo.2022.169319.

[3] M. Cedillo-Hernandez, A. Cedillo-Hernandez, M. Nakano-Miyatake, H. Perez-Meana, Improving the management of medical imaging by using robust and secure dual watermarking, Biomed. Signal Process. Control 56 (2020) 101695, https://doi.org/10.1016/j.bspc.2019.101695.

[4] H. Wang, Z. Yuan, S. Chen, Q. Su, Embedding color watermark image to color host image based on 2D-DCT, Optik 274 (2023) 170585, https://doi.org/10.1016/j.ijleo.2023.170585.

[5] T. Huang, X.u. Jia, T.u. Shixin, B. Han, Robust zero-watermarking scheme based on a depthwise overparameterized VGG network in healthcare information security, Biomed. Signal Process. Control 81 (2023) 104478, https://doi.org/10.1016/j.bspc.2022.104478.

[6] R. Xiang, G. Liu, K. Li, J. Liu, Z. Zhang, M. Dang, Zero-watermark scheme for medical image protection based on style feature and ResNet, Biomed. Signal Process. Control 86 (Part A) (2023) 105127, https://doi.org/10.1016/j.bspc.2023.105127.

[7] T.S. Nguyen, C.C. Chang, X.Q. Yang, A reversible image authentication scheme based on fragile watermarking in discrete wavelet transform domain, AEU-Int. J. Electron. C 70 (8) (2016) 1055–1061, https://doi.org/10.1016/j.aeue.2016.05.003.

[8] E. Gul, S. Ozturk, A novel hash function based fragile watermarking method for image integrity, Multimed. Tools Appl. 78 (2019) 17701–17718, https://doi.org/10.1007/s11042-018-7084-0.

[9] G.-D. Su, C.-C. Chang, C.-C. Lin, Effective self-recovery and tampering localization fragile watermarking for medical images, IEEE Access 8 (2020) 160840–160857, https://doi.org/10.1109/ACCESS.2020.3019832.

[10] R. Bhardwaj, An enhanced reversible patient data hiding algorithm for E-healthcare, Biomed. Signal Process. Control 64 (2021) 102276, https://doi.org/10.1016/j.bspc.2020.102276.

[11] G. Azizoğlu, A. N. Toprak, A novel reversible fragile watermarking in DWT domain for tamper localization and digital image authentication, 9th International Symposium on Digital Forensics and Security (ISDFS), Elazig, Turkey, 2021, pp. 1-6. https://doi.org/10.1109/ISDFS52919.2021.9486339.

[12] T.A. Tamal, C. Saha, M.F. Hossain, S. Rahman, "Integer Wavelet Transform Based Medical Image Watermarking for Tamper Detection, International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6. https://doi.org/10.1109/ECACE.2019.8679152.

[13] F. Nejati, H. Sajedi, A. Zohourian, Fragile watermarking based on QR decomposition and fourier transform, Wirel. Pers. Commun. 122 (2022) 211–227, https://doi.org/10.1007/s11277-021-08895-1.

[14] R. Bouarroudj, F. Souami, F.Z. Bellala, Fragile watermarking for medical image authentication based on DCT technique, 2023 5th International Conference on Pattern Analysis and Intelligent Systems (PAIS), Sétif, Algeria 2023 1–6. https://doi.org/10.1109/PAIS60821.2023.10322029.

[15] H.a. Zhang, S. Sun, F. Meng, A high-capacity and reversible patient data hiding scheme for telemedicine, Biomed. Signal Process. Control 76 (2022) 103706, https://doi.org/10.1016/j.bspc.2022.103706.

[16] G. Melendez-Melendez, R. Cumplido, Reversible image authentication scheme with blind content reconstruction based on compressed sensing, Eng. Sci. Technol. Int. J. 34 (2022) 101080, https://doi.org/10.1016/j.jestch.2021.101080.

[17] T.S. Nguyen, Fragile watermarking for image authentication based on DWT-SVD-DCT techniques, Multimed. Tools Appl. 80 (2021) 25107–25119, https://doi.org/10.1007/s11042-021-10879-z.

[18] X. Xiong, S. Zhong, Y. Lu, Separable and reversible data hiding scheme for medical images using modified Logistic and interpolation, Biomed. Signal Process. Control 87 (Part B) (2024) 105521, https://doi.org/10.1016/j.bspc.2023.105521.

[19] C. Li, A. Zhang, Z. Liu, L. Liao, D. Huang, Semi-fragile self-recoverable watermarking algorithm based on wavelet group quantization and double authentication, Multimed. Tools Appl. 74 (2015) 10581–10604, https://doi.org/10.1007/s11042-014-2188-7.

[20] T. Tuncer, M. Kaya, A novel image watermarking method based on center symmetric local binary pattern with minimum distortion, Optik 185 (2019) 972–984, https://doi.org/10.1016/j.ijleo.2019.04.038.

[21] P. Lefèvre, P. Carré, C. Fontaine, P. Gaborit, J. Huang, Efficient image tampering localization using semi-fragile watermarking and error control codes, Signal Process. 190 (2022) 108342, https://doi.org/10.1016/j.sigpro.2021.108342.

[22] L. Rakhmawati, W. Wirawan, S. Suwadi, A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability, J. Image Video Proc. 2019 (2019) 61, https://doi.org/10.1186/s13640-019-0462-3.

[23] A. Anand, A.K. Singh, Watermarking techniques for medical data authentication: a survey, Multimed. Tools Appl. 80 (2021) 30165–30197, https://doi.org/10.1007/s11042-020-08801-0.

[24] R. Thabit, Review of medical image authentication techniques and their recent trends, Multimed. Tools Appl. 80 (2021) 13439–13473, https://doi.org/10.1007/s11042-020-10421-7.

[25] K. Fares, A. Khaldi, K. Redouane, E. Salah, DCT & DWT based watermarking scheme for medical information security, Biomed. Signal Process. Control 66 (2021) 102403, https://doi.org/10.1016/j.bspc.2020.102403.

[26] Z. Li, H. Zhang, X. Liu, C. Wang, X. Wang, Blind and safety-enhanced dual watermarking algorithm with chaotic system encryption based on RHFM and DWT-DCT, Digital Signal Process. 115 (2021) 103062, https://doi.org/10.1016/j.dsp.2021.103062.

[27] Z. Liang, Q. Qin, C. Zhou, An image encryption algorithm based on Fibonacci Q-matrix and genetic algorithm, Neural Comput. & Applic. 34 (2022) 19313–19341, https://doi.org/10.1007/s00521-022-07493-x.

[28] K.M. Hosny, S.T. Kamal, M.M. Darwish, G.A. Papakostas, New image encryption algorithm using hyperchaotic system and Fibonacci Q-matrix, Electronics 10 (9) (2021) 1066, https://doi.org/10.3390/electronics10091066.

[29] K Scott Mader, CT Medical Images Dataset, kaggle, 2017. https://www.kaggle.com/datasets/kmader/siim-medical-images.

[30] Mohamed Hany, Chest CT-Scan images Dataset, Kaggle, 2020. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.

[31] Masoud Nickparvar, Brain Tumor MRI Dataset, Kaggle (2021), https://doi.org/10.34740/kaggle/dsv/2645886.

[32] C.S. Hsu, S.F. Tu, Enhancing the robustness of image watermarking against cropping attacks with dual watermarks, Multimed. Tools Appl. 79 (2020) 11297–11323, https://doi.org/10.1007/s11042-019-08367-6.

[33] H.C. Wu, W.L. Fan, C.S. Tsai, J.J.C. Ying, An image authentication and recovery system based on discrete wavelet transform and convolutional neural networks, Multimed. Tools Appl. 81 (2022) 19351–19375, https://doi.org/10.1007/s11042-021-11018-4.

[34] M. Holliman, N. Memon, Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes, IEEE Trans. Image Process. 9 (3) (2000) 432–441, https://doi.org/10.1109/83.826780.

[35] N.R.N. Raj, R. Shreelekshm, A survey on fragile watermarking based image authentication schemes, Multimed. Tools Appl. 80 (2021) 19307–19333, https://doi.org/10.1007/s11042-021-10664-y.

[36] C.C. Chang, Y.H. Fan, W.L. Tai, Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery, Pattern Recogn. 41 (2) (2008) 654–661, https://doi.org/10.1016/j.patcog.2007.06.003.

[37] S.H. Lee, S.G. Kwon, B.J. Jang, J.H. Kim, K.R. Kwon, D.K. Kim, E.J. Lee, 3D Animation Watermarking Scheme using Orientation Interpolator, 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2007, pp. 1223–1226. https://doi.org/10.1109/ICME.2007.4284877.