

# A Prediction Error Nonlinear Difference Expansion Reversible Watermarking for Integrity and Authenticity of DICOM Medical Images

David Muigai<sup>1\*</sup>, Elijah Mwangi<sup>2</sup>, Edwell T. Mharakurwa<sup>3</sup>

Department of Electrical and Electronics Engineering, Dedan Kimathi University of Technology, Nyeri, Kenya<sup>1,3</sup>  
Faculty of Engineering, University of Nairobi, Kenya<sup>2</sup>

**Abstract**—It is paramount to ensure the integrity and authenticity of medical images in telemedicine. This paper proposes an imperceptible and reversible Medical Image Watermarking (MIW) scheme based on image segmentation, image prediction and nonlinear difference expansion for integrity and authenticity of medical images and detection of both intentional and unintentional manipulations. The metadata from the Digital Imaging and Communications in Medicine (DICOM) file constitutes the authentication watermark while the integrity watermark is computed from Secure Hash Algorithm (SHA)-256. The two watermarks are combined and compressed using the Lempel Ziv (LZ) -77 algorithm. The scheme takes advantage of the large smooth areas prevalent in medical images. It predicts the smooth regions with zero error or values close to zero error, while non-smooth areas are predicted with large error values. The binary watermark is encoded and extracted in the zero-prediction error using a nonlinear difference expansion. The binary watermark is concentrated more on the Region of non-interest (RONI) than the Region of interest (ROI) to ensure a high visual quality while maintaining a high capacity. The paper also presents a separate low degradation side information processing algorithm to handle overflow. Experimental results show that the scheme is reversible and has a remarkable imperceptibility and capacity that are comparable to current works reported in literature.

**Keywords**—Medical Image Watermarking (MIW); Digital Imaging and Communication in Medicine (DICOM); region of interest (ROI) and region of non-interest (RONI); prediction error (PE); nonlinear difference expansion (NDE); authenticity; integrity

## I. INTRODUCTION

Medical images and patient data are often shared in e-diagnosis over open communication channels. The transmission of such data is prone to intentional and unintentional manipulations, affecting confidentiality, integrity and authenticity. Such manipulations can result in misdiagnosis and even lead to loss of life hence the need to ensure reliability [1-2].

Medical images and patient information are transmitted, stored, retrieved, printed, processed and displayed through Digital Imaging and Communications in Medicine (DICOM) standards [3]. In DICOM, metadata which is the patient report and information that connects to the image ensures reliability of medical images data. The metadata is saved in the image file header [3]. This technique is insecure as the metadata can be

easily modified, destroyed, or disconnected from the medical image [4].

Digital image watermarking, a branch of information hiding technology where a secret message is hidden in public data, can overcome these challenges. The secret message can be the metadata, a hospital logo, an electronic signature, or any other identifier in medical images. The requirements for medical image watermarking are reversibility, imperceptibility and reliability [2].

Digital image watermarking is classified into several classes based on the method of embedding the secret message, reversibility, application and region(s) used to encode the secret message [5]. It can be either frequency or spatial domain based on the method of encoding the secret message. Spatial domain techniques [4], [6] changes the pixel intensities of the image directly to implant the secret message. Frequency domain techniques [7-8] implants the secret message by changing the coefficient values of the transformed image. Digital image watermarking techniques are classified as reversible, semi-reversible or irreversible based on reversibility. In reversible techniques [9-10], the original image and the secret message are losslessly restored, while in irreversible techniques [11], the secret message and the original image cannot be losslessly restored. Semi-reversible techniques [12] restore some regions of the original image while others cannot be restored. Therefore, reversible techniques are preferred for watermarking medical images. Depending on the application of digital image watermarking, the schemes are classified as either fragile or robust. Robust watermarking techniques emphasize the robustness of the encoded message. The encoded watermark can resist legitimate and illegitimate attacks during image transmission in robust schemes. Therefore, robust watermarking strategies [13-14] are mainly used for copyright protection of images. Fragile watermarking methods emphasize on detection of manipulations during image transmission. Hence, fragile watermarking techniques [4], [6] are used to confirm the integrity of images. The watermarking can be further classified as ROI or RONI based. The anatomical details are contained in ROI whereas RONI carries the uninformative background usually black in color [15]. In ROI techniques [4],[16] the secret message is hidden in ROI while RONI techniques [17-18] hides it in RONI.

In this paper, a novel reversible MIW scheme based on prediction error nonlinear difference expansion for authenticity

\*Corresponding Author.

and integrity of medical images is proposed. The scheme has the following objectives:

1) Predicting the medical image with zero error or values close to zero for smooth regions while non-smooth regions are predicted with large error values. Smooth regions are characterized by zero or slight differences in adjacent pixel intensities. Hiding the watermark in these regions is less visible.

2) Concentrating the watermark mainly in RONI to ensure excellent visual quality on the ROI while maintaining its security.

3) Determining an optimum point for trade-off between capacity and imperceptibility for prediction error nonlinear difference expansion watermarking.

4) Attaining good visual quality of watermarked images that supersede the benchmark value and the perceptual boundary.

5) Providing a low degradation approach to handle the overflow.

6) Lossless recovery of the medical image without the need of a location map.

The rest of this paper is organized as follows; The second section provides medical image watermarking schemes reported in the literature. The third section describes the proposed work. The fourth section presents the results and discussion. The fifth section, which is the final section, presents the conclusion and suggestion for further work.

## II. RELATED WORK

Researchers have recently presented Medical Image Watermarking (MIW) techniques. This section analyzes some of these techniques. Roček *et al.* [12] presented a new MIW strategy that merges RONI watermarking method with zero-watermarking principle and reversibility features. The scheme uses a reversible watermarking in the RONI, which achieves a high capacity and implants data using the zero-watermarking principle. A Dual Tree Complex Wavelet Transform (DT-CWT) is used to merge these techniques. The limitation of the approach is the need for a location map at extraction to recover the image and the encoded watermark.

Gao *et al.* [19] presented a reversible MIW approach that achieves tamper detection and enhances ROI contrast. It utilizes Otsu's thresholding method to differentiate the RONI from the ROI. The scheme expands the peak-pairs of ROI histogram to achieve data encoding alongside a less distortion contrast enhancement. This approach creates a feature bit matrix from ROI and encodes it in the least significant bits of RONI to guarantee ROI reversibility. The limitations of the scheme are; (i) the scheme is semi-reversible as it can restore only the ROI at reception and (ii) the need for implanting the feature bit matrix.

Atta-ur-Rahman *et al.* [20] presented a reversible MIW approach for the integrity of medical images and the secrecy of patient data. The watermark is created chaotically and encoded using a chaotic key in selected pixels. The selected pixels are divided using a primitive polynomial of degree four and the

remainder appended to the secret message. At the reception, the computed remainder validates the watermark. In this approach, a high imperceptibility was exhibited. The approach's limitations were; the hiding capacity was not measured and the method is not region-based hence making it impossible to select hiding regions.

Liu *et al.* [21] presented a novel robust reversible MIW to protect the integrity and authenticity of medical images. This method addresses the challenge of losing information in watermark embedding due to image segmentation. It avoids biases during diagnosis by designing a recursive dither modulation (RDM) based watermarking. RDM is later combined with Singular Value Decomposition (SVD) and Slantlet Transform (ST) to protect image authenticity. The RONI and ROI are divided to generate the watermark encoded into the whole image, thus avoiding risk related to image segmentation.

Swaraja *et al.* [22] presented a MIW technique that conceals a dual watermark on RONI blocks for authenticity and tamper recognition in medical images. This procedure uses the lossless Lempel-Ziv-Welch compression algorithm to compress the dual watermark, thus increasing capacity. The embedding blocks are chosen based on the human visual systems characteristics, integrating Discrete Wavelet Transform (DWT) and Schur transform alongside Particle Swarm Bacterial Foraging Optimization Algorithm (PSBFO). The scheme is robust against signal attacks and compression and shows transparency from the simulation results.

Fares *et al.* [23] proposed a MIW approach based on Discrete Cosine Transform (DCT) and DWT for protecting patient data. The scheme proposes two approaches. The first approach combines DCT and Schur Decomposition (SD) and performs integration in medium frequencies thus achieving a good compromise between visual quality and robustness. The second approach combines SD and DWT to achieve a robust watermark distribution. The proposed schemes maintain good visual quality and are robust against attacks. The capacity of the first approach is 682 bits which correspond to 85 characters only. Therefore, the capacity of the first approach is limited. The second approach conceals 1024 bits equivalent to 128 characters. Therefore, the capacity of the second approach is certainly reduced.

## III. METHODOLOGY

A fragile Medical Image Watermarking (MIW) scheme is presented for detecting both intentional and unintentional manipulations and ensuring the authenticity and integrity of DICOM medical images. The approach also presents a separate low degradation side information processing algorithm to handle overflow. The sub-sections are as follows.

### A. Watermark Creation and Compression

There exist several ways for creating an authentication watermark [24]. The DICOM files consist of image data and metadata in a single *.dcm* file. The metadata contains patient information, image dimensions, parameters of modality acquisition and operator identification [24]. The scheme uses the metadata of DICOM images as the authentication watermark. It also employs the Secure Hash Algorithm (SHA)

-256 to compute the integrity watermark. This is a patent cryptographic hash function used in data integrity and digital certificates [25]. The output is a 64-digit hexadecimal number and is strong, and easy to compute [25]. The method detects manipulations by comparing the hidden and extracted integrity watermark. It also combines the authentication and integrity watermark to form the total watermark. The text string is compressed to a binary string using the Lempel Ziv -77 compression algorithm. Table I shows a summary of the watermark creation and compression data features.

TABLE I. WATERMARK CREATION AND COMPRESSION DATA FEATURE

Type of Watermark	Creation	Minimum size in bits	Maximum size in bits
Integrity	SHA-256	500	550
Authenticity	Metadata	22000	25000
Total Watermark	Authenticity + Integrity	22500	25550

### B. Image Prediction

Any image pixel is predictable using an expression that constitutes its neighboring pixels [26]. The hiding capacity of prediction-error expansion depends on how close the predicted image resembles the original image. Medical images are characterized by a large smooth area, unlike other images and hiding the watermark in these areas is less distinguishable by the human visual system [4]. The scheme predicts the smooth areas of the medical image with zero error or values close to zero and non-smooth areas with large error values using (1). The predicted pixel is  $p_{ij}$ .

$$p_{ij} = \begin{cases} \frac{p_{i-1j}p_{ij-1}}{p_{i-1j-1}} & 0 \notin (p_{i-1j}, p_{ij-1}, p_{i-1j-1}) \\ p_{i-1j} + p_{ij-1} - p_{i-1j-1} & \text{otherwise} \end{cases} \quad (1)$$

### C. Image Segmentation

There exist several techniques to segment medical images into ROI and RONI. The techniques can be manual, like use of polygons and freehand sketching, or automatic. A radiographer divides a medical image using a technique of his/her choice. The scheme automatically segments the medical image using a thresholding technique that utilizes the mean of pixels and morphological operations. The specific medical image segmentation procedure is as follows.

- 1) Load the medical image  $M_I$ .
- 2) Compute the mean of all pixels and consider it as the initial threshold  $T_0$ .
- 3) Divide the pixels into two groups such that pixels greater than  $T_0$  form the ROI, otherwise RONI.
- 4) Compute the mean of ROI  $M_{ROI}$  and RONI  $M_{RONI}$  separately.
- 5) Compute the new threshold  $T_1$  as the average of  $M_{ROI}$  and  $M_{RONI}$ .
- 6) Repeat steps (iv) and (v) until the new threshold converges.
- 7) Convert the medical image into a binary image using the last threshold by making all pixels less than it black otherwise white.

- 8) Perform morphological filtering on the binary image obtained from step (vii).
- 9) Perform region filling on the binary image from step (viii).
- 10) Display the binary image.
- 11) Obtain the indices  $M_{I0}$  and  $M_{I1}$  corresponding to the pixels with values 0 and 1 of the binary image, respectively.
- 12) The RONI and ROI of the medical image correspond to  $M_{I0}$  and  $M_{I1}$  respectively.

### D. Prediction Error Nonlinear Difference Expansion

The basic prediction error difference expansion first proposed by [27] is given as follows;

Let  $p_{ij}$  be the pixels of the original image and  $\hat{p}_{ij}$  be the predicted pixel. The prediction error is computed as;

$$e_{ij} = p_{ij} - \hat{p}_{ij} \quad (2)$$

Let  $b_i$  be a binary watermark. The watermark bits are embedded by expanding the prediction error as;

$$e'_{ij} = 2e_{ij} + b_i \text{ where } b_i = 0 \text{ or } 1 \quad (3)$$

Let  $T > 0$ , be the threshold to increase capacity and control degradation. It is directly proportional to degradation and capacity but inversely proportional to visual quality. The threshold  $T$  can be varied from 1 to the maximum possible gray intensity value of an image. The watermarked pixels are given by (4).

$$p'_{ij} = \hat{p}_{ij} + e'_{ij} \text{ if } e_{ij} < T \quad (4)$$

If  $|e_{ij}| \geq T$ , the pixels cannot carry a watermark bit and are shifted to provide a greater prediction error than the carrier pixels at detection using (5).

$$p'_{ij} = \begin{cases} p_{ij} + T & \text{if } e_{ij} \geq T \\ p_{ij} - (T - 1) & \text{if } e_{ij} < -T \end{cases} \quad (5)$$

The embedding threshold and the predicted image are transmitted as side information alongside the watermarked image to the receiver [26]. The problem of overflow and underflow is re-solved by creating a location map or using flag bits [26-28]. At detection, if the same predicted value for the original image is available, then the error is computed as in (6).  $\bar{p}_{ij}$  is the received image.

$$\bar{e}_{ij} = \bar{p}_{ij} - \hat{p}_{ij} \quad (6)$$

The prediction error differentiates the embedded and shifted pixels. If  $-2T \leq \bar{e}_{ij} \leq 2T + 1$ , then it is a carrier pixel and the error is computed as  $\bar{e}_{ij} = 2e_{ij} + b$ , where  $b$  is the least significant bit of  $\bar{e}_{ij}$ . The original and shifted pixels are recovered as in (7) and (8) respectively.

$$p_{ij} = \frac{\bar{p}_{ij} + \hat{p}_{ij} - b}{2} \quad (7)$$

$$p_{ij} = \begin{cases} \bar{p}_{ij} - T \\ \bar{p}_{ij} + (T - 1) \end{cases} \quad (8)$$

To overcome the challenge of high degradation caused by increasing  $T$  while increasing capacity, the transmission of the predicted image and  $T$  as side information, overflow and underflow in the basic prediction error linear difference expansion, the scheme uses a prediction error nonlinear difference expansion. It uses systematic multiple predictions and expansions of the zero - prediction error. The binary watermark is divided into sections equal to the number of expansions. The original image is predicted using (1). The error is computed using (2). The carrier and non-carrier pixels are embedded and shifted as in (9) and (10) respectively.

$$p'_{ij} = p_{ij} + b_i \text{ if } e_{ij} = 0 \text{ where } b_i = 0 \text{ or } 1 \quad (9)$$

$$p'_{ij} = \begin{cases} p_{ij} + 1 & e_{ij} \geq 1 \\ p_{ij} & e_{ij} \leq -1 \end{cases} \quad (10)$$

The first row and column are not used in embedding. The rest of the rows and columns are embedded in intervals of two pixels, with the initial pixels being  $P_{22}$ ,  $P_{23}$ ,  $P_{32}$  and  $P_{33}$  until all the other pixels are used in embedding the watermark. The image is predicted in each initial pixel embedding stage. This is the prediction error-linear difference expansion (PE-LDE) with a threshold  $T=1$ . To increase capacity with low distortion, unlike the basic prediction error difference expansion which increases  $T$ , the scheme uses a prediction error quadratic difference expansion (PE-QDE) considered as a cascade of the PE-LDE described as follows:

The image  $p'_{ij}$  is used again to perform PE-QDE and is predicted using (1) to obtain  $\hat{p}'_{ij}$ . The error is computed as;

$$e'_{ij} = p'_{ij} - \hat{p}'_{ij} \quad (11)$$

Let  $b'_i$  be the watermark to be embedded in PE-QDE. The carrier and non-carrier pixel are embedded and shifted as in (12) and (13) respectively.

$$p''_{ij} = p'_{ij} + b'_i \text{ if } e'_{ij} = 0 \text{ where } b'_i = 0 \text{ or } 1 \quad (12)$$

$$p''_{ij} = \begin{cases} p'_{ij} + 1 & e'_{ij} \geq 1 \\ p'_{ij} & e'_{ij} \leq -1 \end{cases} \quad (13)$$

The embedding in PE-QDE follows the same order as in PE-LDE. The process is repeated so as to increase the capacity at low distortion until saturation is reached.

The watermarked image is predicted at the reception using the same prediction technique used in embedding. During extraction, at any instance of image prediction that corresponds to an image prediction during embedding, the same predicted image is obtained. This guarantees reversibility. The error is computed as in (14).  $\check{p}_{ij}$  is the obtained predicted image.

$$\check{e}_{ij} = p''_{ij} - \check{p}_{ij} \quad (14)$$

The error obtained is the same as the expanded error in the last embedding stage of initial pixel  $P_{33}$  in PE-QDE. If the  $\check{e}_{ij}$  is either 0 or 1 then it is a carrier pixel, else a non-carrier pixel. For carrier pixels, the watermark is extracted as follows.

$$b = \check{e}_{ij} \quad (15)$$

The carrier and non-carrier pixels are recovered as;

$$p'_{ij} = \begin{cases} p''_{ij} - b & \text{if } \check{e}_{ij} \text{ is } 0 \text{ or } 1 \\ p''_{ij} - 1 & \text{if } \check{e}_{ij} > 1 \\ p''_{ij} & \text{otherwise} \end{cases} \quad (16)$$

The first row and column remain unaltered as it is not used in embedding. The rest of the rows and columns are extracted in intervals of two pixels with the initial pixels being  $P_{33}$ ,  $P_{32}$ ,  $P_{23}$  and  $P_{22}$ , an inverse order to that of embedding. This is the inverse PE-QDE. The inverse PE-LDE is as follows;

The image  $p'_{ij}$  is predicted using the same technique and the error is computed as in (17).  $\check{p}_{ij}$  is the obtained predicted image.

$$\check{e}_{ij} = p'_{ij} - \check{p}_{ij} \quad (17)$$

If  $\check{e}_{ij}$  is either 0 or 1 then it is a carrier pixel else a non-carrier pixel. For carrier pixels, the watermark is extracted as;

$$b = \check{e}_{ij} \quad (18)$$

The original carrier and non-carrier pixels are recovered as;

$$p_{ij} = \begin{cases} p'_{ij} - b & \text{if } \check{e}_{ij} \text{ is } 0 \text{ or } 1 \\ p'_{ij} - 1 & \text{if } \check{e}_{ij} > 1 \\ p'_{ij} & \text{otherwise} \end{cases} \quad (19)$$

The inverse PE-LDE follows the same order as inverse PE-QDE. Therefore, the binary watermark and the original image are restored.

### E. Watermark Encoding

Fig. 1 shows the flow chart of the watermark embedding and the specific procedure is as follows:

- 1) Read the DICOM file.
- 2) Segment the image data using the segmentation procedure to obtain the binary mask.
- 3) Obtain the authentication watermark (AW) from the DICOM metadata.
- 4) Compute SHA-256 on the image data to get the integrity watermark (IW).
- 5) Concatenate the authentication and integrity watermark to form the total watermark.
- 6) Compress the total watermark using LZ-77 compression algorithm.
- 7) Divide the binary watermark into sections equal to the knee point less one.
- 8) Embed the first and second section of the binary watermark using the PE-LDE and PE-QDE respectively as described in section D to obtain the partial watermarked image.
- 9) Using the binary mask obtained in step (ii), segment the partial watermarked image such that the RONI and ROI region of the partial watermarked image corresponds to '0' and '1' of the binary watermark respectively.
- 10) Embed the third, fourth and fifth sections of the binary watermark using prediction error - third order difference

expansion (PE-TODE), prediction error -fourth order difference expansion (PE-FODE) and prediction error - fifth order difference expansion (PE-FIODE) respectively on RONI only. Consider these expansions as cascades of PE-LDE.

11) Obtain the watermarked image which may exhibit overflow.

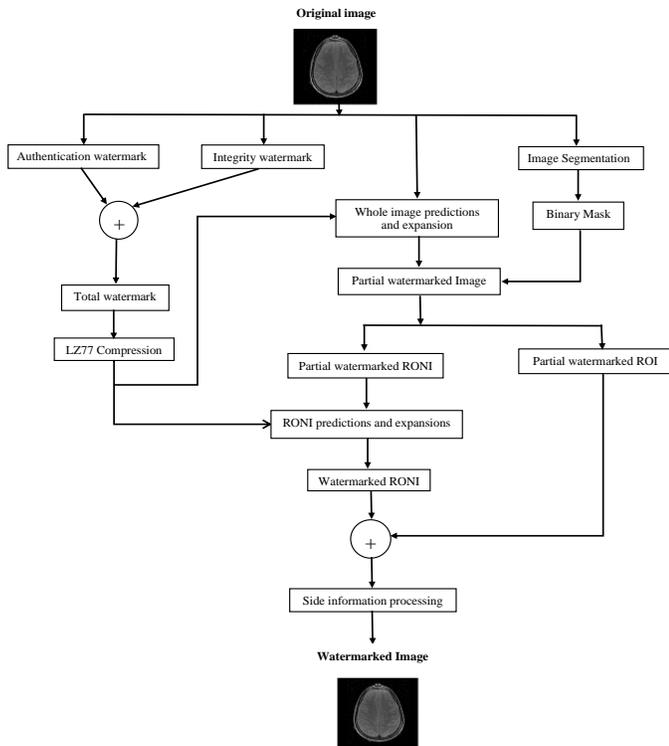


Fig. 1. Flowchart of Watermark Embedding.

#### F. Side Information Processing

Overflow and underflow problems in reversible watermarking are essential as it can lead to irreversibility or heavy distortion. Overflow occurs when the maximum gray level is exceeded, while underflow occurs when the minimum gray level is exceeded. The scheme does not exhibit underflow as pixel intensity are not decreased during processing. However, the approach can exhibit overflow. The maximum overflow is equals to the number of expansions used during embedding. The scheme preserves the last prediction error expansion to hide the side information. The scheme considers the length of the watermark, segmentation threshold, and locations of maximum gray level and overflowed pixels as side information. The procedure for Side Information Processing (SIP) can be described as follows:

- 1) Scan the watermark image and record the locations of maximum and overflowed gray levels.
- 2) Modify the maximum and overflowed gray pixels by subtracting the number of expansions used in embedding.
- 3) Concatenate the locations, segmentation threshold and length of the watermark and consider it as side information.
- 4) Compress the side information.

5) Hide the side information on the whole image using the last prediction error difference expansion.

6) Finally, obtain the watermarked image that carries the watermark and side information without exhibiting overflow for transmission.

#### G. Side Information Recovery

The side information is first recovered at extraction and the maximum and overflowed pixels are restored before extracting the watermark. The side information is recovered using the following steps.

- 1) Perform the sixth inverse prediction error difference expansion on the whole image to obtain the side information binary watermark.
- 2) Decompress the side information binary watermark.
- 3) Modify the locations obtained in step (ii) by adding the number of expansions used in embedding to restore the overflowed watermark image.

#### H. Watermark Extraction

Fig. 2 shows a flow chart of the watermark extraction and the specific procedure is as follows:

- 1) Using the side information recovery procedure, recover the side information and restore the maximum and overflowed pixels.
- 2) Separate the RONI and ROI of the watermarked image using the segmentation procedure. Use the recovered segmentation threshold
- 3) Extract the fifth, fourth and third sections of the binary watermark using inverse PE-FIODE, PE-FODE and PE-TODE respectively on the RONI only to recover the partial watermarked image
- 4) Extract the second and first sections of the binary watermark using inverse PE-QDE and PE-LDE respectively on the whole image to recover the original image
- 5) Concatenate the recovered sections of the binary watermark and decompress using LZ-77 decompression algorithm
- 6) Compare the recovered and hidden authenticity watermark for authentication verification
- 7) Compute the SHA-256 of the recovered image and compare it with the hidden integrity watermark to verify that the image has been transmitted without manipulation.

#### I. Performance Measures

The proposed scheme is evaluated in terms of capacity, imperceptibility, reversibility and robustness to provide a fair comparison with other relevant schemes. The capacity of the proposed scheme in bits per pixel (bpp) is computed using (20). The capacity of the scheme depends on the accurate prediction of the image. To increase capacity, systematic multiple predictions and expansions are employed.

$$capacity = \frac{Number\ of\ pixels\ for\ embedding}{size\ of\ original\ image} \quad (20)$$

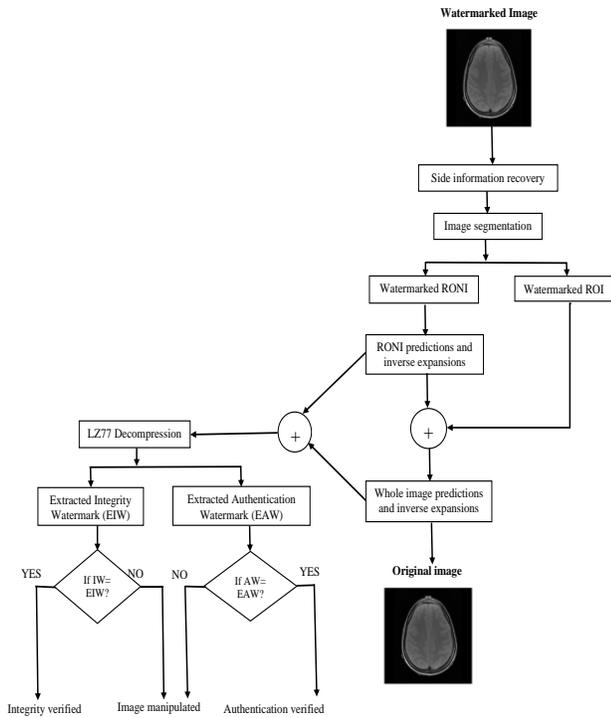


Fig. 2. Flowchart of Watermark Extraction and Verification.

The imperceptibility of the scheme is evaluated using PSNR, SSIM and Image Fidelity (IF) between the original and watermarked image. The PSNR, SSIM and IF are computed using (21), (22) and (23) respectively. The  $I$  and  $I'$  of the equations represent the original and watermarked image respectively and  $(i, j)$  are the coordinates of pixels in these images. The images are of dimensions  $M \times N$ .

$$PSNR(I, I') = 10 \times \log_{10} \frac{HP^2 \cdot M \cdot N}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I(i, j) - I'(i, j))^2} \quad (21)$$

Where  $HP$  is the highest possible pixel value for the images. PSNR values range between 0 and  $+\infty$ . A higher PSNR value shows low image distortion and high visual quality.

$$SSIM(I, I') = \frac{(2\mu_I \mu_{I'} + C_1)(2\sigma_{II'} + C_2)}{(\mu_I^2 + \mu_{I'}^2 + C_1)(\sigma_I^2 + \sigma_{I'}^2 + C_2)} \quad (22)$$

Where  $\mu_I, \mu_{I'}$  are the averages of  $I, I'$  respectively,  $\sigma_I, \sigma_{I'}$  are variances of  $I, I'$  respectively,  $C_1, C_2$  are balancing constants and  $2\sigma_{II'}$  are the covariance for  $I, I'$  respectively. The SSIM is a quality measure based on Human Visual System to measure image distortion in structural information. Its values lie between 0 and 1. An SSIM of 1 indicates complete similarity.

$$IF = 1 - \frac{\sum_{i=1}^{N-1} \sum_{j=0}^{M-1} (I(i, j) - I'(i, j))}{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I(i, j))^2} \quad (23)$$

The IF parameter measures similarity between two images. An IF value of 1 between two images indicates that they are similar.

At extraction, the proposed scheme evaluates the reversibility of the image and the watermark. In evaluating the reversibility of the image, PSNR and Root Mean Square Error (RMSE) between the extracted and original image are used to verify that two images are 100% numerically identical. The RMSE is computed using (24).

$$RMSE(I, I') = \sqrt{\frac{\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I(i, j) - I'(i, j))^2}{M \cdot N}} \quad (24)$$

A PSNR and RMSE of  $+\infty$  and 0 respectively indicate that the image has been recovered without any loss else otherwise. In evaluating the watermark reversibility, Accuracy Ratio (AR) and Bit Error Rate (BER) between the embedded and extracted binary string is used. The BER and AR are computed as;

$$BER = \frac{\text{Number of error bits}}{\text{Total number of bits}} \quad (25)$$

$$AR = \frac{\text{Number of correct bits}}{\text{Total number of bits}} \quad (26)$$

A BER and AR value of 0 and 1 respectively indicate that the watermark has been recovered without any loss. The scheme also computes the BER between the original and extracted binary watermark as in (25) to evaluate robustness against attacks. A BER value closer to 0 indicate stronger robustness.

#### IV. RESULTS AND DISCUSSION

##### A. Experimental Setup

The experimental results were obtained using a PC with an Intel CPU of 2.6 GHz and 8GB RAM. The scheme was implemented in MATLAB 2021a to test the reversibility, imperceptibility, robustness and capacity. A set of 270 medical images in DICOM format comprising of 30 brain Magnetic Resonance Images (MRI) images, 30 cervix MRI images, 30 kidney Computed Tomography (CT) images, 30 lung CT images, 30 chest Digital Radiography (DX) images, 30 breast Mammography (MG) images, 30 liver Ultrasound (US) images, 30 chest Computed Radiology (CR) images and 30 headneck Positron emission tomography (PT) images were obtained from [29-30]. All images were 16 bpp and were re-sized to 512 x 512.

##### B. Imperceptibility

The imperceptibility of the scheme was evaluated using PSNR, SSIM and IF and the results given in Table II. The minimum PSNR value between the original and watermarked images is 83.0 dB, which is above the acceptable benchmark value of 40 dB [5] and the perceptual boundary of 82 dB [4] for the human visual system. Qasim *et al.* [4] conducted a relative Visual Grading Analysis (VGA) trial and determined that the modification of images to 82 dB or higher is unnoticeable to all observers. Therefore, the original and watermarked image are visually indistinguishable by the human eye. The SSIM and IF are either unity or close to unity indicating that the watermarked was hidden invisibly. A small set of the medical images and their corresponding watermarked images are shown in Fig. 3.

TABLE II. EVALUATION OF THE WATERMARKED IMAGES

Body Part Examined (Modality)	No of images used	Capacity (bpp)		PSNR		SSIM	IF	
		Minimum	Maximum	Minimum	Maximum		Minimum	Maximum
1. Brain (MRI)	30	0.47	0.93	84.5	85.5	1.00	0.93	0.99
2. Cervix (MRI)	30	0.98	1.21	83.8	85.0	1.00	0.96	0.98
3. Kidney (CT)	30	0.44	0.76	83.8	84.4	1.00	0.97	0.99
4. Lung (CT)	30	0.43	0.45	83.3	84.5	1.00	1.00	1.00
5. Chest (DX)	30	0.41	0.52	83.2	85.9	1.00	1.00	1.00
6. Breast (MG)	30	0.55	1.00	83.6	86.0	1.00	1.00	1.00
7. Liver (US)	30	0.62	0.85	83.8	85.6	1.00	0.95	0.99
8. Chest (CR)	30	0.41	0.53	83.8	86.2	1.00	1.00	1.00
9. Headneck(PT)	30	0.66	0.87	83.0	84.3	1.00	1.00	1.00
<b>Overall Performance</b>	<b>270</b>	<b>0.41</b>	<b>1.21</b>	<b>83.0</b>	<b>86.2</b>	<b>1.00</b>	<b>0.93</b>	<b>1.00</b>

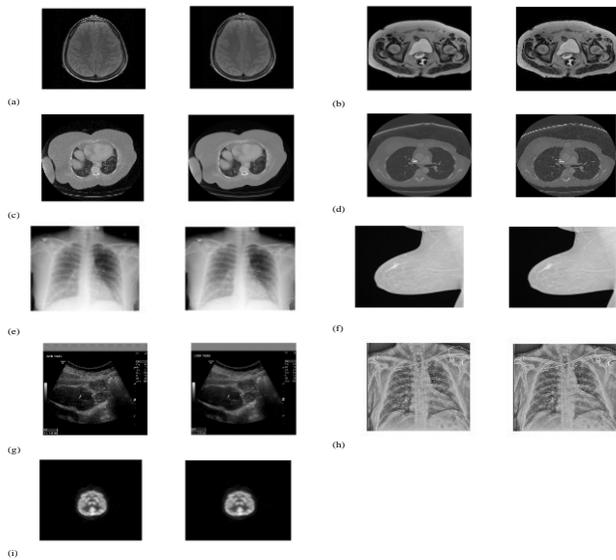


Fig. 3. Sample Medical Images and their Corresponding Watermarked Images. (a) MRI Brain, (b) MRI Cervix, (c) CT Kidney, (d) CT Lung, (e) DX Chest, (f) MG Breast, (g) US Liver, (h) CR Chest, (i) PT Headneck, [32-33].

### C. Reversibility

The reversibility of the scheme was assessed at reception for both the extracted image and the watermark. The BER and AR values between the embedded and extracted binary watermark were used to evaluate the watermark reversibility. The values of zero and one respectively were obtained, demonstrating that the watermark was extracted without loss. This is a confirmation of the integrity and authenticity of the watermark. At the reception, the medical image is restored for diagnosis. The PSNR, RMSE, SSIM and IF between the original and extracted image were used to evaluate the image reversibility. The PSNR and RMSE values were positive infinity and zero respectively indicating that the two images are 100% numerically identical. The SSIM and IF were both unity demonstrating that the extracted image is identical to the original image.

### D. Capacity

The binary watermark is encoded in the zero error pixels. To increase capacity, the number of iterations is increased. For the first and second iteration, that is the PE-LDE and PE-QDE, the binary watermark is encoded in the whole image, while for iterations above two, the binary watermark is encoded in the RONI only. A binary image obtained from the image segmentation procedure is used to distinguish the ROI from the RONI. This controls degradation in ROI as increasing capacity distorts the watermarked image. Fig. 4 shows the capacity of the scheme versus the number of iterations.

In Fig. 4, the capacity of the scheme increases steadily with an increase in the number of iterations until the knee point. After the knee point, increasing the number of iterations results in a low increase in capacity but still degrades the image due to the shifting of non-carrier pixels. The complexity of the approach and consequently the computational time increases with an increase in the number of iterations. After 30 iterations, an increase in the number of iterations increases capacity by less than 0.001bpp. This is the saturation point of the scheme. The scheme limits the number of iterations to 6 which correspond to the knee point. The capacity created in the 6th iteration hides the side information and is performed on both RONI and ROI. Fig. 5 shows PSNR as a function of capacity.

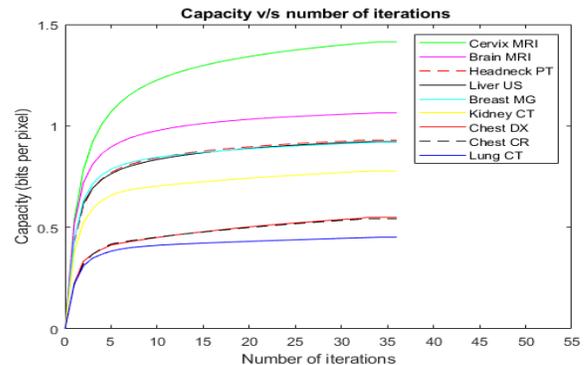


Fig. 4. Capacity v/s Number of Iterations.

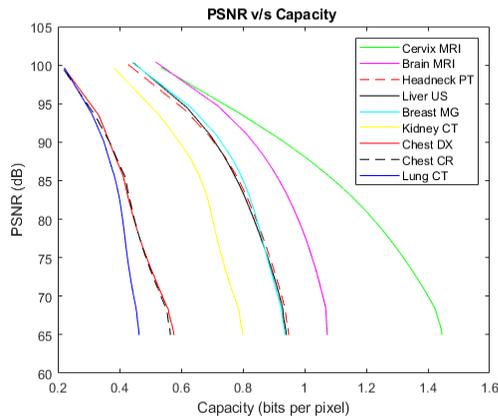


Fig. 5. PSNR v/s Capacity.

From Fig. 5, it can be noted that an increase in capacity distorts the watermarked image. An increase in capacity degrades the image until saturation. After saturation, increasing the number of iterations results in a negligible increase in capacity but continues to degrade the image due to the shifting of non-carrier pixels. Table III shows the specific number of bits in Kbytes hidden and the resultant PSNR for the images in Fig. 3. The cervix MRI image had the highest capacity of 36.57 Kbytes, while the lung CT had the lowest capacity of 12.87 Kbytes. Fig. 6 shows a sample binary image of MRI image (a) shown in Fig. 3.

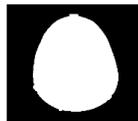


Fig. 6. Binary Image of MRI Image in Fig. 3 (a).

**E. Robustness**

The robustness of the scheme was assessed using BER between the encoded and extracted binary watermark under various attacks. The average PSNR and BER of the scheme under various intentional and unintentional attacks are shown

in Table IV. In this table, adding or removing a region illustrate intentional attacks while the rest demonstrate unintentional attacks. The integrity and authenticity of the medical images are confirmed when the encoded and extracted watermark are completely identical. However, intentional and unintentional manipulations on the medical image result in a mismatch between the encoded and extracted watermark. The latter are malicious manipulations where a region is added or removed while the former are accidental manipulations occurring during transmission. The reversibility of the medical image under various attacks was also assessed using PSNR. The PSNR obtained for each image under attack was not positive infinity, indicating that the original medical image was not recovered. The BER obtained under various attacks for each image was non-zero, demonstrating a mismatch between the encoded and extracted watermark. Therefore, the authenticity and integrity of the medical images under attack is not confirmed. This shows that the approach is fragile to manipulations.

**F. Comparison with Relevant Schemes**

The developed scheme has been compared with other relevant MIW schemes as shown in Table V and Table VI.

The approaches are compared in terms of location map, reversibility, capacity and visual quality. The developed approach uses a non-linear difference expansion that does not need a location map at the extraction for recovery of the watermark and the image to guarantee reversibility thus is better than the approach [12]. The side information of the developed approach is hidden in the last iteration leading to a low degradation approach to handle the overflow. The developed scheme performs better than the scheme reported in [19] in terms of reversibility as it recovers the whole image while the former recovers only the ROI. The medical image is restored in its pristine state in the developed approach. Additionally, it gives a higher performance than schemes [19] and [22-23] in terms of capacity as it achieves a higher capacity at a lower degradation. Also, the developed scheme achieved the highest PSNR of 83.0 to 86.2 dB and a SSIM of unity compared to the scheme [12] and [19-23], demonstrating a better visual quality.

TABLE III. SPECIFIC NUMBER OF BITS HIDDEN IN EACH ITERATION AND THE RESULTANT PSNR FOR THE MEDICAL IMAGES SHOWN IN FIG. 3. ‘C’ REPRESENTS CAPACITY IN K BYTES

Body part examined	PE-LDE		PE-QDE		PE-TODE		PE-FODE		PE-FIODE		S I P		Total Bits Hidden
	C (kB)	PSNR (dB)	C (kB)	PSNR (dB)	C (kB)	PSNR (dB)	C (kB)	PSNR (dB)	C (kB)	PSNR (dB)	C (kB)	PSNR (dB)	
1. Brain (MRI)	16.96	100.3	6.67	94.6	2.96	91.0	1.70	88.4	1.08	86.4	0.79	<b>84.8</b>	<b>30.16</b>
2. Cervix (MRI)	17.53	99.6	8.20	93.8	4.47	90.3	2.86	87.8	2.00	85.9	1.51	<b>84.3</b>	<b>36.57</b>
3. Kidney (CT)	12.49	99.6	4.72	93.6	2.20	90.1	1.35	87.6	0.80	85.0	0.53	<b>84.0</b>	<b>22.09</b>
4. Lung (CT)	7.25	99.6	2.92	93.6	1.25	90.1	0.63	87.6	0.50	85.6	0.32	<b>84.0</b>	<b>12.87</b>
5. Chest (DX)	7.34	99.4	3.58	93.4	1.15	89.9	0.80	87.4	0.66	85.5	0.29	<b>83.9</b>	<b>13.82</b>
6. Breast (MG)	14.43	100.3	6.28	94.4	2.72	90.8	1.44	88.2	0.92	86.2	0.61	<b>84.5</b>	<b>26.40</b>
7. Liver (US)	14.57	100.3	5.68	94.5	2.47	90.9	1.41	88.3	0.91	86.3	0.69	<b>84.7</b>	<b>25.73</b>
8. Chest (CR)	7.19	99.4	3.34	93.4	1.55	89.9	0.70	87.4	0.73	85.4	0.26	<b>83.9</b>	<b>13.77</b>
9. Headneck (PT)	14.04	100.0	5.91	94.2	2.73	90.7	1.56	88.2	1.01	86.2	0.71	<b>84.5</b>	<b>25.96</b>

TABLE IV. AVERAGE PSNR AND BER OF THE PROPOSED SCHEME UNDER ATTACKS

Attack type	Parameter	PSNR	BER
Resizing	0.8	-	0.47
Rotation and cropping	-3°	51.9	0.46
Brightness adjustment	-	12.7	0.56
Histogram equalization	-	9.3	0.52
Gaussian noise	Mean=0.003 Variance=0.001	31.2	0.49
Salt and pepper noise	Density=0.09	19.1	0.58
JPEG compression	Quality factor=80	48.9	0.41
Median filter	Window= 5 x 5	68.7	0.48
Average filter	Window= 3 x 3	68.6	0.43
Adding a region	-	57.2	0.42
Removing a region	-	57.1	0.44

TABLE V. PERFORMANCE COMPARISON OF THE PROPOSED SCHEME WITH OTHER RECENT RELEVANT SCHEMES. RCSP STANDS FOR RESIDUE WITH CHAOTICALLY SELECTED PIXELS

Approach Year	Medical Image Segmentation	Hiding Region	Hiding Technique	Location map	Reversible	Capacity (bpp)	Visual Quality	
							PSNR	SSIM
Roček et al. [12]	Automatic	ROI RONI	DT-CWT LSB	YES	YES	ROI Dependent	Average = 81	Average = 1
Gao et al. [19]	Automatic	ROI RONI	HS LSB	NO	YES ONLY ROI	0.08 to 0.35	24.5 to 30.5	0.92 to 0.98
Atta-ur-Rehman et al. [20]	No segmentation	Whole Image	RCSP	NO	YES	---	Average = 71.1	---
Liu et al. [21]	Automatic	Whole Image	ST SVD	NO	YES	---	Average = 41.3	Average = 0.96
Swaraja et al. [22]	Automatic	ROI RONI	DWT PSBFO	NO	YES	0.76 to 1.00	Average = 34.5	---
Fares et al. [23]	No segmentation	Whole Image	DCT- Schur	NO	YES	6.50 x 10-4	Average = 48.0	Average = 1.
			DWT- Schur	NO	YES	9.77 x 10-4	Average = 49.2	Average = 1
<b>Proposed Approach</b>	<b>Automatic</b>	<b>ROI RONI</b>	<b>PE NDE</b>	<b>NO</b>	<b>YES</b>	<b>0.41 to 1.21</b>	<b>83.0 to 86.2</b>	<b>Average = 1</b>

TABLE VI. FURTHER COMPARISON OF THE PROPOSED SCHEME TO OTHER RELEVANT SCHEMES

Description	Jun [28]	Jaiswal [31]	Kang [32]	Proposed Scheme
Capacity (bpp)	0.25	0.48	1.00	<b>0.70</b>
PSNR	58.8489	49.4970	51.6190	<b>84.34</b>

The developed scheme is further compared to a linear difference expansion scheme [28], a prediction-error linear difference expansion scheme [31] and a fragile SVD with grouped block-based scheme [32]. The average performance in terms of PSNR and Capacity in bpp for the medical images shown in Fig. 3 are used for comparison. The scheme has a better performance than schemes in [28] and [31] in terms of capacity as it has a higher payload. The scheme has a better performance than the scheme [28] and [31-32] in terms of visual quality as it has the highest PSNR value.

## V. CONCLUSION

An imperceptible and reversible watermarking scheme based on prediction error and non-linear difference expansion to ensure integrity, authenticity and detect manipulations on

DICOM medical images has been proposed. The experimental results obtained demonstrate that the approach is reversible and provides remarkable visual quality and capacity. The scheme yields PSNR values which are above the benchmark value and the perceptual boundary, demonstrating that it is imperceptible. The approach is fragile to manipulations making it suitable for detecting them. The approach also yields superior performance in terms of visual quality and compares favorably in terms of capacity to schemes available in the literature. It also yields superior results compared to other linear difference expansion-based schemes, demonstrating that the nonlinear difference expansion is superior to linear difference expansion in capacity and visual quality. The future work will involve developing a watermarking system that will not only detect manipulation but also restore the tampered regions.

## REFERENCES

- [1] P. Aparna and P. V. Kishore, "A Blind Medical Image Watermarking for secure E-healthcare application using crypto-watermarking system," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1558–1575, 2019, doi:10.1515/jisys-2018-0370.
- [2] S. M. Mousavi, A. Naghsh, and S. A. Abu-Bakar, "Watermarking techniques used in medical images: A survey," *Journal of Digital Imaging*, vol. 27, no. 6, pp. 714–729, 2014, doi: 10.1007/s10278-014-9700-5.

- [3] M. Larobina and L. Murino, "Medical Image File Formats," *Journal of Digital Imaging*, vol. 27, no. 2, pp. 200–206, 2013, doi: 10.1007/s10278-013-9657-9.
- [4] A. F. Qasim, R. Aspin, F. Meziane, and P. Hogg, "ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16433–16463, 2018, doi: 10.1007/s11042-018-7029-7.
- [5] N. A. Memon and A. Alzahrani, "Prediction-based reversible watermarking of CT scan images for content authentication and copyright protection," *IEEE Access*, vol. 8, pp. 75448–75462, 2020, doi: 10.1109/ACCESS.2020.2989175.
- [6] S-C Liew and J. M. Zain, "Reversible medical image watermarking for tamper detection and recovery," 2010 3rd International Conference on Computer Science and Information Technology, Chengdu, China, 2010, doi: 10.1109/ICCSIT.2010.5564078.
- [7] J. Liu, J. Huang, Y. Luo, L. Cao, S. Yang, D. Wei, and R. Zhou, "An optimized image watermarking method based on HD and SVD in DWT domain," *IEEE Access*, vol. 7, pp. 80849–80860, 2019, doi: 10.1109/ACCESS.2019.2915596.
- [8] O. M. Al-Qershi and B. E. Khoo, "Authentication and data hiding using a hybrid ROI-based watermarking scheme for DICOM images," *Journal of Digital Imaging*, vol. 24, no. 1, pp. 114–125, 2009, doi: 10.1007/s10278-009-9253-1.
- [9] B. Lei, E.-L. Tan, S. Chen, D. Ni, T. Wang, and H. Lei, "Reversible watermarking scheme for medical image based on differential evolution," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3178–3188, 2014, doi:10.1016/j.eswa.2013.11.019.
- [10] T.-S. Nguyen, C.-C. Chang, and N.-T. Huynh, "A novel reversible data hiding scheme based on difference-histogram modification and optimal EMD algorithm," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 389–397, 2015.
- [11] J. H. Wu, R.-F. Chang, C.-J. Chen, C.-L. Wang, T.-H. Kuo, W. K. Moon, and D.-R. Chen, "Tamper detection and recovery for medical images using near-lossless information hiding technique," *Journal of Digital Imaging*, vol. 21, no. 1, pp. 59–76, 2007, doi: 10.1007/s10278-007-9011-1.
- [12] A. Roček, K. Slavíček, O. Dostál and M. Javorník, "A new approach to fully-reversible watermarking in medical imaging with breakthrough visibility parameters", *Biomedical Signal Processing and Control*, vol. 29, pp. 44-52, 2016, doi: 10.1016/j.bspc.2016.05.005.
- [13] C. Gong, J. Li, U. A. Bhatti, M. Gong, J. Ma, and M. Huang, "Robust and secure zero-watermarking algorithm for medical images based on Harris-surf-DCT and chaotic map," *Security and Communication Networks*, vol. 2021, pp. 1–13, 2021, doi: 10.1155/2021/3084153.
- [14] K.-H. Chiang, K.-C. Chang-Chien, R.-F. Chang, and H.-Y. Yen, "Tamper detection and restoring system for medical images using wavelet-based reversible data embedding," *Journal of Digital Imaging*, vol. 21, no. 1, pp. 77–90, 2007, doi: 10.1007/s10278-007-9012-0.
- [15] F. Shih and Y. Wu, "Robust watermarking and compression for medical images based on genetic algorithms," *Information Sciences*, vol. 175, no. 3, pp. 200–216, 2005, doi: 10.1016/j.ins.2005.01.013.
- [16] H. L. Khor, S.-C. Liew, and J. M. Zain, "Region of interest-based tamper detection and lossless recovery watermarking scheme (ROI-DR) on Ultrasound Medical Images," *Journal of Digital Imaging*, vol. 30, no. 3, pp. 328–349, 2017, doi: 10.1007/s10278-016-9930-9.
- [17] Priyanka and S. Maheshkar, "Region-based Hybrid Medical Image Watermarking for secure telemedicine applications," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 3617–3647, 2016, doi: 10.1007/s11042-016-3913-1.
- [18] R. Eswarajah and E. Sreenivasa Reddy, "Robust medical image watermarking technique for accurate detection of tamperers inside region of interest and recovering original region of interest," *IET Image Processing*, vol. 9, no. 8, pp. 615–625, 2015, doi: 10.1049/iet-ipr.2014.0986.
- [19] G. Gao, X. Wan, S. Yao, Z. Cui, C. Zhou, and X. Sun, "Reversible data hiding with contrast enhancement and tamper localization for medical images," *Inf. Sci.*, vols. 385–386, pp. 250–265, Apr. 2017, doi: 10.1016/j.ins.2017.01.009.
- [20] Atta-ur-Rehman, K. Sultan, N. Aldhafferi, A. Alqahtani, and M. Mahmud, "Reversible and fragile watermarking for medical images," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–7, Jul. 2018, doi: 10.1155/2018/3461382.
- [21] X. Liu, J. Lou, H. Fang, Y. Chen, P. Ouyang, Y. Wang, B. Zou, and L. Wang, "A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images," *IEEE Access*, vol. 7, pp. 76580–76598, 2019, doi: 10.1109/ACCESS.2019.2921894.
- [22] K Swaraja, K Meenakshi, and P Kora, "An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101665, doi: 10.1016/j.bspc.2019.101665.
- [23] K. Fares, A. Khaldi, K. Redouane, and E. Salah, "DCT & DWT based watermarking scheme for medical information security," *Biomedical Signal Processing and Control*, vol. 66, p. 102403, 2021, doi: 10.1016/j.bspc.2020.102403.
- [24] S. Padmanaban, K. Thiruvankadam, S. T. Padmapriya, and R. A. M. Kumar, "An medical image file formats and Digital Image conversion," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1S3, pp. 74–78, 2019, doi: 10.35940/ijeat.A1093.1291S419.
- [25] D. Rachmawati, J. T. Tarigan, and A. B. Ginting, "A comparative study of message digest 5(MD5) and sha256 algorithm," *Journal of Physics: Conference Series*, vol. 978, p. 012116, 2018, doi :10.1088/1742-6596/978/1/012116.
- [26] V. Kumar and N. V., "Hybrid local prediction error-based difference expansion reversible watermarking for medical images," *Computers & Electrical Engineering*, vol. 53, pp. 333–345, 2016, doi: 10.1016/j.compeleceng.2015.11.033.
- [27] D. M. Thodi and J. J. Rodriguez, "Prediction-error based reversible watermarking," 2004 International Conference on Image Processing, Singapore, 2004. ICIP '04, doi: 10.1109/ICIP.2004.1421361.
- [28] J. Tian, "Reversible data embedding using a difference expansion", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890-896, 2003, doi: 10.1109/TCSVT.2003.815962.
- [29] "TCIA collections," The Cancer Imaging Archive (TCIA). [Online]. Available: <https://www.cancerimagingarchive.net/collections/>. [Accessed: 20-Nov-2021].
- [30] "NBIA - National Biomedical Imaging Archive," National Institutes of Health. [Online]. Available: <https://imaging.nci.nih.gov/ncia/login.jsf>. [Accessed: 20-Nov-2021].
- [31] S. Jaiswal, O. Au, V. Jakhetiya, Y. Guo, A. Tiwari and K. Yue, "Efficient adaptive prediction based reversible image watermarking", 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 2013.
- [32] Q. Kang, K. Li, and H. Chen, "An SVD-based fragile watermarking scheme with grouped blocks," *Proceedings of 2nd International Conference on Information Technology and Electronic Commerce*, Dalian, China, 2014, doi: 10.1109/ICITEC.2014.7105595.