A Novel Deep Learning-Assisted SVD-based Method for Medical Image Watermarking

Saima Kanwal, Feng Tao*, Rizwan Taj
School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China

***Abstract*—In the present era, the administration of medical images faces various security challenges that necessitate the authentication of image source and origin for accurate patient identification. With the increasing exchange of medical images between hospitals to facilitate informed decision-making, the adoption of digital watermarking techniques has emerged as an efficient solution to address the imperceptibility and robustness requirements in medical imaging watermarking. This research work introduces a technically advanced approach that combines singular value decomposition (SVD) watermarking with deep learning segmentation models to enhance the security of medical image sharing and transfer. The primary objective is to seamlessly integrate the watermark while minimizing distortion to preserve critical medical information within the image. The proposed methodology involves utilizing a ResNet-based U-Net segmentation model to segment X-Ray radiographs into the Region of Interest (ROI) and the Region of Non-Interest (RONI). The watermark data is then encoded into the ROI using singular value decomposition. Subsequently, the ROI and RONI are merged to reconstruct the complete image, preserving its original identity. Additionally, XOR encryption is applied to the watermarked image to enhance data integrity and copyright protection. On the other side of the methodology, the reconstructed image is once again separated into ROI and RONI. The ROI is decoded to recover the original transferred content. To assess the efficacy of the proposed method, a publicly available X-Ray radiograph dataset is employed, and evaluation metrics demonstrate an impressive segmentation accuracy of 98.27%. The proposed approach ensures information integrity, patient confidentiality during data sharing, and robustness against various conventional attacks, demonstrating its effectiveness in the field of medical image watermarking.**

***Keywords*—Singular value decomposition; medical image watermarking; digital watermarking; deep learning**

I. INTRODUCTION

With the introduction of 5G networks as an example, communication technologies have advanced quickly, dramatically altering many facets of daily life [1]. This shift in paradigm has been facilitated by simultaneous significant progressions in the fields of data analytics, computing in the cloud, and online storage. Emerging within the framework of the Internet of Medical Things (IoMT), novel approaches for both diagnosis and treatment have surfaced. These encompass telemedicine, web medicine, systematic diagnosis, and smart medicine. These cutting-edge trends are accompanied by highly advanced diagnostic and therapeutic tools, state-of-the-art sensors for medical, immersive virtual reality technologies, and complex artificial surgical methods [2]. Advancements in healthcare tech merge for better diagnosis & treatment, reshaping the ecosystem, empowering doctors, and enhancing overall care. Medical imaging transforms through e-diagnosis workflows, forming core modern healthcare structures. Hospital systems utilize HIS and advanced imaging platforms for seamless management of varied digital images (X-ray, Ultrasound, MR, and CT) [3]. PACS in hospitals securely store and retrieve these images, while DICOM standardizes their acquisition, transmission, storage, and exchange with patient data. By synergistically leveraging the functionalities of HIS, medical imaging platforms, PACS, and adhering to the DICOM standard, healthcare professionals are empowered with swift access to an extensive collection of high-fidelity digital images [4]. Seamless integration ensures precise diagnosis, treatment planning, and overall care. Interchange of medical images aids diagnosis, therapy, education, and consultations within established PACS workflows. Yet, external transmission to third parties poses potential risks of unauthorized image alterations, impacting accuracy and potentially endangering patients' lives [5]. As a result, maintaining the integrity and authenticity of medical images has become of utmost importance, calling for the implementation of strong controls and strict protections both within internal systems and during the transmission and exchange processes with external systems. A secure and reliable framework for the lifecycle management of medical images must be established, which necessitates the adoption of cutting-edge methods and technology.

Several strategies are used to handle the challenges associated with copyright protection, the confidentiality of healthcare images, and the diagnostic and personal information of patients. For these goals, methods like digital picture watermarking, cryptography, and steganography are frequently used. One of these techniques, digital picture watermarking, has attracted a lot of attention from researchers due to its unique characteristics that might offer solutions for protecting copyright and safeguarding medical data. Digital image watermarking makes it possible to incorporate subtle or noticeable watermarks into medical photographs, allowing copyright enforcement, integrity verification, and authentication [6]. Watermarking conceals metadata in images, safeguarding patient privacy and ensuring traceability. It operates in spatial and frequency domains: spatially by adjusting pixel grey levels and frequency-wise via methods like Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) to embed the watermark into the image [7]. The secret information is then incorporated by tampering with the frequency coefficients after conversion. The amplitudes and phases of the various frequency components that are present in the image are represented by these coefficients. By

*Corresponding Author.

changing these coefficients, the watermark is smoothly included into the image's representation in the frequency domain. Based on how the embedded watermark is extracted, watermarking methods are further divided into different categories. For non-blind watermarking, the source image and the watermarked image are both necessary in order to obtain the watermark. As a result, both the original host image and the matching watermark image must be accessible during the extraction process [8]. In contrast, semi-blind watermarking entails using both the watermarked image and the secret key throughout the extraction process. The watermark is included in the watermarked image and the secret key acts as a cryptographic parameter. Finally, only the secret key is required to extract the watermark in blind watermarking techniques. Both types of watermarking techniques, robust watermarking and fragile watermarking, each serve a particular function in the context of data protection and verification [9]. Watermarking aims for resilient data embedding, crucial for protecting against tampering during transfer, especially for copyright protection. Resilience determines resistance to outsider attacks or unintentional distortions during transmission. Fragile watermarking, on the other hand, detects alterations in watermarked content, ensuring integrity and authenticity. It verifies material reliability by identifying inconsistencies, serving as a vital tool for content authentication and integrity checks [10]. When discussing watermarking, as we are dealing with images, the terms "frequency domain" and "spatial domain" in the context of imaging apply to various images' representations [11]. In imaging, the term "spatial domain" describes how a picture is represented as a grid of pixels, where each pixel holds information about color or intensity and correlates to a particular place. On the other hand, the image is evaluated in terms of its frequency's components, which show differences in brightness or color across different frequencies, in the frequency domain [12]. In comparison to spatial domain watermarking, frequency domain watermarking provides higher robustness to routine image processing operations, improved imperceptibility, increased embedding capacity, resistance to geometric changes, and improved security against attacks. Frequency domain approaches are less prone to perceptual degradation and information loss brought on by spatial processes since the watermark is embedded in the frequency coefficients. Higher imperceptibility is ensured by the judicious distribution of watermark energy over several frequency bands. By using more coefficients for watermark insertion, frequency domain watermarking also offers a higher embedding capacity [13]. Additionally, it demonstrates enhanced security against statistical assaults and improved resistance to geometric transformations. Frequency domain watermarking is a favored option for many applications in digital data protection and authentication due to these benefits.

This article introduces a blind watermarking method enabling watermark embedding and extraction without the original content. This approach ensures robustness and authenticity. XOR encryption fortifies security, preventing unauthorized access or changes to the watermarked image. Encryption safeguards content, necessitating a decryption key for access. Leveraging techniques like Singular Value Decomposition (SVD), the host media undergoes transformations, embedding the watermark imperceptibly [14]. During extraction, retrieved components and a secret key recover the watermark. This blind technique shows resilience against various attacks like compression and noise, ideal for copyright protection, authentication, integrity verification, and digital forensics. Integrating U-Net with ResNet50 enhances the method. Their combo ensures precise segmentation, preserving details in the region of interest and aiding targeted watermarking. This model balances accuracy and efficiency, suitable for real-time or large datasets. Leveraging this advanced segmentation boosts the proposed method's performance and effectiveness in watermarking [15]. Through this approach, we enhance the quality of segmentation, facilitate effective watermarking, an advanced level of security and integrity for digital media, and enable efficient image analysis and protection.

Here are our primary contributions.

- Introducing a novel blind watermarking approach that seamlessly integrates U-Net segmentation with pre-trained ResNet50, enabling watermark extraction without original content access.
- Achieving precise ROI and RONI segmentation through U-Net, ensuring efficient and targeted watermark embedding while minimizing interference.
- Incorporating XOR encryption for watermarked image protection, preventing unauthorized access and modification.
- Addressing limitations in medical image sharing, applicable in copyright protection, content authentication, data integrity verification, and digital forensics.

The article is structured as follows: Section I introduces the research problem and outlines the proposed methodology, including the key contributions. Section II reviews related work in watermarking and segmentation models. Section III elaborates on the proposed blind watermarking technique, integration of U-Net and ResNet50, and XOR encryption. Section IV details the experimental setup, dataset, evaluation metrics, and discusses results, comparing with existing methods. Section V concludes by summarizing contributions and implications.

II. Related Work

Digital watermarking encompasses four fundamental concerns: imperceptibility, robustness, capacity, and security, all of which play an important role in the design and evaluation of watermarking algorithms. Researchers have invested considerable efforts in improving these aspects, especially in blind scenarios where access to the original host images is limited. However, existing watermarking methods often exhibit deficiencies in terms of robustness, transparency, and payload capacity under blind conditions. In this context, a hybrid domain approach combining three methodologies, first the discrete wavelet transforms (DWT), second the discrete cosine transforms, and the third singular value decomposition was proposed by [16] to implement a non-blind watermarking technique. Photography and text were used as watermarks and

incorporated into medical photographs. To encrypt the text watermark, a low encryption technique was used in order to streamline encryption-decryption procedures and shorten computational time. The watermark image used had dimensions of 256 x 256, whereas the cover image used was 512 x 512. Additionally, a 50-character text watermark was successfully inserted into the cover photo. Peak signal-to-noise ratio (PSNR) results for the proposed watermarking system were 35.84 dB, confirming the effectiveness and high caliber of the adopted strategy.

In order to improve decision-making processes, [17] effort proposes a novel wavelet-based digital watermarking technology designed for the transfer of medical images between institutions. Digital watermarking is essential in the medical industry for maintaining the dependability, accessibility, and privacy of images used for treatment and diagnosis. Although several approaches utilizing the spatial and transform domains have been put forth, current systems continue to run into issues with data fabrication during picture interchange. The work presents a wavelet-based digital watermarking method for medical photographs to address this problem. To help doctors make informed decisions, the scheme includes a three-level discrete wavelet transform and BCH coding. In the context of IoMT, researchers in [18], study introduces a revolutionary validated watermarking algorithm created exclusively for healthcare data. Healthcare volume data must now be transmitted and stored in a secure, dependable manner due to the growing use of IoMT technology in healthcare. In the context of the Internet of IoMT, this research introduces a reliable zero-watermarking technique-based on 3D hyper chaos and 3D dual-tree complicated wavelet transform. To create a reliable binary sequence as the watermark, the approach makes use of enhanced perceptual hashing algorithms and selective binarization of low-frequency components. Using zero embedding and blind extraction techniques, the system effectively defends against attacks and geometric distortions while maintaining the authenticity of the original clinical volume data. In terms of normalized correlation value under geometric assaults, it outperforms existing algorithms and offers bandwidth efficiency while still meeting the security criteria for sending and storing large volumes of clinical information. Utilizing the features of the Human Visual System (HVS), a blind watermarking technique was used in study [19] to insert numerous watermarks in a cover image. The watermark values were created by applying a specified threshold value to the first column of the orthogonal U matrix that was produced by applying singular value decomposition. By balancing the Normalized Cross Correlation (NCC) and the invisibility of the resulting watermarked image, the ideal threshold was found. The experimental results illustrate the proposed scheme's robustness and highlight its notable resistance to a variety of attack scenarios, demonstrating the value of using numerous watermarks. To enhance the security of the sharing and transmission of medical images, our research diligence proposes a novel and technically complex approach that combines singular value decomposition watermarking with deep learning segmentation models. The segmentation task is prioritized as a crucial component of the introduced methodology in addition to watermarking. In order to precisely define anatomical features and anomalies in medical images, a variety of segmentation methodologies have been introduced by specialists in the field of healthcare imaging. In this context, for precise segmentation of brain tumors disease from MR images, the study in [20] presented a hybrid technique that combines the DenseNet and U-Net segmentation algorithms. This study's main goal is to use deep learning techniques to precisely locate and define brain tumors in MR images. The U-Net architecture, a well-known deep learning network, is combined with a pre-trained DenseNet121 architecture in the hybrid model to improve segmentation. Smaller tumor sub-regions with intricate structural properties are given more consideration throughout the training and testing phases. The proposed approach is evaluated using the publicly available BRATS 2019 brain tumor dataset, encompassing both high-grade and low-grade glioma tumors. An advanced deep learning methodology was put forth for the segmentation of pneumothorax in chest X-ray pictures in another noteworthy study [21]. The strategy makes use of the effective EfficientNet and ResNet architectures, as well as the strong and powerful U-Net architecture. They unveiled a brand-new end-to-end semantic segmentation model for medical pictures called Ens4B-UNet. To provide incredibly accurate segmentation results, this novel method combines the power of four U-Net topologies with backbone networks that have already been trained. Ens4B-UNet improves the U-Net framework by using nearest-neighbor up-sampling in the decoders and using strong convolutional neural networks (CNNs) as the foundation for the U-Net encoders. The segmentation network's formulation, which achieves excellent performance, is a weighted average ensemble of the four encoder-decoder models. They claim an outstanding mean Dice similarity coefficient (DSC) of 0.8608 on the test data, the Ens4B-UNet model placed among the top 1% of systems in the prestigious Kaggle competition.

The existing methodologies in digital watermarking encounter several unresolved challenges that necessitate the development of advanced approaches. One major limitation is the reliance on access to the original unmarked content during the extraction time, hindering the extraction of watermarks when the original content is unavailable. This restricts the applicability of existing methods in scenarios where independent extraction is required. Moreover, many current approaches lack robustness against common interruptions resulting in the degradation of watermark quality and integrity. Furthermore, there is a pressing need for accurate and efficient segmentation techniques to precisely delineate the region of interest, facilitating targeted watermark embedding. To address these unresolved issues, our proposed methodology combines blind watermarking, utilizing the U-Net model with ResNet50 as a bottleneck, and encryption of the watermarked image. This integrated approach offers enhanced security, integrity, and resilience, effectively mitigating the limitations of existing methodologies.

III. PROPOSED METHODOLOGY

The hybrid strategy for digital watermarking in this proposed methodology combines the U-Net model with pre-trained ResNet50 for image segmentation and the singular value decomposition method for watermarking. The foundation for precise and thorough segmentation of the input images into

separate areas of interest is the U-Net model, which is renowned for its efficiency in semantic segmentation tasks. The highly effective feature extraction capabilities of the pre-trained ResNet50 are used to increase segmentation accuracy. The appropriate watermark information is then embedded within the segmented ROIs using the SVD-based watermarking technique, assuring imperceptibility and resilience. To accomplish reliable and secure watermark embedding, the process entails several crucial steps.

A. Watermark Generation

In this proposed approach, the watermark utilized encompasses multiple components, each serving a specific purpose within the extraction process, shown in Fig. 1. The first component is dedicated to patient identification and includes details such as the patient's name, gender, contact number, and address. These attributes aid in uniquely identifying the individual associated with the medical image. The second component focuses on providing information related to the acquisition of the medical image. This includes data such as the medical center responsible for generating the image, the attending doctor's name, and the timestamp indicating when the data was gathered. By incorporating this information, the source and author of the image can be effectively identified. By combining these two components, the proposed watermarking technique enhances the overall security and traceability of medical images. It allows for accurate patient identification and ensures the authenticity and origin of the image data, enabling efficient management and reliable attribution in medical imaging scenarios.

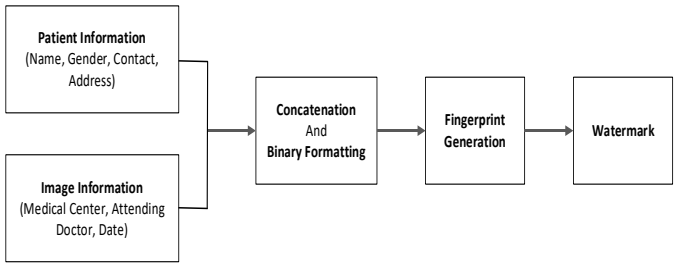

Fig. 1. Watermark generation.

B. U-Net Model

- U-Net model is frequently employed for image segmentation tasks [21]. It has an encoder-decoder structure with skip links that makes it possible to localize object boundaries with accuracy. The encoder extracts context and information from source image, while the decoder module builds a segmentation map at the pixel level. Here is a detailed description of how the U-Net model functions and avoids combining SI and CGS units, such as current in amperes and magnetic field in oersted. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

1) Encoder: Convolutional layers are succeeded by max-pooling layers, constituting the encoder. To extract various information from source data, each convolutional layer uses a separate set of filters. As we dig deeper into the encoder, the number of filters often rises, enabling the model to capture more intricate information. The max-pooling layers shrink the feature maps' spatial dimensions, which aids in expanding the receptive field and lightening the computational burden.

Let's refer to the encoder as consisting of N encoder blocks, each represented by the function Ei, and the input picture as X. Ei(X) stands for the ith encoder block's output.

2) Bottleneck: The output is then sent to the bottleneck layer after going through each encoder block. The bottleneck layer is frequently made up of several convolutional layers that aid in further capturing high-level and abstract characteristics from the input. In our proposed methodology, the pre-trained ResNet50 is utilized as the bottleneck for U-Net architecture. The term "bottleneck" refers to a specific component within ResNet50 that serves as a feature extractor. By integrating ResNet50 as the bottleneck in the U-Net model, we leverage its powerful feature representation capabilities to enhance the segmentation process. In this context, the ResNet50 model is responsible for extracting high-level semantic features from the input medical images. It is an essential transitional stage between the U-Net architecture's encoder and decoder modules. The ResNet50 bottleneck layer decreases the number of dimensions of the map features while retaining crucial data, facilitating the efficient extraction of meaningful features. The integration of ResNet50 as the bottleneck within the U-Net architecture enables our proposed methodology to benefit from the rich representation capabilities of ResNet50 for feature extraction. This combination enhances the U-Net model's ability to accurately segment medical images, leveraging the precise localization capabilities of U-Net along with the comprehensive feature representation of ResNet50.

Let's write B(X) for the bottleneck layer's output.

3) Decoder: Convolutional layers are applied after a sequence of upsampling and concatenation steps make up the decoder. The upsampling layers, responsible for augmenting the spatial dimensions of the feature maps, enable the decoder to produce a segmentation map pixel by pixel, matching the dimensions of the input image. For both local and global context information, the concatenation procedure merges the feature maps from the encoder and decoder.

Let's write Di for the ith decoder block and Di(B(X)) for the ith decoder block's output.

4) Final segmentation map: Applying a 1x1 convolutional layer to the final decoder block's output yields the final segmentation map. In order to match the number of classes in the segmentation task to the number of channels, a 1x1 convolutional layer is used. The resulting segmentation map will be referred to as S(X), with X standing for the input image. The U-Net model's forward pass can be described as follows:

$$S(X) = 1x1Conv(DN(B(X))) \quad (1)$$

In this equation, DN represents the Nth decoder block, which consists of up sampling and convolutional layers to recover spatial resolution. B represents the bottleneck layer, which serves as the bridge between the encoder and decoder, capturing the high-level semantic information. In order to limit the number of streams to the appropriate number of classes, the result of the bottleneck layer is then routed to the 1x1 convolutional layers, abbreviated as 1x1Conv. The segmentation map S(X) that is produced has the same spatial dimensions as the original image. A class probability vector, which indicates how likely it is that each pixel in the map belongs to a particular class, is present for each pixel. To obtain the ultimate pixel-by-pixel segmentation outcome, every pixel is attributed to the class exhibiting the highest probability.

C. Singular Value Decomposition

In order to facilitate numerous operations and analysis, the matrix factorization method named singular value decomposition (SVD) decomposes a matrix into three distinct matrices [22]. In the context of image processing and watermarking, SVD plays a crucial role in embedding and extracting watermarks while preserving the integrity of the source image. Mathematically, given an m × n matrix A, the SVD of A can be represented as follows:

$$A = U\Sigma V^T \quad (2)$$

where:

- U is an m × m orthogonal matrix, representing the left singular vectors.
- Σ is an m × n diagonal matrix with non-negative elements, known as singular values.
- V^T is the transpose of an n × n orthogonal matrix V, representing the right singular vectors.

The singular values in Σ are ordered in descending order, indicating their significance in capturing the image's energy or information. The higher singular values correspond to the most essential features of the image. During the watermarking process, the SVD technique is applied to the selected regions of interest (ROIs) within the image. These ROIs are represented as matrices, which are decomposed using SVD. The watermark data is then embedded into the singular values of the ROI matrix while preserving the orthogonal matrices U and V^T. The watermark embedding process involves modifying the singular values in Σ by adding or subtracting a certain value or pattern based on the watermark information. The modified Σ, along with the original U and V^T matrices, is used to reconstruct the watermarked ROI. To extract the watermark, the SVD is applied to the watermarked ROI, yielding the modified Σ matrix. By comparing the modified Σ with the original Σ obtained from the original ROI, the watermark information can be retrieved.

D. XOR Encryption

The bitwise exclusive OR (XOR) function is used to encrypt binary data using the straightforward XOR encryption method. It involves performing an XOR operation between the binary data of the watermarked image and a secret key to produce the encrypted version of the image.

The XOR encryption process can be represented as follows:

- Let E be the encrypted image, I be the original watermarked image, and K be the secret key.
- Convert the image I and the secret key K into binary representations.
- Perform the XOR operation between each corresponding bit of I and K.
- The result of the XOR operation is the corresponding bit of the encrypted image E.

By applying the XOR operation to each bit of the image and the secret key, the encrypted image is obtained. This process ensures that the encrypted image cannot be easily understood without knowledge of the secret key. Fig. 2 depicts the entire methodology and gives an extensive overview of the procedure. The segmentation stage, when the data is integrated into the image, is shown in Fig. 2 (a) and Fig. 2 (b) then depicts the decoding procedure, in which the embedded data is retrieved from the watermarked image.

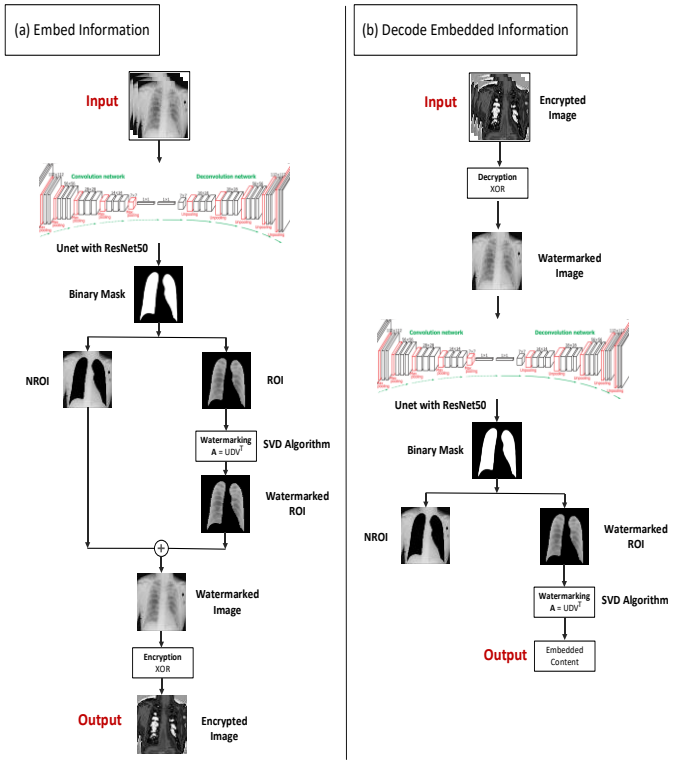

Fig. 2. Proposed architecture.

In order to properly recover the original content from the watermarked image and guarantee accurate retrieval of the encoded data, this phase makes use of sophisticated algorithms and computations. These two subfigures together shows how segmentation and decoding are seamlessly integrated into the proposed methodology. It involves two algorithms that are pivotal in achieving the desired outcomes. Algorithm 1, titled "Segmentation and Watermarking," elucidates the intricate processes illustrated in Fig. 2(a). This algorithm comprehensively outlines the step-by-step execution of segmenting the image and embedding the watermark within the

identified regions using techniques such as U-Net and ResNet50. Commencing with preprocessing the input image along with the pertinent information intended for watermark embedding, the resultant image is fed into a segmentation model. This input image, denoted as I, is then passed through a U-Net segmentation model, yielding a segmented binary mask. This mask effectively segregates the region of interest from the region of non-interest. Subsequently, the watermark W is transformed into an appropriate format and embedded into the region of interest within the initial image. Following the watermark embedding process, the regions of interest and non-interest are amalgamated to undergo an encryption procedure. The outcome of this encryption is an encrypted image, primed for transmission.

Algorithm 1: Segmentation and Watermarking

Inputs:
- Image (I): Input image for segmentation and watermarking
- Watermark (W): Information to be embedded as a watermark.
- EncryptionKey (EK): Key used for encryption

Outputs:
- Segmented Image (SI): Image with segmented regions
- Watermarked Image (WI): Image with embedded watermark

Step 1: Procedure PreprocessImage(I):
a. Normalize the pixel values of the image I.
b. Perform any necessary image enhancement or noise reduction techniques.

Step 2: Procedure ApplySegmentationModel(I):
a. Pass the pre-processed image I through the U-Net architecture with a pre-trained ResNet50 as the bottleneck.
b. Obtain the segmented regions by applying a threshold or post-processing techniques to the output of the segmentation model.
c. Generate the segmented image SI by overlaying the segmented regions on the original image I.

Step 3: Procedure SegmentImage(I, SI):
a. Subtract the non-segmented regions from the original image I to obtain the ROI.
b. Subtract the segmented regions from the original image I to obtain the RONI.

Step 4: Procedure EmbedWatermark(ROI, W):
a. Convert the watermark W into a suitable format for embedding, such as a binary sequence or a transform domain representation.
b. Iterate over the pixels in the ROI:
For each pixel P in the ROI:
Modify the pixel value of P to embed the corresponding watermark bit.

Step 5: Procedure CombineImage(ROI, NROI):
a. Generate the watermarked image WI by combining the modified ROI with the RONI from the original image I.

Step 6: Procedure EncryptionImage(WI, EK):
a. Output the encrypted watermarked image EWI by applying XOR watermarking technique.

Step 7: Procedure SegmentationAndWatermarking(I, W):
PreprocessImage(I)
ApplySegmentationModel(I)
SegmentImage(I, SI)
EmbedWatermark(ROI, W)
CombineImage(ROI, NROI)
EncryptionImage(WI)
SegmentationAndWatermarking(I, W)
Return EWI
End of Algorithm

On the other hand, Algorithm 2, titled "Content Extraction from Watermarked Image," delineates the procedures depicted in Fig. 2 (b).

Algorithm 2: Content Extraction from Watermarked Image

Input:
- Encrypted Watermarked Image (EWI): Encrypted image with embedded watermark
- Pre-trained ResNet50 Model (M): Pre-trained ResNet50 model for segmentation
- Decryption Key (DK): Key used for Decryption

Output:
- Extracted Watermark (EW): Embedded content extracted from the watermarked image

Step 1: Procedure DecryptWatermarkedImage(EWI, DK):
a. Apply decryption to output the watermarked image WI.

Step 2: Procedure Segmentation(WI, M):
1. Apply the pre-trained ResNet50 model to the Watermarked Image to obtain the Segmentation Map.
2. Apply a threshold to the Segmentation Map to obtain a binary mask using the Segmentation Threshold.

Step 3: Procedure SegmentImage(WI):
a. Subtract the non-segmented regions from the original image I to obtain the ROI.
b. Subtract the segmented regions from the original image I to obtain the RONI.

Step 4: Procedure ExtractWatermark(ROI):
1. Apply singular value decomposition to the watermarked ROI using the SVD Parameters.
2. Retrieve the modified SVD coefficients from the Watermarked Image.
3. Initialize an empty array to store the extracted content.
4. For each pixel in the Watermarked Image:
a. Check if the corresponding pixel in the binary mask is non-zero.
b. If non-zero, extract the content from the modified SVD coefficients corresponding to the pixel.
c. Append the extracted content to the array.

Step 5: Procedure ContentExtraction(WI, M, DK):
DecryptWatermarkedImage(EWI, DK):
Segmentation(WI, M):
SegmentImage(WI):
ExtractWatermark(ROI):
ContentExtraction(WI, M, DK)
Return EW
End of Algorithm.

This algorithm intricately illustrates the sequential operations involved in extracting the information from the embedded watermarked image, enabling the reconstruction of the original content. Subsequently, the reverse sequence is initiated, commencing with the application of a decryption

algorithm on the encrypted input image. This decryption process yields a watermarked image, which is subsequently provided as input to the segmentation model to retrieve the regions of interest and non-interest. Following this, the SVD watermarking algorithm is implemented to extract the watermark information from the watermarked region of interest.

IV. Experimental Results and Discussions

For the experiments, we trained our algorithm on a medical imaging dataset consisting of chest radiographs. The training was performed on a Google Colab GPU to leverage its computational power. The dataset was carefully curated and pre-processed to ensure high-quality and standardized images. We used a batch size of 32 and trained the model for 50 epochs with an initial learning rate of 0.0001. We employed the Adam optimizer with a learning rate decay schedule to facilitate convergence. To maximize both segmentation accuracy and border delineation, a loss function that combines binary cross-entropy loss and dice loss was employed. We partitioned the dataset randomly into training and testing sets, with an 80:20 split, in order to assess the effectiveness of our algorithm. We carried out numerous studies and presented the findings using a variety of evaluation metrics, including the dice coefficient, intersection over union, and precision-recall curves. The experimental setup ensured rigorous validation and comparison of our proposed system against state-of-the-art methods in medical image segmentation and watermarking.

A. Datasets

The experimental setup involved training our algorithm the ChestX-Det10 dataset [23], a subset of the NIH ChestX-14 dataset [24], which is a widely used and comprehensive dataset for chest radiograph analysis and contains instance-level annotations. This dataset comprises a total of 3,543 images, with 2,779 images depicting various diseases and 764 images representing healthy x-rays. It offers a diverse range of diseases, including atelectasis, calcification, consolidation, effusion, emphysema, fibrosis, fracture, mass, nodules, and pneumothorax, making it suitable for training a fully supervised segmentation method. Fig. 3 shows a sample from the dataset.

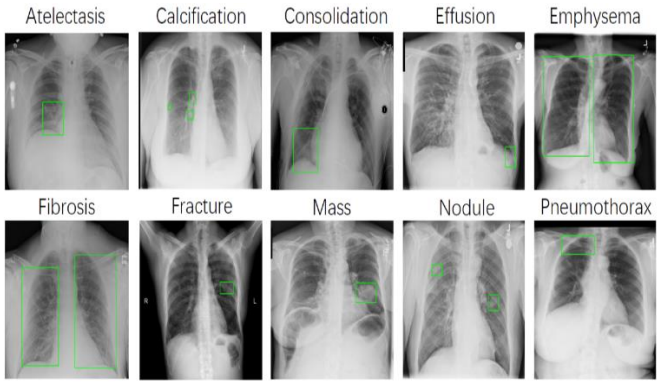

Fig. 3. Samples from dataset.

B. Performance Comparison

The quantitative results of our strategy were compared to those of existing methodologies in the comparison results, and the outcomes are listed in Table I. It is important to note that some methodologies exhibited PSNR values exceeding 50 dB, which can be attributed to differences in the input data format used for evaluation. However, it is crucial to consider that the proposed method maintained a consistent and comparable average PSNR value within the range of 49.32 to 50 dB, indicating its robustness and effectiveness in preserving image quality. The PSNR values per image are depicted in Fig. 4, showcasing that our model attains a PSNR value of 50.0 for image 1, 49.73 for image 2, 47.17 for image 3, 49.79 for image 4, and 49.91 for image 5.

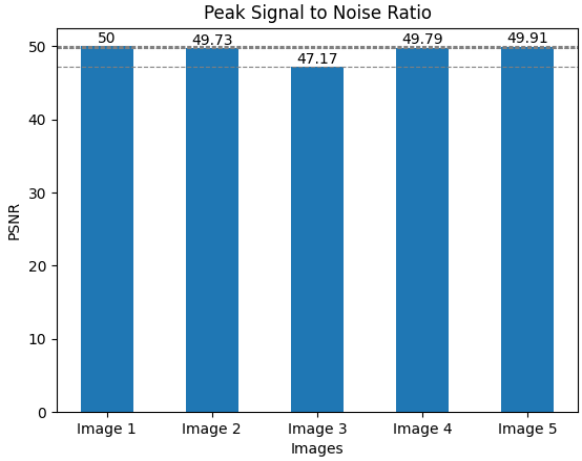

Fig. 4. PSNR values after watermarking.

Furthermore, the average SSIM values obtained by the proposed method demonstrated a high level of similarity and structural preservation with the original image, outperforming several existing methodologies. Similarly, the SSIM values per image are illustrated in Fig. 5, revealing that our model achieves an SSIM value of 1.0 for image 1, 0.99 for image 2, 0.99 for image 3, 0.98 for image 4, and 0.99 for image 5.

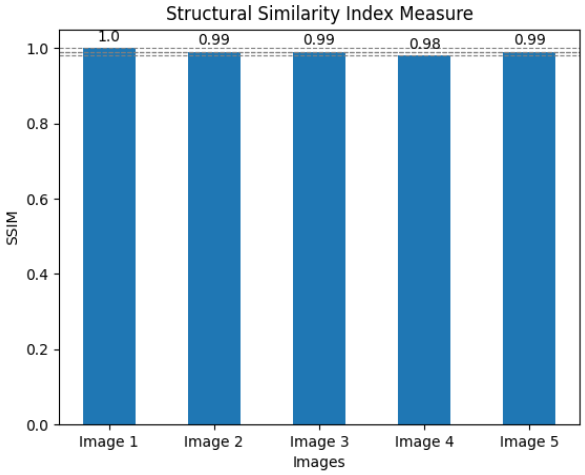

Fig. 5. SSIM values after watermarking.

These results underscore the superiority of the proposed method in terms of both PSNR and SSIM metrics, highlighting its potential for accurate and reliable watermarking in medical image applications.

TABLE I. COMPARISON OF PSNR AND SSIM RESULTS WITH EXISTING STUDIES

References			
	Method	***PSNR/dB***	***SSIM***
Soni, M., et al. [17]	DWT	98.58	Not applied
Ernawan, F., et al. [19]	SVD	Not applied	0.88
Balasamy,K.,& Ramakrishnan, S. [25]	DWT and PSO	49.00	0.99
Balasamy, K., et all. [26]	DWT and SVD	49.5	Not applied
Kahlessenane, F., et all. [27]	DWT	49.20	0.99
Balasamy, K., et all. [28]	SVD	49.30	Not applied
Wang, L., et all. [29]	DWT, HMD, and SVD	44.90	Not applied
Sanivarapu, P. V., et all. [30]	DWT, SVD, RSA	39.42	Not applied
Khaldi, A., et all. [31]	DWT, IWT	58.09	0.99
Apostolidis, K. D., et all. [32]	Krawtchouk Moments	Not Applied	0.99
Proposed	**U-Net, SVD, XOR**	**49.32**	**0.99**

C. Qualitative Evaluation

We used a variety of evaluation metrics, such as Intersection over Union (IoU), Dice coefficient, and F1 score, in our qualitative evaluation of the proposed approach. These metrics offer useful information about the effectiveness and precision of the segmentation findings. The Dice coefficient quantifies how similar the two sets are while the intersection over union assesses the overlap between the anticipated and real-world masks. The F1 score also assesses how well precision and memory are balanced. We were able to fully comprehend the algorithm's segmentation abilities and its capacity to precisely outline the regions of interest in the medical images by making use of these several assessment matrices. The plot of training and validation loss, shown in Fig. 6, provides valuable insights into the learning progress of our model. During the training phase, the model undergoes iterative optimization, where the loss is minimized to improve its performance. As the training progresses, the model's training loss steadily decreases. In our case, the training loss reaches a remarkable value of approximately 0.02, indicating that the model has learned to capture and generalize patterns effectively from the training data.

Similarly, the validation loss, which measures the model's performance on unseen data, also decreases during training. An approximate validation loss of 0.03 signifies effective generalization of the model to novel data, demonstrating its capability to provide accurate predictions even for previously unseen instances. The convergence of both training and validation loss, shown in Fig. 6, to such low values demonstrates the effectiveness of our model in capturing complex patterns and achieving high accuracy in segmentation tasks. It highlights the model's ability to generalize well and indicates its potential for robust performance in real-world scenarios.

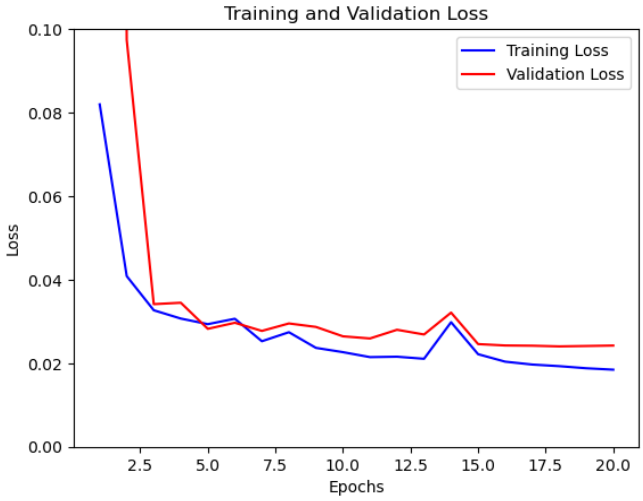

Fig. 6. Training and validation loss.

The findings and details of the matrices are given below:

1) Intersection over union: A frequently used evaluation statistic for gauging the precision of segmentation models is intersection over union. It quantifies the degree of overlap between the anticipated segmentation mask and the obtained segmentation mask. It determined mathematically by dividing the intersection area of the predicted mask (P) and the ground truth mask (G) by the union area of these masks:

$$\text{IoU} = |P \cap G| / |P \cup G| \quad (3)$$

where, $|P \cap G|$ represents the area of intersection between P and G, and $|P \cup G|$ represents the area of union.

The intersection over the union metric spans a range from 0 to 1. A score of 0 denotes a complete absence of overlap between the predicted and ground truth masks, while a score of 1 signifies a flawless alignment. In the context of our proposed methodology, we utilized IoU as the evaluation metric to assess the results produced by our U-Net model with ResNet50 as the backbone. As shown in Fig. 7, we were capable of objectively evaluating the accuracy of the segmentation by computing the IoU score for each segmented area and quantitatively measuring the degree of concordance between the expected and actual masks.

2) Dice coefficient: In the context of image segmentation, the Dice coefficient is a frequently used evaluation metric for determining how similar two sets are. The Dice coefficient is determined mathematically as the reciprocal of the intersection area between the P and the G divided by the sum of the P and G intersection areas:

$$\text{Dice Coefficient} = 2 * |P \cap G| / (|P| + |G|) \quad (4)$$

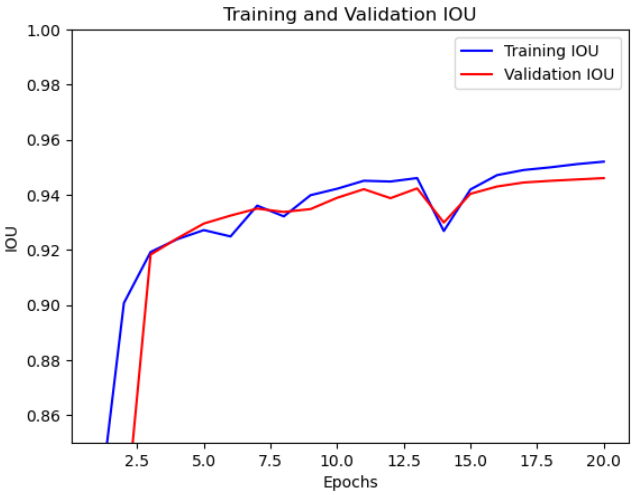

Fig. 7. IoU result.

where, |P ∩ G| represents the area of intersection between P and G, |P| represents the area of P, and |G| represents the area of G. The range of the Dice coefficient is 0 to 1. Between the anticipated and ground truth masks, a value of 0 indicates there is no overlap and a value of 1 indicates full congruence. In the context of our proposed methodology, we employed the Dice coefficient as the assessment metric to gauge the precision and excellence of the segmentation outcomes. Through the utilization of the Dice coefficient, we conducted a rigorous quantitative assessment of the segmentation performance within our proposed methodology. By computing the Dice coefficient for each segmented region, we were able to precisely evaluate the extent of concurrence between the predicted and the ground truth, thereby delivering a reliable metric for segmentation accuracy. As depicted in Fig. 8, our evaluation showcased exceptional performance, with the resulting plot graph indicating a Dice coefficient of approximately 0.98, which represents the highest achievement on the validation dataset. This outcome underscores the efficacy and robustness of our approach in achieving superior segmentation outcomes.

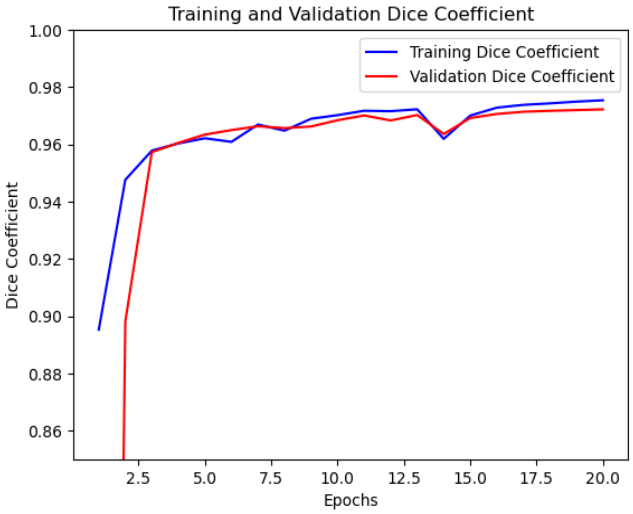

Fig. 8. Dice coefficient result.

3) F1 score: In the proposed methodology, the F1 score was utilized as an evaluation metric to assess the performance of the segmentation model. The performance is evaluated using the F1 score, which takes precision and recall into account. It provides a balanced assessment by taking into account the trade-off between correctly identifying positive samples (precision) and capturing all actual positive samples (recall). Mathematically, the F1 score is calculated as follows:

$$F1\ Score = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall} \quad (5)$$

Recall corresponds to the ratio of true positive predictions to the total number of positive predictions, while precision quantifies the ratio of true positive predictions to the overall instances identified as positive. An increase in the F1 score, which runs from 0 to 1, suggests improved segmentation performance. By using the F1 score as an evaluation metric, we were able to evaluate the segmentation model's capability to precisely detect the regions of interest within the medical images, considering both precision and recall simultaneously. This metric provided a comprehensive measure of the model's performance in capturing the relevant features while minimizing false positives and false negatives pixels. The evaluation of our proposed model on the validation data demonstrated outstanding performance, with an achieved F1 score of 0.96, shown in plot of Fig. 9. This high F1 score indicates the model's exceptional precision and recall values, highlighting its efficacy in achieving accurate and reliable segmentation results. The superior performance of our proposed model underscores its capability to effectively handle complex medical imaging data and accurately delineate regions of interest.

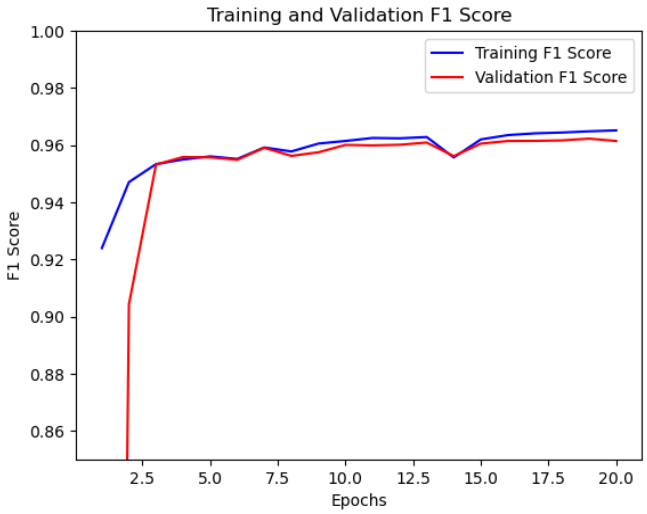

Fig. 9. F1 score result.

4) Segmentation pictorial results: The figure illustrates the input image and the comparative analysis of the output binary mask predicted by the proposed segmentation algorithm, and the ground truth binary mask. In this analysis, the input image represents the original image that was fed into the segmentation algorithm for processing. The segmentation

algorithm produces a binary mask, where each pixel is assigned a value of either 0 or 1, indicating the presence or absence of the region of interest. The ground truth binary mask serves as the expected segmentation result, obtained through manual annotation or another reliable source. By placing these three images side by side in Fig. 10, we can visually assess the performance of the segmentation algorithm by comparing the agreement between the algorithm's output and the ground truth. This comparative analysis provides valuable insights into the accuracy and Efficacy of the suggested segmentation algorithm in precisely outlining the area of interest within the input image.

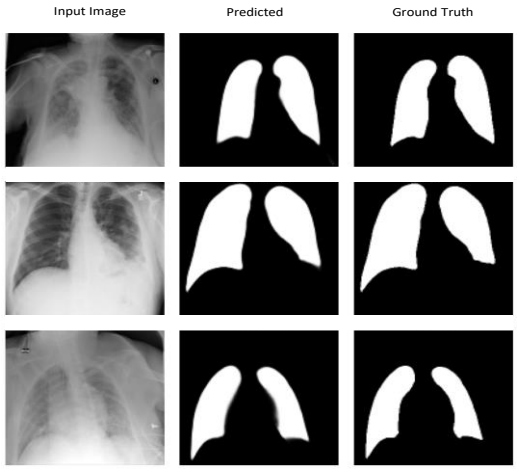

Fig. 10. Comparative result.

D. Quantitative Evaluation

In this study, two quantitative evaluation metrics, namely PSNR, and SSIM, were employed to assess the effectiveness of the introduced approach. PSNR measures the quality of the watermarked image by quantifying the ratio of the maximum signal to noise power. It produces a number that indicates how faithful and similar the watermarked image is to the original image. On the other hand, SSIM assesses the similarity between the watermarked image and source image by considering the luminance, contrast, and structural information. It offers a value between 0 and 1, with practically 1 denoting a perfect match. By utilizing both PSNR and SSIM, a comprehensive evaluation of the proposed method was conducted, allowing for an accurate assessment of its effectiveness in context of image quality preservation and similarity to the original image.

1) Peak signal-to-noise ratio: PSNR quantifies the level of distortion or noise in an image by comparing it to a reference or original image. The PSNR is determined as the mean square error of source and distorted images divided by the peak signal power. It is often expressed in decibels (dB) and provides a numerical value that indicates the similarity between the two images. Higher PSNR values indicate better image quality with less distortion or noise. Mathematically, the PSNR can be computed as:

$$PSNR = 10 * \log10((MAX2) / MSE) \quad (6)$$

where, term MAX indicates the highest pixel value that can be achieved and MSE denotes the mean square error between the original and warped image. The maximum possible PSNR value depends on the pixel intensity range of the images. For images with pixel intensities in the range of 0 to 255 for grayscale or RGB images, the highest PSNR value is typically around 30-50 dB. The proposed method attained a PSNR value of 49.329, indicating its efficacy in maintaining the fidelity and quality of the original grayscale image. By effectively embedding the watermark, the method minimizes distortion and preserves crucial visual information. The high PSNR value underscores the robustness and potential of this approach for secure and reliable image watermarking applications.

2) Structural similarity index: The proposed methodology incorporates the use of SSIM as an evaluation metric for the watermarked image. To mathematically determine SSIM, the mean, standard deviation, and covariance of the pixel intensities in the reference (original) image and the warped (watermarked) image are compared. The SSIM index ranges between -1 and 1, where 1 indicates maximum similarity and a value close to -1 indicates significant dissimilarity. The average SSIM value of 0.99 achieved by the proposed watermarking method serves as a strong testament to its remarkable efficiency in preserving the structural similarity and quality of the original image. This outstanding result demonstrates the validity of the proposed system in embedding the watermark while ensuring minimal distortion and preserving the visual integrity of the image.

V. Conclusion

In conclusion, this article presented a novel watermarking methodology that combines the strengths of the SVD algorithm and the U-Net architecture with a pre-trained ResNet50 model as the bottleneck. The proposed methodology demonstrated remarkable efficiency in preserving the fidelity and quality of the original image while effectively embedding the information using XOR encryption to ensure data integrity and copyright protection. Through rigorous experimentation and evaluation on a chest radiograph dataset, the algorithm showcased its effectiveness in accurately segmenting regions of interest and embedding watermarks while maintaining the coherence of the medical images. The integration of the U-Net model with the pre-trained ResNet50 model as the bottleneck proved to be a powerful combination, enabling the algorithm to leverage the deep learning capabilities of ResNet50 for feature extraction and the U-Net's architectural design for precise segmentation. This hybrid approach contributed to the algorithm's exceptional segmentation accuracy and its ability to preserve crucial medical information. Furthermore, the application of the SVD algorithm for watermarking provided a robust and imperceptible means of embedding and extracting information within the segmented regions. The algorithm successfully achieved secure and reliable watermarking while ensuring minimal distortion to the original medical images. The experimental results showed that, in terms of segmentation accuracy and watermark robustness, the suggested methodology outperformed previous strategies. The algorithm's high Dice coefficient, F1 score, and intersection over union

values substantiated its efficacy and accuracy in segmenting medical images and extracting embedded information.

We also have plans to evaluate the approach in many scenarios in the future, including multimedia applications and other medical imaging. Furthermore, evaluating how well the method works on various systems and maintaining compatibility will be crucial components of future research. The ultimate objective is to improve the method by utilizing cutting-edge strategies that strike a balance between security, effectiveness, and usability, opening the door for its smooth incorporation into useful domains.

REFERENCES

[1] Venkatachalam, K., Prabu, P., Alluhaidan, A. S., Hubálovský, S., & Trojovský, P. (2022). Deep belief neural network for 5G diabetes monitoring in big data on edge IoT. *Mobile Networks and Applications*, *27*(3), 1060-1069.

[2] Cao, L., Li, J., Liu, J., & Chen, Y. W. (2023, May). Robust Watermarking Algorithm for Medical Volume Data Based on PJFM and 3D-DCT. In *International KES Conference on Innovation in Medicine and Healthcare* (pp. 215-232). Singapore: Springer Nature Singapore.

[3] Aljabri, M., AlAmir, M., AlGhamdi, M., Abdel-Mottaleb, M., & Collado-Mesa, F. (2022). Towards a better understanding of annotation tools for medical imaging: A survey. *Multimedia tools and applications*, *81*(18), 25877-25911.

[4] Madhusudhan, K. N., & Sakthivel, P. (2021). A secure medical image transmission algorithm based on binary bits and Arnold map. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 5413-5420.

[5] Gull, S., & Parah, S. A. (2023). Advances in medical image watermarking: a state-of-the-art review. *Multimedia Tools and Applications*, 1-41.

[6] Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295-336). Academic Press.

[7] Wang, B., Wang, W., Zhao, P., & Xiong, N. (2022). A Zero-Watermark Scheme Based on Quaternion Generalized Fourier Descriptor for Multiple Images. *Computers, Materials & Continua*, *71*(2).

[8] Shehab, A., Elhoseny, M., Muhammad, K., Sangaiah, A. K., Yang, P., Huang, H., & Hou, G. (2018). Secure and robust fragile watermarking scheme for medical images. *IEEE access*, *6*, 10269-10278.

[9] Kelkar, V., Mehta, J. H., & Tuckley, K. (2018). A novel robust reversible watermarking technique based on prediction error expansion for medical images. In *Proceedings of 2nd International Conference on Computer Vision & Image Processing: CVIP 2017, Volume 1* (pp. 131-143). Springer Singapore.

[10] Sultan, K., Aldhafferi, N., Alqahtani, A., & Mahmud, M. (2018). Reversible and fragile watermarking for medical images. *Computational and mathematical methods in medicine*, *2018*.

[11] Yang, J., Ma, Y., Yao, W., & Lu, W. T. (2008). A spatial domain and frequency domain integrated approach to fusion multifocus images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *37*(PART B7).

[12] Liu, B., Li, Y., & Zhang, L. (2022). Analysis and visualization of spatial transcriptomic data. *Frontiers in Genetics*, *12*, 2852.

[13] Cao, F., Wang, T., Guo, D., Li, J., & Qin, C. (2023). Screen-shooting resistant image watermarking based on lightweight neural network in frequency domain. *Journal of Visual Communication and Image Representation*, *94*, 103837.

[14] Luo, Y., Li, L., Liu, J., Tang, S., Cao, L., Zhang, S., ... & Cao, Y. (2021). A multi-scale image watermarking based on integer wavelet transform and singular value decomposition. *Expert Systems with Applications*, *168*, 114272.

[15] Aboudi, F., Drissi, C., & Kraiem, T. (2022, May). Efficient U-Net CNN with Data Augmentation for MRI Ischemic Stroke Brain Segmentation. In *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)* (Vol. 1, pp. 724-728). IEEE.

[16] Singh, A. K., Dave, M., & Mohan, A. (2016). Hybrid technique for robust and imperceptible multiple watermarking using medical images. *Multimedia Tools and Applications*, *75*, 8381-8401.

[17] Soni, M., & Kumar, D. (2020, September). Wavelet based digital watermarking scheme for medical images. In *2020 12th international conference on computational intelligence and communication networks (CICN)* (pp. 403-407). IEEE.

[18] Liu, J., Ma, J., Li, J., Huang, M., Sadiq, N., & Ai, Y. (2020). Robust watermarking algorithm for medical volume data in internet of medical things. *IEEE Access*, *8*, 93939-93961.

[19] Ernawan, F., Liew, S. C., Mustaffa, Z., & Moorthy, K. (2018). A blind multiple watermarks based on human visual characteristics. *International Journal of Electrical and Computer Engineering*, *8*(4), 2578.

[20] Cinar, N., Ozcan, A., & Kaya, M. (2022). A hybrid DenseNet121-UNet model for brain tumor segmentation from MR Images. *Biomedical Signal Processing and Control*, *76*, 103647.

[21] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[22] Chandra, D. S. (2002, August). Digital image watermarking using singular value decomposition. In *The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.* (Vol. 3, pp. III-III). IEEE.

[23] Liu, J., Lian, J., & Yu, Y. (2020). Chestx-det10: chest x-ray dataset on detection of thoracic abnormalities. *arXiv preprint arXiv:2006.10550*.

[24] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

[25] Balasamy, K., & Ramakrishnan, S. (2019). An intelligent reversible watermarking system for authenticating medical images using wavelet and PSO. *Cluster Computing*, *22*, 4431-4442.

[26] Balasamy, K., & Suganyadevi, S. (2021). A fuzzy based ROI selection for encryption and watermarking in medical image using DWT and SVD. *Multimedia tools and applications*, *80*(5), 7167-7186.

[27] Kahlessenane, F., Khaldi, A., Kafi, R., & Euschi, S. (2021). A robust blind medical image watermarking approach for telemedicine applications. *Cluster computing*, *24*(3), 2069-2082.

[28] Balasamy, K., Krishnaraj, N., & Vijayalakshmi, K. (2022). Improving the security of medical image through neuro-fuzzy based ROI selection for reliable transmission. *Multimedia Tools and Applications*, *81*(10), 14321-14337.

[29] Wang, L., & Ji, H. (2022). A watermarking optimization method based on matrix decomposition and DWT for multi-size images. *Electronics*, *11*(13), 2027.

[30] Sanivarapu, P. V., Rajesh, K. N., Hosny, K. M., & Fouda, M. M. (2022). Digital Watermarking System for Copyright Protection and Authentication of Images Using Cryptographic Techniques. *Applied Sciences*, *12*(17), 8724.

[31] Khaldi, A., Kafi, M. R., & Meghni, B. (2022). Electrocardiogram signal security by digital watermarking. *Journal of Ambient Intelligence and Humanized Computing*, 1-13.

[32] Apostolidis, K. D., & Papakostas, G. A. (2022). Digital watermarking as an adversarial attack on medical image analysis with deep learning. *Journal of Imaging*, *8*(6), 155.