A Robust Watermark-based Model for Image Content Integrity Verification and Tampering Detection

Jayanti Rout, Sangram Pati, Minati Mishra*
P. G. Department of Computer Science, Fakir Mohan University, Balasore, India
E-mail: minatiminu@gmail.com
*Corresponding author

Keywords: Tamper detection, watermarking, wavelet transform, DCT, SVD, splicing, cloning

Received: May 9, 2025

The rise of digital image tampering has increased the inability to check authenticity of images. This has made its identification as well as the localization the tampered regions a significant challenge. Watermark-based content authentication methods provides an effective solution to it. This study proposes a strong and reliable watermark-based authentication algorithm. It uses two watermarks that are created using the Haar wavelet transform, discrete cosine transform, and singular value decomposition. A tamper detection mask is also used to highlight the tampered regions of the image. The embedded watermarks were observed to resist various attacks, including compression, cropping, noise addition, and blurring. This highlights the robustness of the approach. The experimental evaluations show that the proposed method achieves high imperceptibility with various near-ideal metrics. It achieved Peak Signal-to-Noise Ratio (PSNR) values > 38 dB, Structural Similarity Index Measure (SSIM) > 99%, and Normalized Cross-Correlation (NCC) of ∼99.8%. The method effectively detects various types of manipulation and attacks in images. The proposed technique shows strong applicability in various domains, including digital forensics, copyright enforcement, and the protection of sensitive multimedia contents.

Povzetek: Študija predlaga robusten in zanesljiv algoritem za preverjanje pristnosti na osnovi vodnega žiga. Metoda uporablja dva vodna žiga, ustvarjena s pomočjo Haarove valovne transformacije, diskretne kosinusne transformacije in metode SVD.

1 Introduction

The use of digital media has increased significantly in recent years. They are considered a suitable choice to easily share information. Among them, images are an impactful source of information. According to a survey, more than 2.1 trillion images are taken annually for content creation [1]. However, easy access to image editing tools poses a threat to the authenticity and integrity of digital content [2]. The reliability of media is very important, especially in the domain of journalism, publishing, law enforcement, medical diagnostics, etc. For example, in the medical field, doctors use medical images for accurate diagnosis and treatment planning. On the other hand, digital images and videos are frequently submitted as evidence in legal settings. In such scenarios, the authenticity of digital content is directly related to important decision-making. As a result, detecting manipulation or tampering of these digital media has become an important research challenge [3, 4].

The wide use of advanced digital cameras and user-friendly editing software provides a seamless editing experience. It provides flexibility for even non-expert users for alteration. In addition, these edits are done using such sophisticated methods that usually leave no noticeable visual artifacts. It can mislead viewers and can rapidly spread misinformation. This poses serious social, political, and economic risks [5]. An example of such image manipulation in 2017 is presented in Figure 1. A senior police officer is presented as kneeling before a politician. The investigation revealed that the image had been manipulated by splicing a frame from a 2011 Bollywood film into an authentic photograph [6]. Such incidents highlight the need for reliable image tampering detection techniques. A brief overview of historical cases of image tampering is presented in [7]. Digital image manipulation falls into four main categories based on the type of alteration: copy-move or cloning, splicing, retouching, and resampling [8]. In copy-move tampering, a region of an image is duplicated and pasted elsewhere within the same image. They are usually used

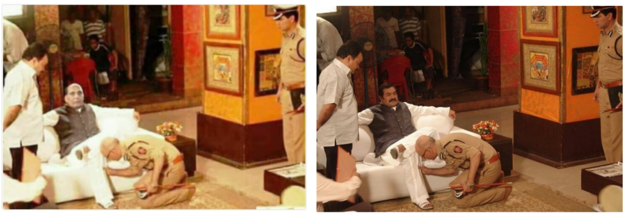


Figure 1: Illustration of image manipulation [6]: left—tampered image; right—original scene.

to hide or replicate an object. Splicing involves merging segments from different images to create a composite that misrepresents reality. Retouching focuses on improving or modifying the image to change its visual impact without changing the image geometrically. Resampling applies geometric operations such as scaling, rotation, affine transformations, etc., to align tampered regions with the surrounding context. An illustration of all these manipulations is shown in Figure 2. These techniques pose unique challenges to detection algorithms. This presents the need to develop specialized approaches for detection.

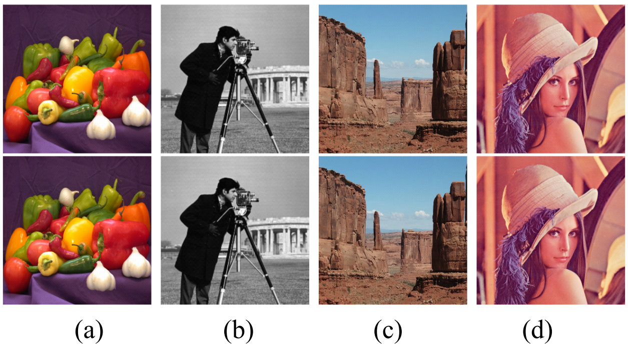


Figure 2: Row 1: the forged images: (a) cloned, (b) spliced, (c) retouched, and (d) resampled. Row 2: the original images

Tamper detection techniques are categorized into two broad types: active and passive methods [9]. Active methods involve the integration of additional information into the media during its creation or transmission. This information consists of watermarks, digital signatures, etc. These features act as integrity verifiers that can be extracted or verified to detect unauthorized alterations. Watermarking techniques can be divided into fragile and robust types. Fragile watermarks are highly sensitive to changes. They can be easily warped by even minor modifications. Although this makes them effective in detecting tampering, they are not built to handle intentional attacks. However, robust watermarking is designed to resist various image processing operations such as compression, filtering, geometric transformations, and noise addition [10]. As discussed in [11], an effective and robust watermarking scheme must meet the criteria of security, imperceptibility, and robustness, collectively known as SIR. Similarly to a watermark, digital signatures are important in active authentication. A digital signature is a cryptographic technique that binds a unique identifier to the content. This allows us to verify the origin and integrity of the image. Unlike watermarking, it does not alter the content of the media itself. This makes them suitable for scenarios where maintaining the original quality of the image is important [12]. Thus, both watermarking and digital signatures play complementary roles in active tamper detection systems.

Passive or blind tampering detection techniques work without embedded information or prior access to the original image [13]. These methods study the visual features of the image and the statistical irregularities introduced within them during manipulation. It assumes that every tampering introduces some detectable traces. These artifacts may be present in the form of irregularities in noise patterns, color distributions, compression artifacts, etc., [14]. Various statistical, machine learning, and signal processing techniques have been proposed to capture and analyze these variations. This analysis provides effective tamper localization and classification [15, 16].

Motivated by the need for an efficient image authentication system, the proposed work presents an improvement over existing dual or hybrid watermarking schemes. It proposes a strategically integrated three-level transform approach: the Discrete Cosine Transform (DCT), the Singular Value Decomposition (SVD), and the Haar wavelet transform. They are combined with a dual watermark structure that consists of a signature and a fingerprint. Unlike traditional methods that focus solely on the verification or authentication of the ownership, this method inserts the signature into the fingerprint. The resultant is then embedded in the host image. This forms a hierarchical and layered authentication mechanism. In addition, a novel tamper detection method is introduced. It uses the embedding of a tamper detection mask into the HL1 and LH2 sub-bands of the Most Significant Bit (MSB) plane using a two-level DWT. This allows authenticity to be checked and the exact manipulated regions to be located. This part was observed to be overlooked in existing work. Integration is not just a fusion of existing techniques but a carefully designed extraction process. It ensures robustness, imperceptibility, and active tamper location. It provides a comprehensive and novel solution for the authentication and protection of image content.

To evaluate the effectiveness of the approach, the following research questions (RQ) are designed to highlight the key contributions:

- RQ1: How does the proposed approach improve the robustness and efficiency of watermark-based content authentication and tampering detection compared to individual techniques?
- RQ2: To what extent does the proposed model maintain high performance on the basis of the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Normalized Cross-Correlation (NCC) when introduced to various attacks?
- RQ3: How accurately can the proposed method detect and locate tampered regions, especially in complex scenarios?

The remaining sections of the paper are organized as follows: Section 2 gives a brief overview of existing technological advancements in the domain. Sections 3 and 4 present an overview of the technologies used and explain the proposed methodology, respectively. Section 5 discusses the experimental setup and analyzes the results obtained. Section 6 concludes the paper with suggestions for future research directions.

2 Literature review

Digital watermarking is the process of inserting unique identifiers into host signals. They are mainly used to verify authenticity, ownership, or both. They have evolved significantly since their prominence in the 1990s based on methodologies and applications [17]. Watermarking is used for various purposes, including copyright protection, prevention of unauthorized duplication, assurance of digital content integrity, etc., [18]. They can be divided into two main groups based on their primary objective: authentication models and tampering detection models.

2.1 Authentication models

The healthcare sector relies on invisible watermarking due to its need for secure, durable, and efficient data protection techniques. A discrete watermarking model for the preservation of copyright in medical data is described in [19]. Continuing in this direction to safeguard medical images, the authors in [20] used DWT and SVD-based methods for invisible watermarking. Poonam and Arora [21] used DWT and SVD-based watermarking techniques. Zermi et al. [22] introduced a blind method for medical image authentication that integrates an electronic patient record (EPR) hash with image acquisition data. This approach ensures data integrity by embedding the EPR hash in medical images, making use of DWT and SVD.

The researchers in [23] suggested a hybrid digital watermarking model for Digital Imaging and Communications in Medicine (DICOM) that combines visible-imperceptible and robust-imperceptible watermarking approaches. This method preserves the integrity of EPR data while authenticating the source. The technique is applied to DICOM images with bit depths greater than 8 bpp. The approach was tested against twenty different geometric and signal processing operations and was found to be robust across all scenarios tested. The researchers in [24] proposed a watermarking model using DCT-DWT and spatial SVD for watermark embedding. It achieved a PSNR of 40 dB, an NCC of 0.9, and an SSIM value of nearly 0.99. The method is found to withstand noise attacks, filtering, geometric deformations, and compression.

Rajput et al. suggested a watermark model combining DWT, Haar Discrete (HD), and SVD that outperformed other existing models in terms of high resistance to attacks and better quality [25]. Hu et al. proposed a watermarking model that resists instruction-driven image editing and withstands instruction-driven image editing (IDIE) [26]. A Partial Instruction-Driven De-noising Sampling Guidance (PIDSG) module with the encoder-decoder training framework was integrated. It improves the strength of the watermark against IDIE and the semantic distortions introduced by IDIE. The researchers have experimentally proved that the watermark offers robustness while maintaining high visual quality and editability.

Neekhara et al. proposed a multipurpose deep learning (DL)-based watermarking technique that can embed a secret text in image pixels [27]. The researchers created an encoder-decoder-based training framework that recovers secrets if the found watermark was found to be untampered with. U-Net CNN architecture is used for the encoder and decoder network that uses 256×256 input images. In addition, a discriminator network is used to distinguish real and tampered images. Its robustness also extends to images and videos.

Nadimpalli and Rattani proposed a facial image watermarking technique for social networks that is impenetrable against adversarial watermark removal attacks [28]. The proposed U-Net CNN-based architecture consisted of five components: encoder, decoder, adversary network, discriminator, and critic network. FaceForensics++, CelebA, and the IMDB-Wiki dataset were used to train and evaluate the model.

The researchers in [29] proposed a wavelet transform-based reversible watermarking algorithm for the authentication of image content. A focus on avoiding tampering's influence on the wavelet domain was considered. Each block of the image (four pixels in each block) is transformed as a carrier signal for watermark information. Two different watermarks are used, encrypted, and scrambled using chaotic mapping with different keys. It increases security and embedding randomness. Different detection and screening levels improve accuracy. Both watermarks reconstruct category and feature information when restoring tampered areas.

Similarly, the researchers in [30] proposed a dual-watermarking scheme using DWT-SVD and two authentication layers. The watermarking scheme uses a non-pixelpping 8x8 pixel block with DWT transform on a 512x512 host image. This generates an authentication bit for tamper localization. The approach was tested and evaluated on the SIPI dataset.

The researchers in [31] introduced a bioinspired watermarking algorithm for retinal fundus images in computer-aided retinopathy diagnosis. It uses the Steered Hermite Transform (SHT) and SVD to create imperceptible watermarks, embed them into RGB fundus images, and encrypt them using the Jigsaw Transform for added security. The MESSIDOR-2 dataset was used to train and evaluate the model. In addition, four DL models were evaluated to detect the application of watermarks in the images. However, the proposed watermarking approach was resilient and remains undetected against DL models.

2.2 Tampering detection models

Hu et al. proposed a tampering identification model using watermarking and alpha mattes to detect forged images [32]. The proposed model calculated the alpha matte of the image using component hue-difference-based spectral matting and used a DWT-DCT-SVD-based watermarking technique to insert the watermarks. The detection method uses the difference in singular values to identify manipulated foreground and background photos by inserting two distinct watermarks based on derived alpha mattes. The technique reliably identifies forgeries produced by image matting or image in-painting, tampered foreground and background photos, and traded foreground and background images. An adaptive threshold was used that makes it suitable for real-world scenarios. Nguyen proposed a fragile model of image authentication that combines DCT, DWT, and SVD, and DCT for feature selection and quantization index modulation (QIM) for the generation and embedding of the authentication code [33]. The coefficients are calibrated using the Gram-Schmidt process to ensure the expected retrieval of the authentication code. The model is found to produce watermarked photos with great visual quality and excellent tamper detection accuracy.

The researchers in [34] proposed a visual cryptography-based watermarking scheme to detect forgeries in images. The approach was tested against various attacks such as noise, sharpening, median filtering, lossy compression, and geometric distortions. The authentication and forgery detection model proposed in [35] integrates vector quantization (VQ), absolute moment block truncation coding (AMBTC), and double matrix encoding to reduce distortions caused to the host due to watermarking. The authors used images from the BOWS dataset to conduct their experiments. The fragile and reversible watermarking model proposed in [36] generates two watermarked versions of the image. It provides complete recovery of the original image when no tampering is present. It ensures high-quality restoration if tampering is detected.

The researchers in [37] proposed a watermarking model to detect image manipulation using adaptive matrices and overcome the limitations of various previous methods. The embedding process uses a two-stage authentication mechanism with a 16-bit validation scheme to ensure precise forgery localization. This adaptive approach is designed to adapt the matrix pattern to the characteristics of the image. At the same time, the utilization of logistic sequencing enables the generation of non-periodic and non-convergent patterns. It improves authentication efficiency and accuracy. The approach was tested using the USC-SIPI dataset. An average PSNR value of 55.90 dB and an SSIM greater than 0.99 were observed.

The researchers in [38] proposed a watermarking technique using a complex wavelet transform of quaternion two trees for the extraction of features from images. The maximal entropy random walk algorithm identifies the best embedding area within the low-frequency subband. To detect and localize tampering, a watermark is created using the Swin transformer model and embedded in the chosen blocks. In addition, a dual scrambled image is encoded using SVD-derived principal component coefficients to authenticate the watermarked image before extraction. The semi-blind extraction method verifies authenticity by comparing the retrieved scrambled watermark with the original regenerated watermark. The approach was evaluated using three datasets, including HDR images, CASIA, and other datasets.

The researchers in [39] proposed an approach to image tampering and localization using the integer wavelet transform (IWT). IWT coefficients were used to generate face recovery information and embed it into cover-face images at the sender's side. Three algorithms are proposed, with IWT showing superior performance. The algorithm efficiently identifies altered blocks and restores unaltered versions to ensure originality, integrity, and security. The approach was tested on various images that differ in dimension and size.

A detailed review of existing studies shows that various traditional watermarking approaches use single-domain techniques, such as standalone DWT, DCT, or SVD. It leads to a failure to balance robustness, imperceptibility, and tampering localization. They usually rely on DWT. As a result, they are vulnerable to frequency-based attacks, such as filtering, compression, or noise addition. This occurs because of the limited stability of wavelet coefficients under such distortions. Similarly, even if DCT-based schemes benefit from energy compaction, they suffer from reduced spatial localization capabilities. Although SVD-based approaches are efficient in maintaining imperceptibility, they lack spatial sensitivity to detect local tampering effectively.

In addition, only a few existing frameworks focus mainly on watermark retrieval without offering a reliable mechanism to locate tampered regions. In particular, limited attention has been given to the idea of embedding auxiliary tamper detection structures within transform domains. They can include spatial markers or canvases that improve the visual identification of manipulations. The proposed approach addresses these issues effectively by proposing a hybrid watermarking-cum-tampering localization model.

3 Background materials

This section highlights the mathematical details of the transforms used: the Discrete Wavelet Transform (DWT), the DCT, and the SVD. They play an important role in various image processing tasks due to their ability to extract, analyze, and manipulate features in the transform domain.

3.1 Discrete wavelet transform

It has the ability to break up an image into its frequency components. It provides analysis at multiple resolutions. Unlike the traditional Fourier Transform, which only captures frequency content, DWT provides both spatial and frequency localization, making it particularly suitable for hierarchical image analysis. Among various wavelet families, the Haar wavelet is often preferred due to its simplicity and computational efficiency.

For two-dimensional (2D) images, the DWT applies a set of low-pass (L) and high-pass (H) filters along the rows and columns of the image, followed by downsampling. This operation results in four sub-band images, each representing different frequency and directional information.

Figure 3 shows the level-1 DWT decomposition of the standard 'cameraman's image using the Haar wavelet. The right image is the original, while the left image displays the four decomposed sub-bands: (a) LL, (b) LH, (c) HL, and (d) HH. The basic operations of the 2D Haar-DWT, specifically, horizontal and vertical filtering, are illustrated in Figure 4.

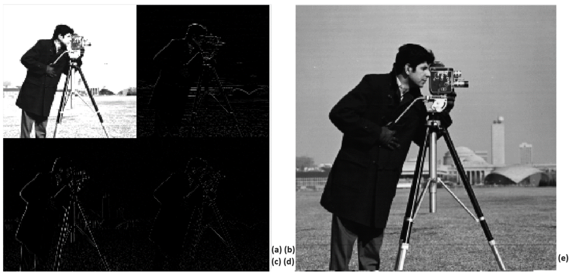


Figure 3: Level-1 DWT decomposition using Haar wavelet

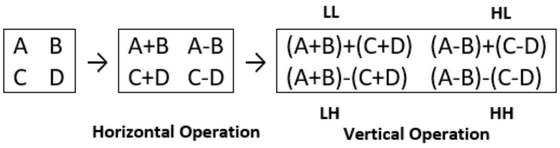


Figure 4: Level-1 Haar-DWT: Horizontal and vertical operations on a 2 × 2 image.

In watermarking applications, embedding data in the LL sub-band enhances robustness but can significantly degrade visual quality. Conversely, embedding in the HH sub-band is less perceptible but more vulnerable to attacks such as compression or filtering. Therefore, most DWT-based watermarking techniques favor the LH and HL sub-bands, striking a balance between imperceptibility and robustness. DWT's ability to provide secure, resilient, and imperceptible embedding makes it ideal for applications in copyright protection, image authentication, and tamper detection.

3.2 Discrete cosine transform

It is widely used in image compression and watermarking because of its energy-compaction property. It transforms spatial-domain image data into frequency-domain coefficients, making it easier to manipulate significant features while removing redundant information.

DCT segments an image into non-overlapping blocks (NOB) of size $B \times B$, where B depends on the image resolution and processing requirements. For each block, the forward DCT coefficients are computed using Equation (1).

$$FDCT(\gamma,\zeta) = \sum_{\alpha=0}^{B-1}\sum_{\beta=0}^{B-1} NOB(\alpha,\beta)\cdot Ker(\alpha,\beta,\gamma,\zeta) \tag{1}$$

Here, $NOB(\alpha,\beta)$ represents the pixel value at position (α,β), and the kernel function $Ker(\alpha,\beta,\gamma,\zeta)$ is evaluated using Equation (2).

$$\begin{aligned} Ker(\alpha,\beta,\gamma,\zeta) = \omega(\gamma)\cdot\omega(\zeta)\cos\left(\frac{(2\alpha+1)\gamma\pi}{2B}\right) \\ \cdot\cos\left(\frac{(2\beta+1)\zeta\pi}{2B}\right) \end{aligned} \tag{2}$$

The scaling factor $\omega(z)$ is calculated using Equation (3)

$$\omega(z) = \begin{cases} \sqrt{\frac{1}{B}} & \text{if } z = 0 \\ \sqrt{\frac{2}{B}} & \text{if } z > 0 \end{cases} \tag{3}$$

To reconstruct the image block from its frequency-domain coefficients, the inverse DCT (IDCT) is applied using Equation (4).

$$RDCT(\alpha,\beta) = \sum_{\gamma=0}^{B-1}\sum_{\zeta=0}^{B-1} FDCT(\gamma,\zeta)\cdot Ker(\alpha,\beta,\gamma,\zeta) \tag{4}$$

Due to its ability to preserve important visual details while minimizing perceptible distortions, DCT is an effective tool for embedding watermarks in mid-frequency coefficients, where changes are less noticeable yet are resistant to compression attacks.

3.3 Singular value decomposition

It is a linear algebra technique that is used in image processing for tasks such as compression, noise reduction, and watermarking. It decomposes a matrix A of size $m \times n$ into three matrices: U, S, and V^T as defined in Equation (5).

$$A_{m\times n} \rightarrow U_{m\times m}S_{m\times n}V_{n\times n}^{T} \tag{5}$$

Here, U is an orthogonal matrix of size $m \times m$ that contains the left singular vectors, S is an diagonal matrix of size $m \times n$ containing singular values that represent the strength of each corresponding vector, and V^T is the transpose of an orthogonal matrix of size $n \times n$ that contains the right singular vectors.

SVD identifies the directions in which the data are most variable and separates the signal from the noise. The original matrix A can be reconstructed from its decomposed components using the inverse operation using Equation (6).

$$A_{m\times n} \leftarrow U_{m\times m}S_{m\times n}V_{n\times n}^{T} \tag{6}$$

In watermarking, SVD is particularly useful because slight changes in singular values do not significantly affect image quality. Embedding watermark data in singular values ensures SIR, making SVD a preferred technique in high-performance watermarking systems.

4 Proposed framework

The proposed pipeline for digital image content authentication and tampering detection framework consists of the following phases:

1. Embedding of dual watermarks and the tamper detection mask into the cover image.
2. Tampering and Attack Simulations.
3. Retrieval of embedded watermarks to verify the integrity and authenticity of the image.
4. Extraction of the tamper detection mask to detect and locate tampered regions.

Each of the phases is designed to systematically address the requirements of secure watermarking and tampering detection, ensuring the integrity and authenticity of the cover image.

4.1 Watermark embedding phase

As presented in Figure 5, the embedding process begins by separately processing two components of the watermark: the signature (W_s) and the fingerprint (W_f). Each watermark undergoes a single-level DWT, yielding four subbands: LL, LH, HL, and HH. The DWT of the signature is represented using Equation (7).

$$W_s \xrightarrow{\text{DWT}} \{LL_s, LH_s, HL_s, HH_s\} \tag{7}$$

The HH_s sub-band is selected for further processing due to its high-frequency detail preservation and imperceptibility. A DCT is applied to HH_s using Equation (8).

$$HH_s \xrightarrow{\text{DCT}} DCT(HH_s) \tag{8}$$

Next, SVD is performed on the transformed sub-band and computed using Equation (9).

$$DCT(HH_s) = U_s S_s V_s^T \tag{9}$$

The same operations are applied to the fingerprint (W_f) and the cover image (I) to obtain their respective singular values using Equation (11).

$$W_f \xrightarrow{\text{DWT}\to\text{DCT}} DCT(HH_f) = U_f S_f V_f^T \tag{10}$$

$$I \xrightarrow{\text{DWT}\to\text{DCT}} DCT(HH_I) = U_I S_I V_I^T \tag{11}$$

Next, the singular values of the signature are embedded into those of the fingerprint using Equation (12).

$$S'_f = S_f + \alpha S_s \tag{12}$$

The transformed singular values of the watermarks are embedded into the singular value (S_I) of the cover image using Equation (13).

$$S'_I = S_I + \beta S'_f \tag{13}$$

Here, α and β are scaling factors controlling the strength of embedding. The inverse SVD, DCT, and DWT operations are then applied sequentially to obtain the partially watermarked image (I_{pw}) using Equation (14).

$$I_{pw} = \text{IDWT}\left(\text{IDCT}\left(U_I S'_I V_I^T\right)\right) \tag{14}$$

To enhance tamper localization, a tamper detection mask (T) is embedded in the LH2 of HL1 sub-bands of the MSB plane using a two-level DWT decomposition using Equations (15) and (16).

$$\begin{aligned} I_{pw} \xrightarrow{\text{2-level DWT}} \{&LL1, LH1, \\ &\{LL2, LH2, HL2, HH2\}, HH1\} \end{aligned} \tag{15}$$

$$HL1'LH2' = HL1LH2 + T \tag{16}$$

Applying inverse DWT yields the final watermarked image I_w using Equation (17).

$$\begin{aligned} I_w = \text{IDWT}(\{&LL1, LH1, \\ &\{LL2, LH2, HL1', LH2', HH2\}, HH1\}) \end{aligned} \tag{17}$$

This ensures a robust and imperceptible watermark with improved tamper localization.

4.2 Tampering and attack simulations

In this phase, the watermarked image is subjected to various types of tamperings such as cloning, splicing, retouching, and resampling, followed by common signal processing attacks such as cropping, compression, noise addition, blurring, and geometric distortions. This step plays a crucial role in evaluating the reliability and robustness of the proposed model under both malicious and nonmalicious modifications/alterations.

4.3 Watermark retrieval phase

The watermark retrieval process as shown in Figure 7 begins by applying DWT and DCT to the HH sub-band of the received watermarked image (I_w), followed by SVD using Equation (18).

$$I_w \xrightarrow{\text{DWT}\to\text{DCT}\to\text{SVD}} DCT(HH_w) = U'_I S'_I V_I'^T \tag{18}$$

The transformed watermark is recovered by subtracting the original singular values of the cover image using Equation (19).

$$S_s = \frac{S'_f - S_f}{\alpha} \tag{19}$$

Where,

$$S'_f = \frac{S'_I - S_I}{\beta}$$

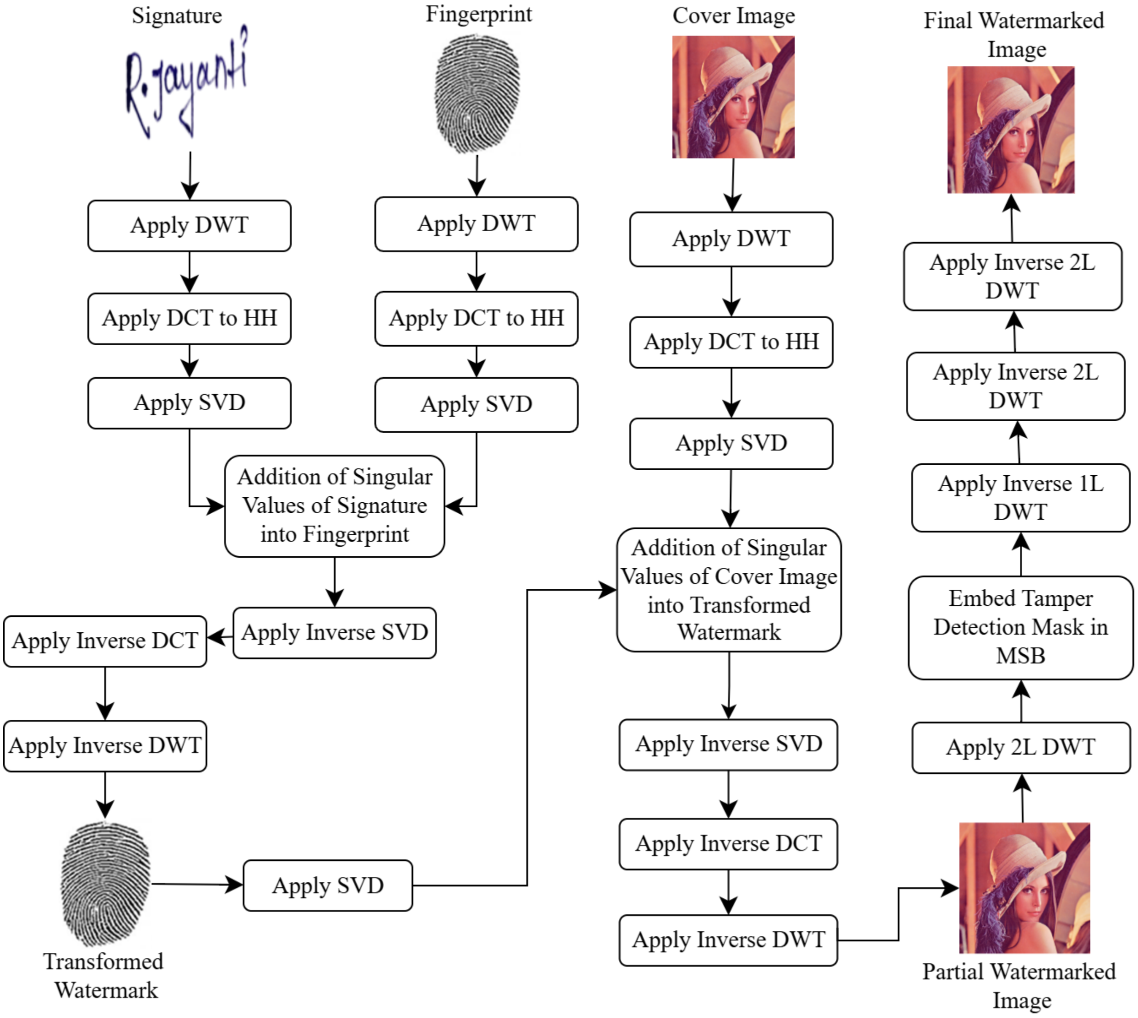


Figure 5: Proposed watermark embedding model.

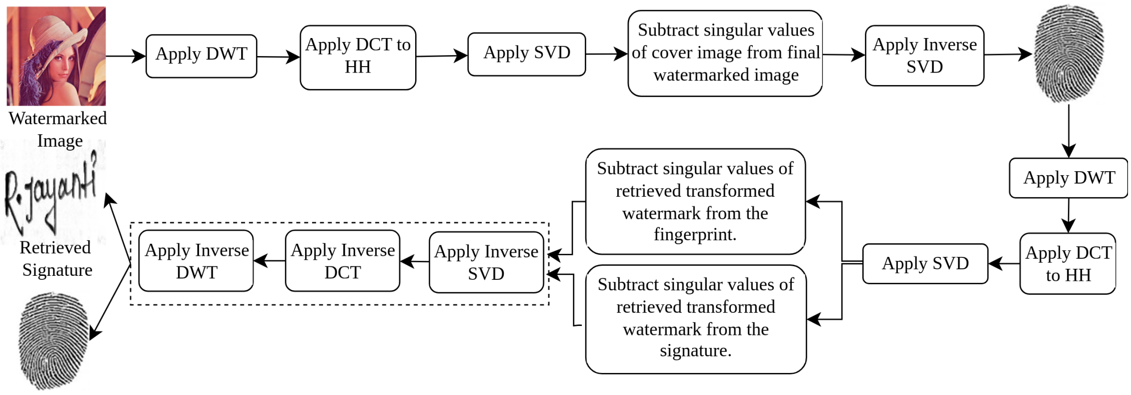

Watermarked Image
Apply DWT
Apply DCT to HH
Apply SVD
Subtract singular values of cover image from final watermarked image
Apply Inverse SVD
Apply DWT
Apply DCT to HH
Apply SVD
Subtract singular values of retrieved transformed watermark from the fingerprint.
Subtract singular values of retrieved transformed watermark from the signature.
Apply Inverse SVD
Apply Inverse DCT
Apply Inverse DWT
Retrieved Signature

Figure 6: Flow of the watermark extraction procedure

The inverse SVD, DCT, and DWT steps reconstruct the fingerprint and signature in the spatial domain are computed using Equations (20) and (21).

$$W_f' = \text{IDWT}(\text{IDCT}(U_f S_f' V_f^T)) \tag{20}$$

$$W_s' = \text{IDWT}(\text{IDCT}(U_s S_s V_s^T)) \tag{21}$$

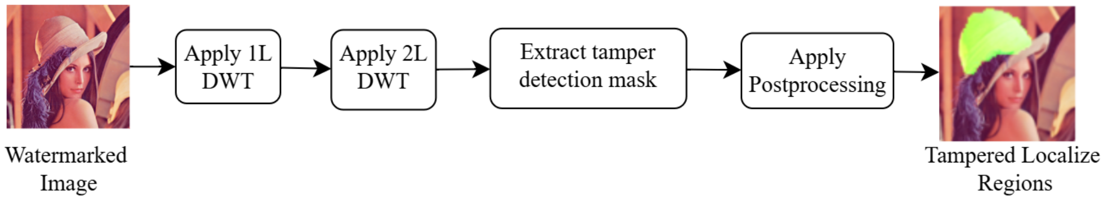


Figure 7: Overall steps of the proposed tampering detection model

4.4 Tampering detection phase

As shown in Figure 7, tampering detection starts with applying a two-level DWT to the final watermarked image using Equation (22).

$$I_w \xrightarrow{\text{2-level DWT}} \{LL1, LH1, \{LL2, LH2, HL2, HH2\}, HH1\} \quad (22)$$

The tamper detection mask T' is extracted from the LH2 of HL1 sub-bands of the MSB using Equation (23).

$$T' = \text{Extract}(LH2) \quad (23)$$

To detect manipulated regions, the difference between the original and extracted tamper mask is evaluated using Equation (24).

$$\Delta T = |T - T'| \quad (24)$$

Any significant non-zero values in ΔT are flagged as tampered pixels. It provides accurate localization of altered regions in the image. Further morphological operations, i.e., opening and closing, are applied to the extracted tamper mask to improve the accuracy of the localization.

Together, these three phases ensure secure and layered watermarking. It supports robust authentication and provides fine-grained localization of tampered regions.

5 Analysis and discussion of the findings

This section presents a comprehensive evaluation of the proposed method through a detailed experimental analysis. Subsection 5.1 describes the experimental setup. Subsection 5.2 describes the details of the dataset. Section 5.3 describes various evaluation metrics used to assess the robustness and reliability of the approach. Subsection 5.4 presents an in-depth discussion of both qualitative and quantitative findings from the results.

5.1 Experimental setup

The performance of the suggested approach is evaluated through various experiments. They were conducted on a 64-bit Windows 11 Pro system (version 23H2) using an Intel Core i7-10700 CPU running at 2.90 GHz with 8 GB of RAM.

In the experimental phase, two distinct watermarks—a signature watermark and a fingerprint watermark—were embedded in digital images along with a tamper detection mask to evaluate the proposed authentication framework. Signature and fingerprint watermarks serve the purpose of verifying the authenticity and ownership of the image content. The tamper detection mask helps to localize the exact regions where manipulations have occurred. These components were embedded using a combination of DWT, DCT, and SVD. It ensures robustness against common image processing attacks. During the verification stage, both watermarks were extracted and matched to the original references to confirm the authenticity. The tamper detection mask can detect any mismatch to pinpoint the manipulated areas.

The following techniques were used during the experiments to evaluate the efficacy of the watermark and tamper detection model:

- Image manipulation operations: Retouching, resampling, splicing, and cloning.
- Noise attacks: Gaussian, salt&pepper, and speckle.
- Image enhancement procedures: Filtering, sharpening, histogram equalization, blurring, and lossy compression.

To quantify the performance, various metrics including, PSNR, SSIM, NCC, and detection accuracy, were used. The results show the efficacy of the proposed approach in both authenticating the image and locating tampered regions.

5.2 Dataset details

The USC-SIPI dataset [40] is used for experimentation. It is organized into volumes according to the image size. It includes grayscale images at 8 bits per pixel (bpp) and color images at 24 bpp. For simulation of experiments, six commonly used images were selected from the dataset: splash, baboon, Lena, sailboat, pepper, and house. As shown in Figure 8, all selected images are 512×512 pixels in size and are 24-bit color images. The watermarks used in the experiments, including the signature and fingerprint, were provided by the authors. The signature, fingerprint, and tamper detection mask are monochrome images of 64 × 64 bits, 128 × 128 bits, and 32 × 32 bits, respectively.

Figure 8: Illustration of test images from the SIPI image database

5.3 Performance metrics

The effectiveness of the proposed model is evaluated using a set of well-established performance metrics. To assess the robustness of the image watermarking process, metrics such as PSNR, SSIM, Mean Squared Error (MSE), and NCC are employed. To evaluate the accuracy of tampering or forgery detection, various classification metrics are used. These indicators provide a comprehensive understanding of both the preservation of visual quality and the detection capability of the proposed framework.

5.4 Result analysis

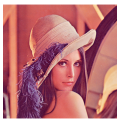
Table 1 presents the quantitative evaluation of the imperceptibility and robustness of the watermark using three widely adopted image quality metrics: PSNR, NCC, and SSIM. These metrics were computed for six standard test images after embedding of the watermark. The PSNR values for all watermarked images range from 38.926dB to 40.322dB. This indicates a high visual quality with minimal perceptible distortion. The NCC values, all exceeding 99.8%, confirm a strong correlation between the original and watermarked images. This signifies excellent robustness and successful preservation of the watermark. Similarly, the SSIM values, ranging from 99.056% to 99.713%, demonstrate that the structural content and perceptual quality of the images remain intact after embedding. These results validate that the proposed scheme maintains a high level of imperceptibility and fidelity across diverse image content. It makes the scheme suitable for secure image authentication applications.

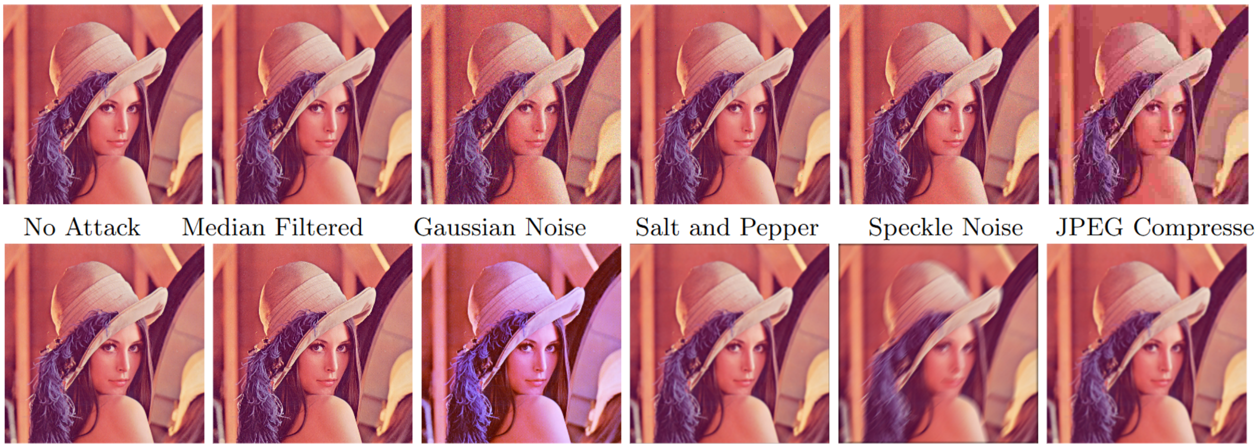
To evaluate the robustness of the proposed watermarking method, various common image processing attacks were applied to the watermarked Lena image. Figure 9 demonstrates the visual impact of these attacks. The applied distortions include median filtering, Gaussian noise, salt&pepper noise, speckle noise, JPEG and JPEG2000 compression, sharpening, histogram equalization, averaging, motion blur, and Gaussian low-pass filtering. These attacks simulate real-world distortions and compression scenarios commonly encountered in digital image transmission and storage. Figure 10 and Figure 11 show the signatures and fingerprints extracted from the attacked images. Despite various distortions, the watermarks remain visually identifiable, demonstrating the model's resilience in preserving embedded watermark information. Although some degradation is visible under severe attacks, such as Gaussian noise and quantization, the embedded signature remains recognizable. This validates the robustness of the hybrid watermarking scheme. The qualitative analysis confirms that the proposed DWT-DCT-SVD-based embedding strategy effectively preserves watermark integrity under typical attack scenarios. This supports the robustness and applicability of the model in real-world multimedia authentication tasks.

Figure 12 presents the quantitative analysis of the proposed watermarking method under various image processing attacks using three standard performance metrics: SSIM, PSNR, and NCC. These metrics assess visual quality, distortion level, and correlation between original and watermarked images with different attacks. As shown in Figure 12, without attack, the SSIM, PSNR, and NCC values are the highest. This indicates perfect preservation of visual quality and watermark integrity. Most attacks, including median filtering, JPEG compression, and sharpening, maintain high SSIM and NCC values (> 98%). This shows strong resistance to common distortions. PSNR values remain above 30 dB in these cases. This ensures minimal perceptual degradation. However, some attacks, including Gaussian noise, histogram quantization, and motion blur, led to a slight reduction in few metrics. PSNR of ∼25 dB and SSIM < 90% were observed. It shows more noticeable distortions. Despite this, the NCC values remain relatively high (above 96% in most cases), demonstrating the robustness of the watermark against noise and filtering attacks. This analysis validates the effectiveness of the proposed scheme in maintaining the fidelity and perceptual quality of the watermark in both mild and moderate attack scenarios.

Figures 14, 13, and 15 illustrate the impact of different values of the scaling parameter α on SSIM, PSNR, and NCC in various common image attacks. These results reveal a significant trade-off between imperceptibility of watermarks and robustness. As α increases, the robustness of the watermark improves. It can be observed by the consistently high NCC values in attacks, particularly for salt&pepper, Gaussian noise, and JPEG compression. However, this affects visible image quality. In Figure 14, higher values α result in reduced PSNR. This indicates a large distortion in the watermarked image. Similarly, the SSIM values in Figure 13 show a downward trend with increasing α, especially for high-distortion attacks such as motion blur and histogram quantization. This highlights

Table 1: PSNR, NCC, and SSIM of the watermarked images

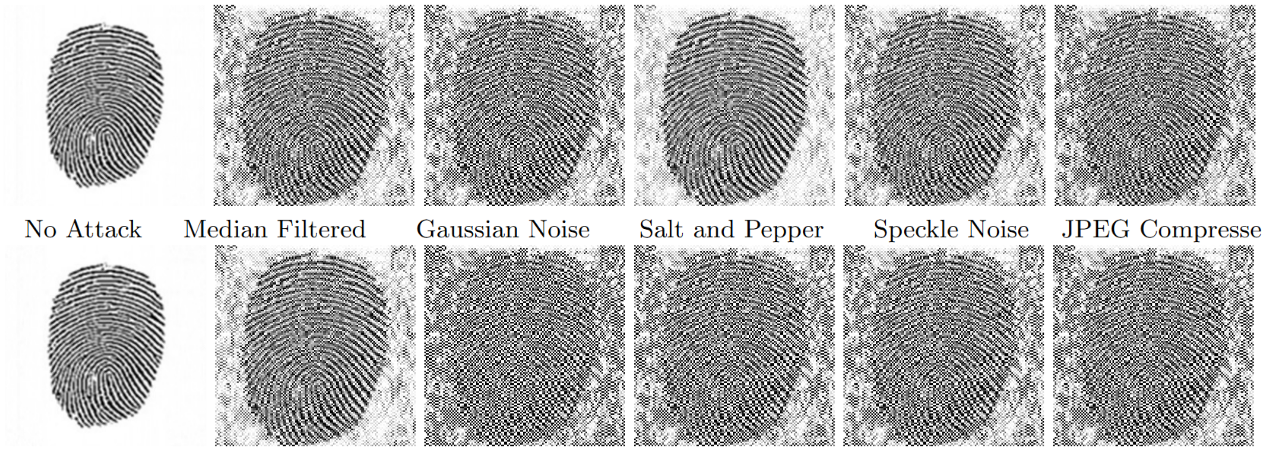
Watermarked Image						
PSNR	40.322	38.926	40.014	39.703	40.133	40.130
NCC	99.944	99.868	99.907	99.925	99.923	99.895
SSIM	99.356	99.449	99.658	99.056	99.713	99.641



No Attack Median Filtered Gaussian Noise Salt and Pepper Speckle Noise JPEG Compressed

JPEG2000 Compressed Sharpened Histogram Quantized Average Filter Motion Blur Gaussian Lowpass

Figure 9: Different attacks were applied to the watermarked Lena image



No Attack Median Filtered Gaussian Noise Salt and Pepper Speckle Noise JPEG Compressed

JPEG2000 Compressed Sharpened Histogram Quantized Average Filter Motion Blur Gaussian Lowpass

Figure 10: Extracted fingerprints from attacked images

the role of α watermarking systems. A smaller one α maintains high visual fidelity but offers less robustness against attacks. On the other hand, the increase in α improves resistance to tampering but reduces the image quality. Therefore, selecting an optimal α value is essential to balance robustness and imperceptibility in real-world applications.

Tables 2 and 3 demonstrate the robustness and imperceptibility of the proposed method under two common image processing attacks: JPEG compression and Gaussian lowpass filtering. The visual results of extracted/retrieved signatures and fingerprints, along with PSNR, NCC, and SSIM metrics, are used to evaluate the performance. As presented in Table 2, the watermarking algorithm shows strong resilience to JPEG compression, even at low quality factors (Q). The NCC remains above 97% in all compression levels, which indicates an accurate retrieval of the embedded watermark. SSIM and PSNR improve steadily as the qual-

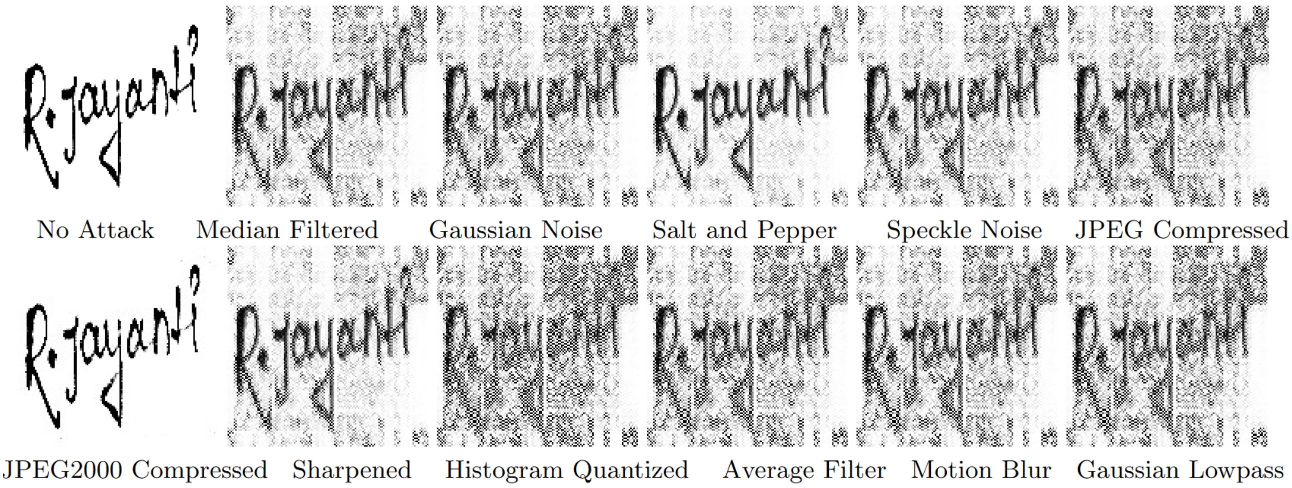


Figure 11: Extracted signatures from attacked watermarked images

Table 2: Watermark extraction from JPEG compressed watermarked images

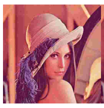
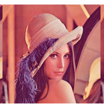
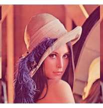
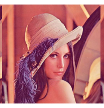
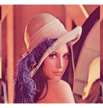
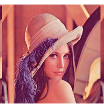
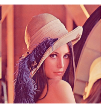
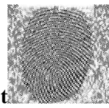
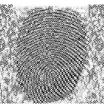
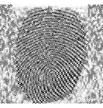
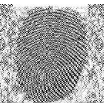
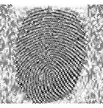
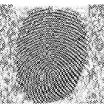
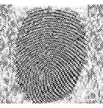
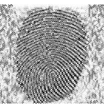
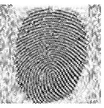
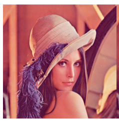
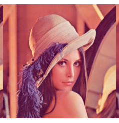
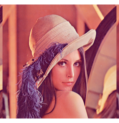
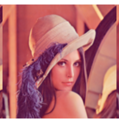
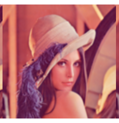
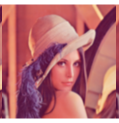
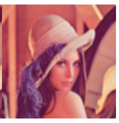
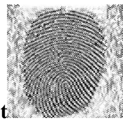
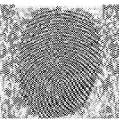
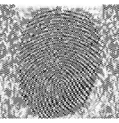
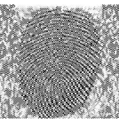
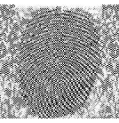
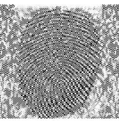
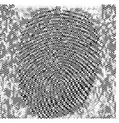
Quality factor (Q)	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**
Watermarked image									
Extracted signature									
Extracted fingerprint									
PSNR	27.531	29.832	30.909	31.537	32.014	32.410	32.907	33.604	34.748
NCC	97.537	99.030	99.243	99.344	99.412	99.463	99.521	99.592	99.687
SSIM	95.931	97.378	97.885	98.133	98.311	98.444	98.601	98.789	99.048

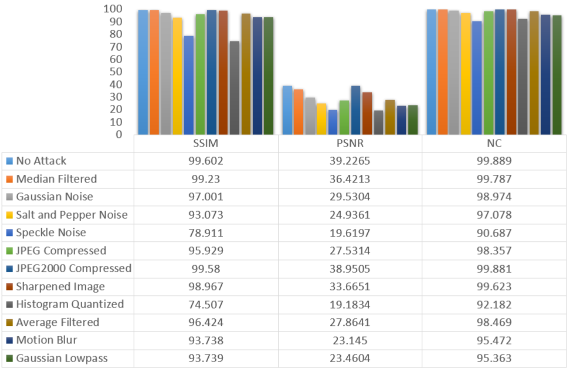
Table 3: Retrieval accuracy after Gaussian low-pass filtering with different sigma values

Sigma	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
Watermarked Image							
Retrieved Signature							
Retrieved Fingerprint							
PSNR	39.592	32.513	29.989	28.4333	27.330	26.480	25.788
NCC	99.898	99.481	99.070	98.667	98.280	97.907	97.544
SSIM	99.683	98.515	97.463	96.571	95.848	95.262	94.777

ity factor increases. This reflects better visual quality and less distortion in the watermarked image. In particular, even at a low Q=10, the watermark remains recoverable, although with slightly reduced visual fidelity. In Table 3, Gaussian low-pass filtering with increasing standard deviation (σ) is applied to assess the robustness of the watermark

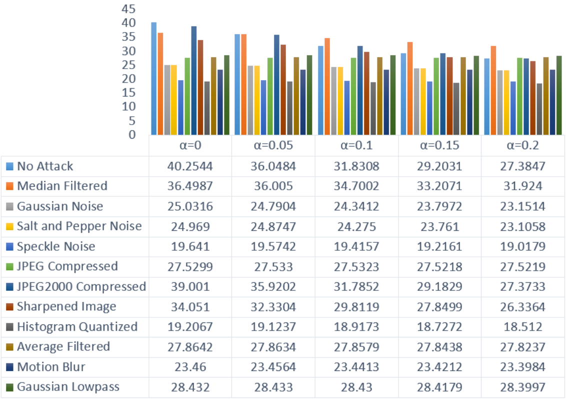
Table 4: PSNR, NCC and SSIM values obtained by the model for six tampered images

Tampered Images						
PSNR	11.200	25.209	27.531	20.126	21.907	16.325
NCC	54.491	97.140	98.357	93.170	95.128	77.110
SSIM	57.973	98.360	95.931	92.081	95.658	82.182



	SSIM	PSNR	NC
■ No Attack	99.602	39.2265	99.889
■ Median Filtered	99.23	36.4213	99.787
■ Gaussian Noise	97.001	29.5304	98.974
■ Salt and Pepper Noise	93.073	24.9361	97.078
■ Speckle Noise	78.911	19.6197	90.687
■ JPEG Compressed	95.929	27.5314	98.357
■ JPEG2000 Compressed	99.58	38.9505	99.881
■ Sharpened Image	98.967	33.6651	99.623
■ Histogram Quantized	74.507	19.1834	92.182
■ Average Filtered	96.424	27.8641	98.469
■ Motion Blur	93.738	23.145	95.472
■ Gaussian Lowpass	93.739	23.4604	95.363

Figure 12: SSIM, PSNR, and NCC of attacked watermarked images



	α=0	α=0.05	α=0.1	α=0.15	α=0.2
■ No Attack	40.2544	36.0484	31.8308	29.2031	27.3847
■ Median Filtered	36.4987	36.005	34.7002	33.2071	31.924
■ Gaussian Noise	25.0316	24.7904	24.3412	23.7972	23.1514
■ Salt and Pepper Noise	24.969	24.8747	24.275	23.761	23.1058
■ Speckle Noise	19.641	19.5742	19.4157	19.2161	19.0179
■ JPEG Compressed	27.5299	27.533	27.5323	27.5218	27.5219
■ JPEG2000 Compressed	39.001	35.9202	31.7852	29.1829	27.3733
■ Sharpened Image	34.051	32.3304	29.8119	27.8499	26.3364
■ Histogram Quantized	19.2067	19.1237	18.9173	18.7272	18.512
■ Average Filtered	27.8642	27.8634	27.8579	27.8438	27.8237
■ Motion Blur	23.46	23.4564	23.4413	23.4212	23.3984
■ Gaussian Lowpass	28.432	28.433	28.43	28.4179	28.3997

Figure 13: PSNR of different attacked watermarked images with varying alpha values

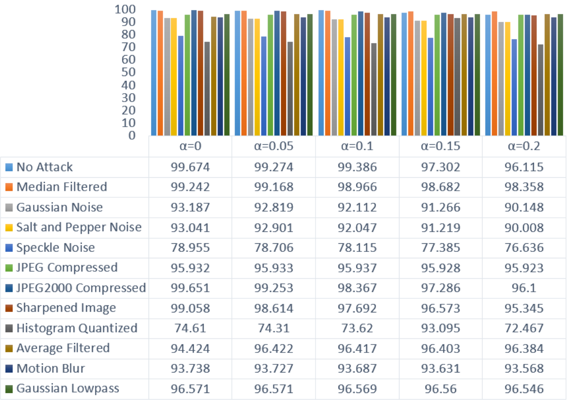
under blurring conditions. As σ increases from 0.5 to 3.5, PSNR and SSIM values decrease. This indicates an increasing distortion and reduced perceptual quality. However, the NCC remains relatively high (above 97%) throughout the study. This shows that the fingerprint can still be accurately retrieved even with significant smoothing. In addition, the robustness of the method is demonstrated against blurring attacks, although some degradation in image quality is unavoidable with stronger filtering. Both tables validate the trade-off between imperceptibility and robustness. The watermarking system maintains high fidelity in light to moderate attacks. It still retrieves the watermark with high accuracy, even under severe distortions. These findings support the algorithm's suitability for practical use where resilience to real-world manipulations is essential.

Table 4 presents the values of different performance metrics for six tampered images used to assess the robustness of the proposed scheme against various image tampering attacks. It includes JPEG compression (QF = 10), rotation by 90° clockwise, retouching, cropping, resampling cloning, retouching cloning, retouching splicing, resampling splicing, and retouching splicing. These scenarios helps to test both visual distortion and the effectiveness of watermark retrieval in manipulated scenarios. The results show significant variability in performance for different types of tampering. The PSNR values range from as low as 11.200 to 27.531. This shows different degrees of visual degradation due to tampering. The lower PSNR (e.g., 11.200) shows severe distortion, as seen in the first image. On the other hand, higher values suggest minor perceptual changes. Despite such variations, the NCC remains relatively strong in most cases. It was observed to be greater than 90% in four of the six altered images. This indicates that the model is capable of retrieving the embedded watermark with high accuracy, even with unwanted modifications. The lowest PSNR observed was 54.491. This shows the correlation between high distortion and reduced retrieval fidelity. SSIM values show that the image structure is better preserved in less tampered images (e.g., 98.360, 95.931) and more affected in heavily tampered ones (e.g., 57.973). This table shows the flexibility of the proposed watermarking system. Even in cases of aggressive tampering, the model achieves high performance. It demonstrates its applicability in real-world content authentication scenarios.

Figures 16 and 17 present the performance of the proposed method against various image tampering operations. Each row in the figures corresponds to a distinct tampering technique. Figure 16 presents a visual overview of the extraction of the watermark and the location of the tamper. For each tampered image, four representations are shown: (i) the forged watermarked image, (ii) the retrieved watermark signature, (iii) the extracted fingerprint watermark, and (iv) the localized tampered region. The watermarks and tamper localization were observed to stay identifiable and accurate despite major modifications. It shows the flexibility of the model for forgeries. Figure 17 further validates this by comparing predicted tamper regions with ground truth masks. Each row presents (i) the forged image, (ii)

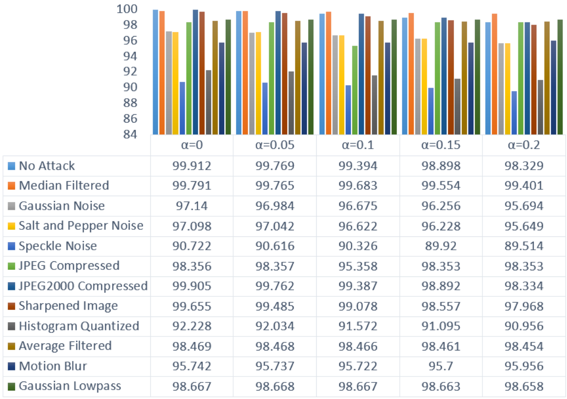
Table 5: Performance analysis of the proposed tamper detection method under various forgery types

Tampered	TP	FP	FN	TN	Precision	Recall	Accuracy	F1 Score
retouched	28274	974	2663	230233	0.9667	0.9139	0.9861	0.9396
cloning with resampling	23075	6173	761	232135	0.7889	0.9681	0.9735	0.8694
cloning with retouching	18676	3148	6472	233848	0.8558	0.7426	0.9633	0.7952
splicing with resampling	33006	2002	1252	225884	0.9428	0.9635	0.9876	0.9530
splicing with retouching	14997	2219	3937	240991	0.8711	0.7921	0.9765	0.8297



	α=0	α=0.05	α=0.1	α=0.15	α=0.2
■ No Attack	99.674	99.274	99.386	97.302	96.115
■ Median Filtered	99.242	99.168	98.966	98.682	98.358
■ Gaussian Noise	93.187	92.819	92.112	91.266	90.148
■ Salt and Pepper Noise	93.041	92.901	92.047	91.219	90.008
■ Speckle Noise	78.955	78.706	78.115	77.385	76.636
■ JPEG Compressed	95.932	95.933	95.937	95.928	95.923
■ JPEG2000 Compressed	99.651	99.253	98.367	97.286	96.1
■ Sharpened Image	99.058	98.614	97.692	96.573	95.345
■ Histogram Quantized	74.61	74.31	73.62	93.095	72.467
■ Average Filtered	94.424	96.422	96.417	96.403	96.384
■ Motion Blur	93.738	93.727	93.687	93.631	93.568
■ Gaussian Lowpass	96.571	96.571	96.569	96.56	96.546

Figure 14: SSIMs of different attacked watermarked images with varying alpha values



	α=0	α=0.05	α=0.1	α=0.15	α=0.2
■ No Attack	99.912	99.769	99.394	98.898	98.329
■ Median Filtered	99.791	99.765	99.683	99.554	99.401
■ Gaussian Noise	97.14	96.984	96.675	96.256	95.694
■ Salt and Pepper Noise	97.098	97.042	96.622	96.228	95.649
■ Speckle Noise	90.722	90.616	90.326	89.92	89.514
■ JPEG Compressed	98.356	98.357	95.358	98.353	98.353
■ JPEG2000 Compressed	99.905	99.762	99.387	98.892	98.334
■ Sharpened Image	99.655	99.485	99.078	98.557	97.968
■ Histogram Quantized	92.228	92.034	91.572	91.095	90.956
■ Average Filtered	98.469	98.468	98.466	98.461	98.454
■ Motion Blur	95.742	95.737	95.722	95.7	95.956
■ Gaussian Lowpass	98.667	98.668	98.667	98.663	98.658

Figure 15: NCC of different attacked watermarked images with varying alpha values

the ground truth tamper mask, and (iii) the tamper map predicted by the proposed method. The close visual match between ground truth and predicted masks shows the reliability of the system in detecting the exact location of the tampering.

Table 5 presents a detailed evaluation of the proposed approach in eight types of image forgery. The details of the correct predictions and errors, along with various performance metrics, are presented. The results show near-ideal performance on different geometric and compression-based attacks.

This indicates that the system is highly effective in detecting these types of manipulations without introducing error. The detection performance was observed to decrease slightly when complex tampering techniques were combined with resampling or retouching. Lower precision and recall values are observed, especially for retouched cloning and retouched splicing. The F1 scores for both scenarios were 0.7952 and 0.8297, respectively. These drops suggest that although the system presented reliable detection, this decrease suggests that these forgeries pose greater challenges to the system. It is due to their fine alterations and blending effects. The method maintains high overall accuracy and consistent F1 scores > 0.79 in all simulated attacks. This illustrates the robustness of the proposed approach for real-world tamper localization and forgery detection tasks.

Tables 6 and 7 present a comprehensive comparative analysis of the proposed method with existing state-of-the-art techniques. As observed in Table 6, the proposed method uses a cover image size of 512 × 512 and a dual watermark size of 128 × 128 and 64 × 64. It can be seen from the tables that the existing methods either lack forgery detection altogether or have not been tested under attack situations. On the other hand, the proposed model successfully authenticates and detect forgeries in various attack situations. Table 7 further evaluates the robustness of the proposed method against various image processing attacks. Consistent high NC values (above 0.92 for all attacks), high PSNR (18.384–39.2265), and SSIM values (mostly above 0.98) were achieved. This approach offers better structural consistency under various distortions compared to existing methods (e.g., [24] and [25]). It performs especially well under Gaussian, median, and speckle noise. In most cases, it achieves higher SSIM and NC values. This indicates better accuracy in watermark retrieval and preservation of visual quality, making it a reliable choice for authentication and tamper detection tasks.

In addition to the above results, Table 8 provides detailed responses to the RQs, formulated to evaluate the effectiveness and novelty of the proposed watermark and tamper detection method.

6 Concluding remarks

This study presents a robust watermark-based framework for digital image authentication and tamper detection. It uses three transformations for robust watermark embed-

Table 6: Comparison of the proposed method with existing approaches in terms of authentication, forgery detection, PSNR, SSIM, types of forgery detected, and robustness under various attacks.

Ref	Authentication	Forgery Detection	PSNR	SSIM	Types of forgery detected	Forgeries detected under types of attacks
[24]	✔	✖	39.45	0.9964	–	–
[25]	✔	✖	31.49	0.8618	–	–
[29]	✔	✔	43.79	–	Only copy-move	–
[30]	✔	✔	46.16	0.9988	Copy-move and splicing	–
[35]	✔	✔	48.21	–	Cropping and copy-move	–
[36]	✔	✔	46.36	0.9963	Cropping up to 50%	–
[37]	✔	✔	55.92	0.9999	Resampling, retouching, copy-move, and splicing	–
Proposed	✔	✔	39.23	0.9960	Copy-move, splicing, retouching, resampling, and cropping	Salt & Pepper Noise,Gaussian Noise,Gaussian Filter,Median Filter,Sharpening,Motion Blur,Average Filter,Histogram Equalization ,JPEG,JPEG2000

Table 7: Comparison of NC, PSNR, and SSIM of the proposed method for various attacks with existing techniques

Attack Type	NC			PSNR			SSIM		
	[24]	[25]	Proposed	[24]	[25]	Proposed	[24]	[25]	Proposed
Salt & Pepper Noise	1	0.9943	0.9708	–	26.7142	24.9361	–	0.931	0.970
Gaussian Noise	0.9988	0.9938	0.9897	–	31.4845	29.5304	–	0.970	0.970
Speckle Noise	0.9999	0.9814	0.9069	–	22.4363	19.6197	–	0.789	0.891
Gaussian Filter	0.9995	0.9995	0.9536	–	26.8799	23.4604	–	0.937	0.939
Median Filter	0.9999	0.9975	0.9979	–	31.6371	36.4213	–	0.992	0.992
Sharpening	0.999	–	0.9962	–	–	33.6651	–	–	0.990
Motion Blur	0.997	–	0.9547	–	–	23.1454	–	–	0.937
Average Filter	0.999	0.9976	0.9847	–	26.7821	27.8641	–	0.964	0.964
Histogram Equalization	0.985	0.9476	0.9218	–	10.3095	19.1834	–	0.906	0.942
JPEG	0.999	–	0.9836	–	–	27.5314	–	–	0.959
JPEG2000	0.999	–	0.9988	–	–	38.9505	–	–	0.996

Table 8: Comprehensive answers to the RQs

RQ No	Answer
RQ1	DWT analyzes spatial-frequency features for localized detection. On the other hand, DCT compresses energy into fewer coefficients. This improves resilience to compression. SVD provides stability against noise, rotation, and scaling that preserves image integrity. In addition, a tamper detection mask is inserted to localize the tampered section. Together, these methods provide reliable image content integrity verification and forgery localization.
RQ2	The suggested model resists attacks with high PSNR, SSIM, and NCC. SSIM and NCC values of $\sim$100% in watermarked images. It highlights the watermarking's durability and the efficiency of the detection model even in noisy environments.
RQ3	Even after different types of attacks on the watermarked photos, the results provided in figures 16 and 17 indicate the effectiveness of the proposed model.

ding: DWT, DCT, and SVD. The model embeds both fingerprint and signature watermarks. It also uses a tamper detection mask in the LH2 sub-band of the LH1 band for accurate localization of manipulated regions. The experimental results show high imperceptibility. PSNR, SSIM, and NCC values greater than 38 dB, 99%, 99.8% were received, respectively. This shows minimal visual distortion and strong watermark fidelity. The system remains highly robust under various attacks and maintains the integrity of the watermark. High F1 scores of 0.9530 and various other near-ideal metrics were received for tamper detection even in complex forgery situations. Various other visualizations further signify the efficacy of the proposed scheme in preserving watermark information for different manipulations. Although the proposed model successfully localized manipulations for various attacks, it faces challenges in dealing with rotation and compression. In addition, the requirement of original watermark data during verification limits the model's applicability in blind authentication scenarios. Future research should focus on blind watermarking techniques and deep learning-based methods to locate tampered regions. This provides improved automation and adaptability. Integration of emerging technologies, including blockchain and artificial intelligence, could fur-

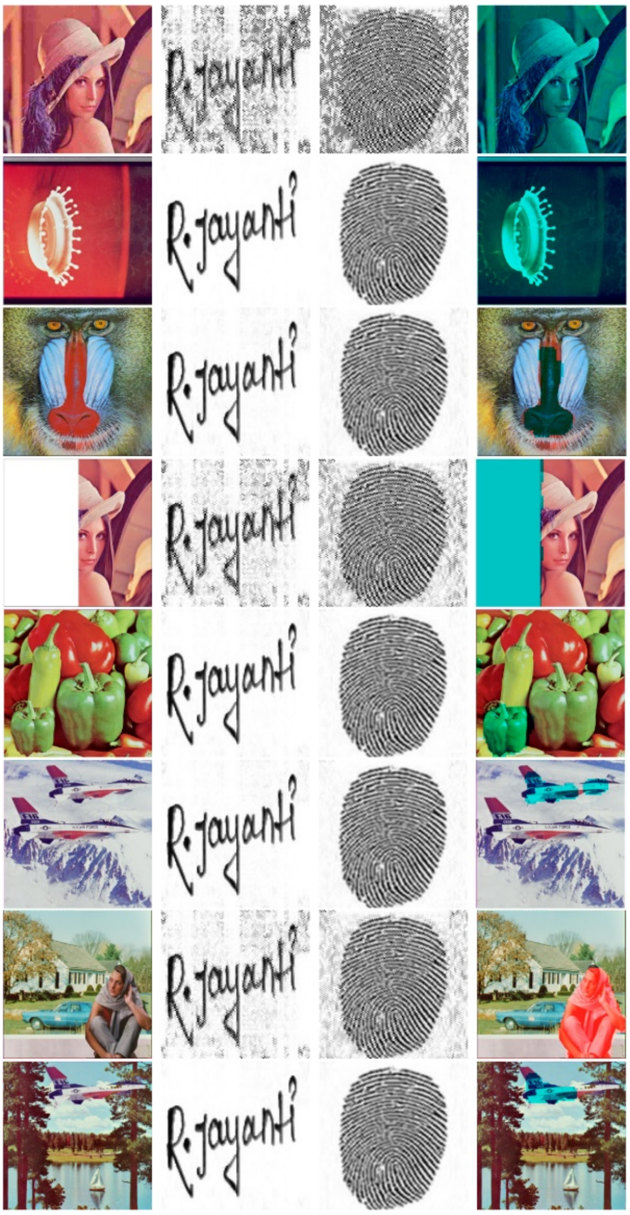


Figure 16: Watermark extraction and tamper localization across various forgery types

ther strengthen the approach. This can help to widen its applicability and address challenges in rotation and compression challenges.

References

[1] Sprout Social. *80+ Must-Know Social Media Marketing Statistics for 2025*. https://sproutsocial.com/insights/social-media-statistics/. Accessed: 2025-11-10. 2025.

[2] Nikitha Sreekanthaswamy et al. "Digital tools and methods". In: *Enhancing School Counseling With Technology and Case Studies*. IGI Global Scientific Publishing, 2025, pp. 25–48. URL: https://doi.org/10.4018/979-8-3693-8392-6.ch002.

[3] Aditya Kumar Sahu et al. "A study on content tampering in multimedia watermarking". In: *SN Computer Science* 4.3 (2023), p. 222. URL: https://doi.org/10.1007/s42979-022-01657-1.

[4] Paweł Duszejko, Tomasz Walczyna, and Zbigniew Piotrowski. "Detection of Manipulations in Digital Images: A Review of Passive and Active Methods Utilizing Deep Learning". In: *Applied Sciences* 15.2 (2025), p. 881. URL: https://doi.org/10.3390/app15020881.

[5] Anna Gruending et al. "Born Too Soon: learning from the past to accelerate action in the next decade". In: *Reproductive Health* 22.2 (2025), pp. 1–15. URL: https://doi.org/10.1186/s12978-025-02044-8.

[6] India Today Web Desk. "FAKE ALERT: No, DGP of Gujarat is not touching Rajnath Singh's feet".

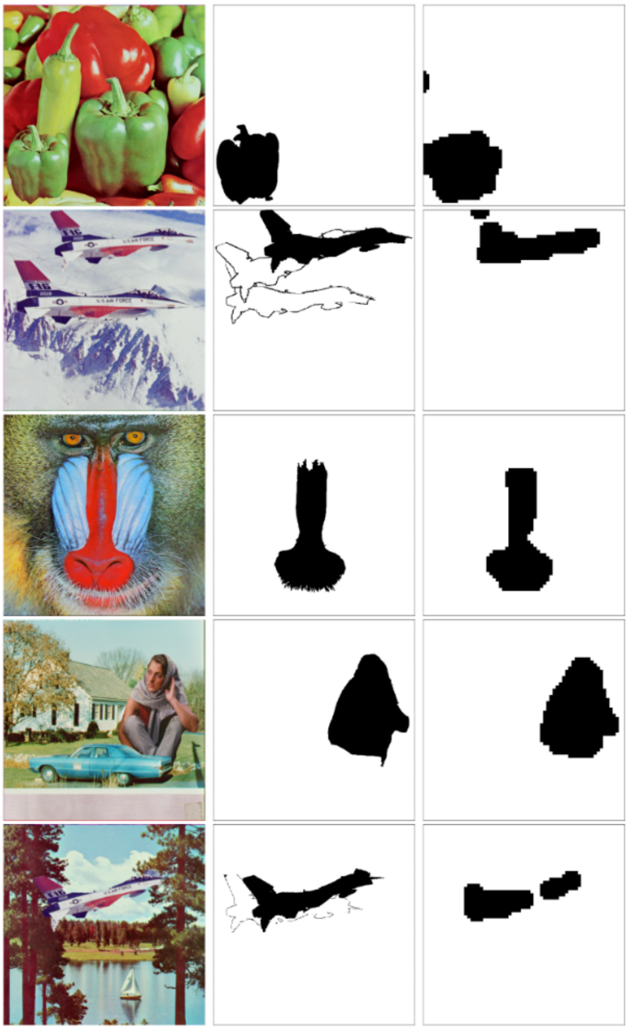


Figure 17: Comparison of predicted and ground truth tampered regions for different manipulations

In: *India Today* (Nov. 2017). Accessed: 2024-05-04. URL: https://www.indiatoday.in/fyi/story/rajnath-singh-fake-image-dgp-gujarat-congress-sanjay-jha-1078093-2017-11-01.

[7] Hany Farid. *Photo Tampering Throughout History*. Online manuscript. Accessed via Georgia Tech faculty website (PDF, 30 pages). 2008. URL: https://faculty.cc.gatech.edu/~beki/cs4001/history.pdf.

[8] Shishir Kumar Raj et al. "An Automated Deep Learning Model for Detecting Image Forgeries". In: *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*. Vol. 1. IEEE. 2024, pp. 1–6. URL: https://doi.org/10.1109/ICEECT61758.2024.10738982.

[9] Santoshini Panda and Minati Mishra. "Passive techniques of digital image forgery detection: developments and challenges". In: *Advances in Electronics, Communication and Computing: ETAEERE-2016*. Springer, 2017, pp. 281–290. URL: https://doi.org/10.1007/978-981-10-4765-7_29.

[10] Wenbo Wan et al. "A comprehensive survey on robust image watermarking". In: *Neurocomputing* 488 (2022), pp. 226–247. URL: https://doi.org/10.1016/j.neucom.2022.02.083.

[11] Sanjeeb Kumar Behera and Minati Mishra. "Steganography–A Game of Hide and Seek in Information Communication". In: *arXiv preprint arXiv:1604.00493* (2016). URL: https://doi.org/10.48550/arXiv.1604.00493.

[12] Qi Peng Yu Feng, De-Biao He, and Min Luo. "A survey on threshold digital signature schemes". In: *Frontiers of Computer Science* 20.4 (2025), p. 2004806. URL: https://doi.org/10.1007/s11704-025-41297-1.

[13] Navneet Kaur, Neeru Jindal, and Kulbir Singh. "Passive image forgery detection techniques: A review, challenges, and future directions". In: *Wireless Personal Communications* 134.3 (2024), pp. 1491–1529. URL: https://doi.org/10.1007/s11277-024-10959-x.

[14] Divya Prathana Timothy and Ajit Kumar Santra. "Detecting Digital Image Forgeries with Copy-Move and Splicing Image Analysis using Deep Learning Techniques". In: *International Journal of Advanced Computer Science & Applications* 15.5 (2024). URL: https://doi.org/10.14569/ijacsa.2024.01505131.

[15] Krishna H Hingrajiya and Chintan Patel. "An Approach for Copy-Move and Image Splicing Forgery Detection using Automated Deep Learning". In: *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE. 2023, pp. 1–5. URL: https://doi.org/10.1109/ESCI56872.2023.10100202.

[16] Saurabh Agarwal, Savita Walia, and Ki-Hyun Jung. "A cohesive forgery detection for splicing and copy-paste in digital images". In: *Multimedia Tools and Applications* 84.1 (2025), pp. 147–163. URL: https://doi.org/10.1007/s11042-024-19825-1.

[17] Ingemar Cox et al. "Digital watermarking". In: *Journal of Electronic Imaging* 11.3 (2002), pp. 414–414. URL: https://doi.org/10.1117/1.1494075.

[18] Mahbuba Begum and Mohammad Shorif Uddin. "Digital image watermarking techniques: a review". In: *Information* 11.2 (2020), p. 110. URL: https://doi.org/10.3390/info11020110.

[19] Ms Shubhangi D Mashalkar and SS Shirgan. "Design of watermarking scheme in medical image authentication using DWT and SVD technique". In: *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE. 2017, pp. 955–960. URL: https://doi.org/10.1109/ICCMC.2017.8282609.

[20] Dengsheng Zhang. "Wavelet transform". In: *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*. Springer, 2019, pp. 35–44. URL: https://doi.org/10.1007/978-3-030-17989-2_3.

[21] Shaifali M Arora et al. "A DWT-SVD based robust digital watermarking for digital images". In: *Procedia computer science* 132 (2018), pp. 1441–1448. URL: https://doi.org/10.1016/j.procs.2018.05.076.

[22] Narima Zermi et al. "A DWT-SVD based robust digital watermarking for medical image security". In: *Forensic science international* 320 (2021), p. 110691. URL: https://doi.org/10.1016/j.forsciint.2021.110691.

[23] David Mata-Mendoza et al. "Secured telemedicine of medical imaging based on dual robust watermarking". In: *The Visual Computer* 38.6 (2022), pp. 2073–2090. URL: https://doi.org/10.1007/s00371-021-02267-3.

[24] Bhargavi Mokashi et al. "Efficient Hybrid Blind Watermarking in DWT-DCT-SVD with Dual Biometric Features for Images". In: *Contrast Media & Molecular Imaging* 2022.1 (2022), p. 2918126. URL: https://doi.org/10.1155/2022/2918126.

[25] Shyam Singh Rajput, Bhaskar Mondal, and Farheen Qamar Warsi. "A robust watermarking scheme via optimization-based image reconstruction technique". In: *Multimedia Tools and Applications* 82.16 (2023), pp. 25039–25060. URL: https://doi.org/10.1007/s11042-023-14363-8.

[26] Runyi Hu et al. “Robust-wide: Robust watermarking against instruction-driven image editing”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 20–37. URL: https://doi.org/10.1007/978-3-031-72670-5_2.

[27] Paarth Neekhara et al. “FaceSigns: semi-fragile watermarks for media authentication”. In: *ACM Transactions on Multimedia Computing, Communications and Applications* 20.11 (2024), pp. 1–21. URL: https://doi.org/10.1145/3640466.

[28] Aakash Varma Nadimpalli and Ajita Rattani. “Social media authentication and combating deepfakes using semi-fragile invisible image watermarking”. In: *Digital Threats: Research and Practice* 5.4 (2024), pp. 1–30. URL: https://doi.org/10.1145/3700146.

[29] De Li et al. “A reversible watermarking for image content authentication based on wavelet transform”. In: *Signal, Image and Video Processing* 18.3 (2024), pp. 2799–2809. URL: https://doi.org/10.1007/s11760-023-02950-z.

[30] Ferda Ernawan et al. “Fragile and Robust Dual Image Watermarking based on DWT-SVD”. In: *2024 10th International Conference on Mechatronics and Robotics Engineering (ICMRE)*. IEEE. 2024, pp. 247–252. URL: https://doi.org/10.1109/ICMRE60776.2024.10532164.

[31] Ernesto Moya-Albor et al. “Bio-inspired watermarking method for authentication of fundus images in computer-aided diagnosis of retinopathy”. In: *Mathematics* 12.5 (2024), p. 734. URL: https://doi.org/10.3390/math12050734.

[32] Wu-Chih Hu et al. “Effective image forgery detection of tampered foreground or background image based on image watermarking and alpha mattes”. In: *Multimedia Tools and Applications* 75 (2016), pp. 3495–3516. URL: https://doi.org/10.1007/s11042-015-2449-0.

[33] Thai-Son Nguyen. “Fragile watermarking for image authentication based on DWT-SVD-DCT techniques”. In: *Multimedia Tools and Applications* 80.16 (2021), pp. 25107–25119. URL: https://doi.org/10.1007/s11042-021-10879-z.

[34] Moataz Z Salim, Ali J Abboud, and Remzi Yildirim. “A visual cryptography-based watermarking approach for the detection and localization of image forgery”. In: *Electronics* 11.1 (2022), p. 136. URL: https://doi.org/10.3390/electronics11010136.

[35] Chia-Chen Lin et al. “Fragile watermarking for tamper localization and self-recovery based on AMBTC and VQ”. In: *Electronics* 12.2 (2023), p. 415. URL: https://doi.org/10.3390/electronics12020415.

[36] Cai Zhan et al. “Reversible Image Fragile Watermarking with Dual Tampering Detection”. In: *Electronics* 13.10 (2024), p. 1884. URL: https://doi.org/10.3390/electronics13101884.

[37] Prajanto Wahyu Adi et al. “Efficient Fragile Watermarking for Image Tampering Detection using Adaptive Matrix on Chaotic Sequencing”. In: *Intelligent Systems with Applications* (2025), p. 200530. URL: https://doi.org/10.1016/j.iswa.2025.200530.

[38] P. Aberna and L. Agilandeeswari. “Optimal Semi-Fragile Watermarking Based on Maximum Entropy Random Walk and Swin Transformer for Tamper Localization”. In: *IEEE Access* 12 (2024), pp. 37757–37781. DOI: https://doi.org/10.1109/ACCESS.2024.3370411.

[39] Asmaa Hatem Jawad et al. “A New Face Image Manipulation Localization and Recovery Algorithm Using Image Watermarking and Integer Wavelet Transform”. In: *Mathematical Modelling of Engineering Problems* 11.9 (2024). DOI: https://doi.org/10.18280/mmep.110920.

[40] USC-SIPI. *USC-SIPI Image Database*. https://sipi.usc.edu/database/. [Accessed: 2024-06-03]. 2024.