

Received: 13 December 2022; Accepted: 21 April, 2023; Published: 9 June, 2023

Vision Transformer-Based Watermark Generation for Authentication and Tamper Detection Using Schur Decomposition and Hybrid Transforms

Aberna P¹, Agilandeswari L², Aashish Bansal³

¹ Research Scholar, School of Information Technology and Engineering,
Vellore Institute of Technology, Vellore, TN 632014, India

aberna.p2020@vitstudent.ac.in

² Professor, School of Information Technology and Engineering,
Vellore Institute of Technology, Vellore, TN 632014, India

agila.l@vit.ac.in

³ Student, School of Information Technology and Engineering,
Vellore Institute of Technology, Vellore, TN 632014, India

aashish.bansal2019@vitstudent.ac.in

Abstract: Multimedia manipulation has increased which demand for security across various applications. Majorly image-oriented security issues such as image authentication, proof of ownership, and copyright protection are highly increased. To authenticate and detect the tampered region and to recover the tampered area vision transformer-based hybrid watermarking model is proposed. We proposed a novel model to achieve image authentication, tamper detection, and localization followed by image recovery. In the proposed model, invariant attention-based watermark feature maps are generated using a Vision transformer. We have generated three different watermarks: first using the SVD eigenvalue generated as an authentication watermark, secondly to detect tampered region average 6MSB of each 2*2 block generated by performing the Schur decomposition method on the biometric image, and to locate and recover the image Vision feature maps are generated and average 6MSB of each block is embedded as the tamper detection watermark. Normally the generated watermark is embedded either using an embedding factor or using a suitable embedding location. In the proposed model, watermark embedding is performed by finding the optimal embedding region using high entropy block region. On the original image, curvelet transform is performed followed by invariant integer wavelet transform. The first authentication eigenvalue is embedded on the LH band singular value diagonal matrix obtained by the SVD model. On the LL band, 2*2 blocks Schur decomposition is performed to embed the 6MSB in the 2LSB bit of upper triangular coefficient values. At last, vision feature maps are embedded in the curvelet approximate coefficient high entropy region and inverse curvelet performed, producing a watermarked image.

Keywords: Vision transformer, Schur decomposition, Integer wavelet transform, DWT, SVD, Entropy.

I. Introduction

Internet technology's growth skyrocketed multimedia data usage and transmission raised multimedia data manipulations. Due to this uncontrollable sharing of multimedia data especially images; securing multimedia data from third-party is a major concern. Among all multimedia data, image sharing and transmission across are higher various applications such as social platforms, the film industry, Healthcare, copy control, etc.... Nowadays image manipulation, replacement, and regeneration are not tough tasks due to the availability of advanced tools. Among all the applications, the social media platform and healthcare industry are extremely prone to cyber criminals and unintentional data leakage. Social media platforms like Facebook, Instagram, Twitter, etc., have attracted people, due to which users' usage of such platforms increased, and also multimedia data sharing increased uncontrollably. In the healthcare system, though electronic healthcare is saving people's lives worldwide, there are still several difficulties that must be addressed to increase the efficiency of this technology. Social media, Telemedicine applications, Biomedical image processing, and authentication/security of biomedical data during the transition are some of the crucial domains that are gaining increased attention [1]. Data transmission whether it might be healthcare-related reports or the user's own multimedia data is shared across the network technology and leads to either intentional or unintentional attacks which demand multimedia data authentication and security. For instance, in the healthcare system, Digital Imaging and Communications in Medicine (DICOM) [2] was one medical image management system that stored and shared medical data evolved in 1993. Multimedia images should be trustworthy and guarantee that the image is authentic. Before making a diagnosis, it is crucial to confirm and authenticate the medical images. Multimedia platforms are very concerned with protecting

images from tamper detection applications, authenticity, and integrity verification.

To ensure multimedia image security such as data authentication, authorization, and copyright protection many researchers have suggested data embedding techniques to achieve multimedia security. Watermarking is the most suitable data embedding technique suggested in the existing research. The watermarking technique was implemented in 1992 by Andrew Terkel and Charles in his paper "Electronic Watermark" [3]. The digital watermarking technique showed a state-of-the-art method for data embedding, and also it has significance to prove ownership, copyright protection, and multimedia data authentication [4]. Embedding watermark data in an image is said to be Watermarking technique and the output is said to be a watermarked image. Watermarking is classified based on two domains: Pixel-based and frequency-based. Embedding the image at the pixel level is known to be a pixel-based domain or spatial domain whereas embedding in spectral coefficients using various frequency methods is known to be a frequency or Transform domain. The spatial coefficient is transformed into a spectral coefficient using a few methods: Discrete Fourier transform (DFT) [5], Discrete Cosine Transform (DCT) [6] and Discrete Wavelet Transform [7, 8, 9], Quaternion Curvelet transform [10], Hilbert Transform [11] [12], Integer Wavelet Transform [13, 12] and Contourlet [14, 15, 16]. Matrix decomposition methods are Singular Value Decomposition (SVD) [17, 18], Schur Decomposition [19]. Based on the domain, the watermarking technique is sub-categorized into three embedding techniques: Robust [20, 21], Fragile watermarking [22, 23], and Semi-fragile watermarking [24]. Robust watermarking refers to a watermark that should remain stable against a variety of unintentional attacks unless the cover image is modified, while fragile watermarking is intended to precisely detect the tampered region and is sensitive to attacks, making it suitable for tamper detection applications. When a watermark is semi-fragile, it is sufficiently stable against unintentional attacks but not against intentional ones wherein authentication and tamper detection are achieved simultaneously, which reflects a combination of both robust and fragile watermarks. Much research has been carried out in the tamper detection field which shows, that a traditional watermarking system with machine learning models attracted attention to various applications. In deep learning techniques, Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) are widely employed for Image classification, computer vision applications, object recognition, etc. A drawback with watermarking-based deep learning techniques is, a small change in the image pixel will affect the neural network performance in terms of fidelity and it tries to fool the network into making wrong predictions. The attention mechanism is an integral part of the transformer model that slowly grabs the attention of computer vision applications in hybridizing with Convolutional Neural Networks (CNN). For large datasets, CNN requires hard inductive bias which is avoided by the transformer. The transformer model is a straightforward and parallel processing method that has shown more state-of-the-art results than CNN, by eschewing convolution instead self-attention plays a key role. Transformer was inspired by the success of the Natural

Language Processing (NLP) task in machine translation [25]. Currently, Transformer models are limited to computer vision applications. Especially transformer named Vision Transformer (ViT) proposed by Google research-Brain team members is slowly grabbing the attention of image processing applications.

The rest of the paper is organized as follows: Section 2 describes the related work; Section 3 explains about preliminary concept followed by an elaboration proposed model in Section 4; Section 5 illustrates the experimental results. Finally, a conclusion is in section 6.

II. Related Works

Tamper detection and localization algorithm using the block-based watermarking method proposed by Campos-Ponce, E et al. [26] for both color and grayscale images. They generated hash key code using the checksum method which is embedded in the LSB coefficient of the low-frequency sub-band of the lifting wavelet transform. The experiment is evaluated against various attacks in terms of imperceptibility and false positive, false negative, and tamper detection rates which attained good imperceptibility with the PSNR value of 51.17 dB.

A hybrid transform-based watermarking model is proposed for the integrity protection of DICOM images proposed by Tiwari, A et al. [27]. First, a 3-level Integer Wavelet Transform is applied to the liver ultrasound image. Further Schur decomposition was performed on the HH sub-band which generated the diagonal upper triangular matrix. The upper triangular matrix is further processed by SVD to obtain singular values. The obtained singular value is embedded in the low-frequency sub-band. Though they have tested for various attacks watermark generated from the HH sub-band won't be attaining better results. As the HH sub-band will capture only the edge components which won't be suitable to locate the tampered region. Security is obtained using the encrypted Arnold chaotic method for its integrity protection. The experimental results show strong robustness against various attacks.

A highly secured invariant Redistributed IWT transform approach is suggested by [28] to generate the invariant domain of the original image. Further on the invariant domain LL and HH sub-band novel QR matrix decomposition has been performed to generate Q and R matrix. Singular values are generated from the R matrix of both the sub-band to embed the watermark image. Another side the watermark image redistributed IWT performed and the LL band is chosen as a watermark data which is embedded in the singular value matrix. Security and reversibility are achieved by encrypting the watermark using gyration transform QR decomposition. The result showed resistance against geometric attacks.

A multiscale watermarking model was proposed by [29] for copyright protection applications using a 1-level Integer wavelet transform. An efficient singular value decomposition technique is applied to all four sub-band coefficients. Later on, each singular value matrix is further subdivided into a non-overlapping matrix based on the watermark size, and the watermark is embedded on the singular value of the LL band using the embedding factor. For the other sub-divided singular value matrix, the watermark is embedded with different strength factors. Further to authenticate the image SHA-1 hash function is

defined to obtain 160-bit watermark data from the signature image which is embedded in the DWT-SVD coefficient values. The experimental results are evaluated by PSNR which resulted in the maximum value of 47.6 dB and SSIM value of 0.999.

A deep learning-based watermarking model is proposed using DCT-CNN by [30]. They generated two features: first to detect tamper region and the second generated image digest watermark feature utilized for image recovery. For tamper detection, a CNN-based authentication watermark was generated. Using Quantization based DCT transform method image digest watermark is generated and an end-to-end CNN model is employed to compress the model and also to maintain the image quality. Error correction code reed-Solomon is used to protect high distortion. To secure the data Arnold transform is applied to the authentication and digest watermark.

Rajput, V. et al. [31] presented a tamper detection model using Discrete Wavelet Transform (DWT). From the original image, the DWT technique was applied, and on the four generated sub-bands second-level DWT was applied which is taken as a watermark image. Using pseudo-random codes, the four reduced LL low-frequency sub-band images are embedded in the 4-LSB of the original image. The result shows better tamper detection accuracy and improve performance of image recovery.

Agilandeewari. L. et. al [46] proposes a robust semi-fragile watermarking system using Pseudo-Zernike moments and dual tree complex wavelet transform for social media content authentication.

Contributions: The main contribution of the paper is

1. Attempted vision transformer model for the first time in the field of watermarking and succeeded in image tampering applications.
2. Watermark data is generated innovatively using the ViT model, which extracts more attentive encoded global feature maps that improve tampered image recovery.
3. Watermark robustness is achieved by a hybrid watermarking algorithm
- 4.

III. Preliminary Concepts

A. Vision Transformer

Vision transformer (ViT) [32] is a trending model for various applications such as image classification and, Prediction [47], object recognition, image captioning, etc. Similarly, its partner model named Swin Trasformer also best suits for image detection and classification [45]. Normally transformer model takes a 1D vector sequence as input whereas in a 2D image split image $I \in \mathbb{Q}^{h \times w \times c}$ into patches P using window size $M \times M$, where $h \times w$ represents image dimension, and C represents the number of channels. From the image I , N number of patches generated using Eq.1 are linearly flattened into a sequence of patches ($p_1, p_2, p_3, \dots, p_n$) of length n . The known embedding matrix, E , is used to linearly project the image patches into a vector with dimension, d represented as $[x_1E, x_2E, \dots, x_nE]$. Classification label CL is attached with the linearly projected vector for classification purposes. Followed by that positional information is added to the vector patches to arrange the image patches as shown in Eq.2

$$N = hw / P^2 C \quad (1)$$

$$PV_0 = [CL; x_1E, x_2E, \dots, x_nE;]_{+ \text{ pos}} \quad (2)$$

Where, $E \in \mathbb{Q}^{(p^2c) \times d}$, $\text{pos} \in \mathbb{Q}^{(n+1) \times d}$

The vision encoder receives the series of embedded image patches, $(PV_1, PV_2, \dots, PV_N)$ as input. Vision Transformer Encoder is stacked up with identical layers: Multihead attention (MAS), fully connected Feed Forward MLP (Multilayer Perceptron). The GeLU activation function is sandwiched between Multihead attention (MHA), and fully connected Feed Forward MLP (Multilayer Perceptron). The two encoder layers collaborate via the normalization layer using the residual connections.

Self-attention layer is the most important layer in the transformer, where it shows the importance of a single patch by combining it with other patches. Self-attention heads generate Key (K), Value (V), and Query (Q) vectors by multiplying each input patch vector (PV_N) with the learned weight matrices W_q, W_k, W_v , which is represented as $Q_i = W_q PV_i, K_i = W_k PV_i, V_i = W_v PV_i$. The weight matrix will be the same for all the input vector $X_{i:n}$ sequences. Self-attention computes the average weighted values [33] by scaling the dot product of the query vector Q_i with all other outcomes key vector $K_{i:n}$ and dividing by key dimension d_k Eq (3). The outcome of scaling dot product is given to the softmax function and multiplied with relevant values element V_i as shown in Eq (4). Each self-attention mechanism referred to as the head performs parallelly and is concatenated together as Multi-Head Attention (MHA) shown in Eq (5).

$$\text{Scaling Dot product (Q, K, V)} = \frac{K_{i:n}^T Q_i}{\sqrt{d_k}} \quad (3)$$

Self-Attention (SA_i) =

$$\text{softmax}\left(\frac{K_{i:n}^T Q_i}{\sqrt{d_k}}\right) V_i, \text{ where } i = 1, 2, \dots, n \quad (4)$$

MultiHead Attention (MHA) = Concat ($SA_1 + SA_2, \dots, SA_n$) (5)

$$Z_i^1 = \text{MHA}(\text{Layernorm}(PV_{N-1})) + \dots PV_{N-1} \quad (6)$$

Where, $N=1, 2, \dots, n$

The normalized outcome of MHA is concatenated together which is provided as input to the MLP classifier of the encoder block to obtain pixel value based on learned feature maps Eq.5.

$$Z'_n = \text{MLP}(\text{Layernorm}(Z_i^1)) + Z_i^1 \quad 1=1, 2, \dots, L \quad (7)$$

$$V_i = (\text{Layernorm}(Z'_n)) \quad (8)$$

B. Discrete Wavelet Transform

Discrete wavelet transform (DWT) is an efficiently employed powerful tool in hierarchical decomposition technique. Spatial localization of an image is transformed to wavelet frequency which is good at frequency resolution and poor at time resolution [34]. In DWT, the spatial coefficient is passed to the wavelet filter which decomposed the spatial image into four sub-bands. So, the first level of four sub-band decomposition is figured out in Figure 1, as low frequency-LL band, high frequency- HH band, mid-frequency- LH, HL band. In the case, of multi-level decomposition, the sub-band is further decomposed into LL1, HH1, HL1, and LH1 as shown in Figure 1. The watermark is highly suggested to be embedded in a low-frequency LL sub-band as it can withstand various attacks and is difficult to extract the watermark. DWT helps to achieve high robustness and imperceptibility property.

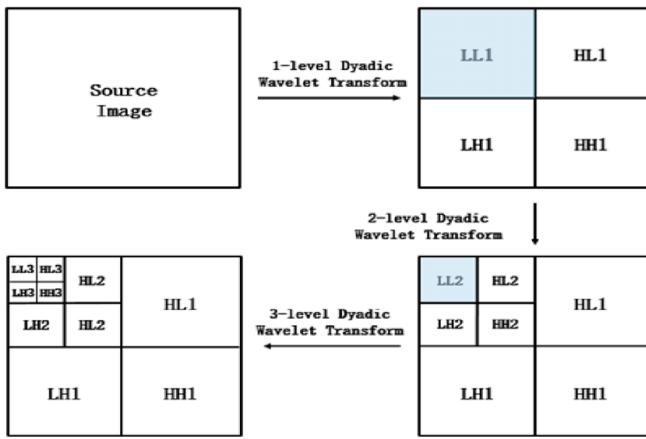


Figure 1: DWT decomposition [35]

C. Singular Value Decomposition

Singular value decomposition (SVD) is a powerful matrix decomposition technique that decomposes the image into orthogonal and singular value components [34, 36]. From a linear algebra point of view, the image is considered to have a non-negative scalar input, that is a matrix. Image is decomposed into the matrix as $I = USV^T$ where U and V are left and right eigenvectors, whereas S is a diagonal eigenvalue matrix. The singular value component S has the durable ability to withstand any perturbation.

$$I = \begin{pmatrix} u_{1,1} & \dots & u_{1,m} \\ u_{2,1} & \dots & u_{2,m} \\ u_{3,1} & \dots & u_{3,m} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} v_{1,1} & \dots & v_{1,n} \\ v_{2,1} & \dots & v_{2,n} \\ v_{3,1} & \dots & v_{3,n} \end{pmatrix} \quad (9)$$

D. Schur Decomposition

Schur decomposition is one of the matrix decomposition methods and an important mathematical linear algebraic tool like the singular value decomposition method [19]. If the image of size $n \times m$, the Schur method is applied to produce $S = UDU^T$ matrix where represents the unitary matrix, D represents the upper triangular matrix and the diagonal of D shows eigenvalue of the image A .

$$S = \begin{pmatrix} u_{1,1} & \dots & u_{1,m} \\ u_{2,1} & \dots & u_{2,m} \\ u_{3,1} & \dots & u_{3,m} \end{pmatrix} \begin{pmatrix} d_{1,1} & d_{1,2} & d_{1,3} \\ 0 & d_{2,2} & d_{2,3} \\ 0 & 0 & d_{3,3} \end{pmatrix} \begin{pmatrix} u_{1,1} & \dots & u_{1,m} \\ u_{2,1} & \dots & u_{2,m} \\ u_{3,1} & \dots & u_{3,m} \end{pmatrix}^T \quad (10)$$

E. Curvelet Transform

Wavelet transforms and the concept of multiresolution is widely used, particularly in the signal and image processing areas. However, due to its inability to give higher direction selectivity, two-dimensional (2-D) discrete wavelet transformations (DWT) cannot capture anisotropic data. A multiscale resolution technique called the curvelet transform enables the best non-adaptive sparse image representation of objects with edges [37]. Curvelet transform, as opposed to wavelet transform, generates a directed feature representation. A new phase of image processing has commenced because of multiscale geometric transformations such as the curvelet transform, which was first introduced by Candes and Donoho et al. [38]. Curvelet

transform is better at capturing the curvature and hyperplane singularities of high-dimensional data. Anisotropy and directionality, key features of curvelets, provide the best formula for the representation of smooth curves in an image, such as edges and region boundaries. In-depth detail of the curvelet transform can be found in [38]. There are two ways to acquire the conventional Curvelet coefficients [34]: (1) the USFFT approach, and (2) the Wrapping method. The production of Curvelet coefficients is done using the Wrapping approach in our proposed work because of its fast work. Based on the wrapping method, the multiscale pyramid is determined at a different angle in the frequency domain [39]. Fast Fourier transform is applied on the given orientation and scale x . Curvelet transform is obtained by performing the inverse fast Fourier transform. We can conceive of the result of these linear digital transformations as a set of coefficients $C^D(x, y, k)$ obtained by the digital analogue, as they accept as input Cartesian arrays of type $f[m, n]$, $0 \leq m, n < t$. The curvelet transform of the function C^D can be expressed as

$$C^D(x, y, k) = \sum_{0 \leq m, n < t} f[m, n] \varphi_{x,y,k}^D[m, n] \quad (11)$$

F. Integer Wavelet Transform

Integer wavelet is a novel wavelet transform that is demonstrated by lifting strategies. The IWT uses three standard lifting techniques [29] as shown in Figure 2: split, predict, and update. Compared with a discrete wavelet transform benefits of an integer wavelet are [40]: (1) Sub-band coefficient values of DWT will be in floating values, during reconstruction losing the float values by rounding it to integers. Whereas the IWT lifting algorithm transforms an integer to integer sub-band values. (2) All calculations are done directly and memory storage is required. These characteristics of IWT can be used to maintain the imperceptibility level and increase the robustness. In split-phase or lazy wavelet, the cover image is divided into even and odd polyphase components. The second predicted phase new odd polyphase generated based on even polyphase components. An old odd polyphase component is replaced by the difference between the odd phase and the predicted value. Further in the update phase, based on the linear combination sample input from the predicting phase determine the new even polyphase component.

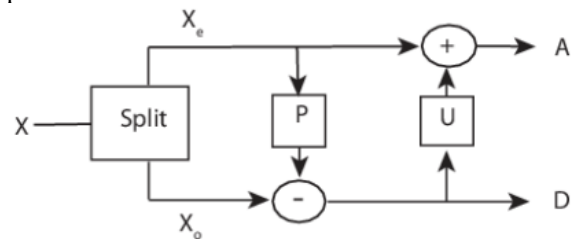


Figure 2: Integer wavelet [40]

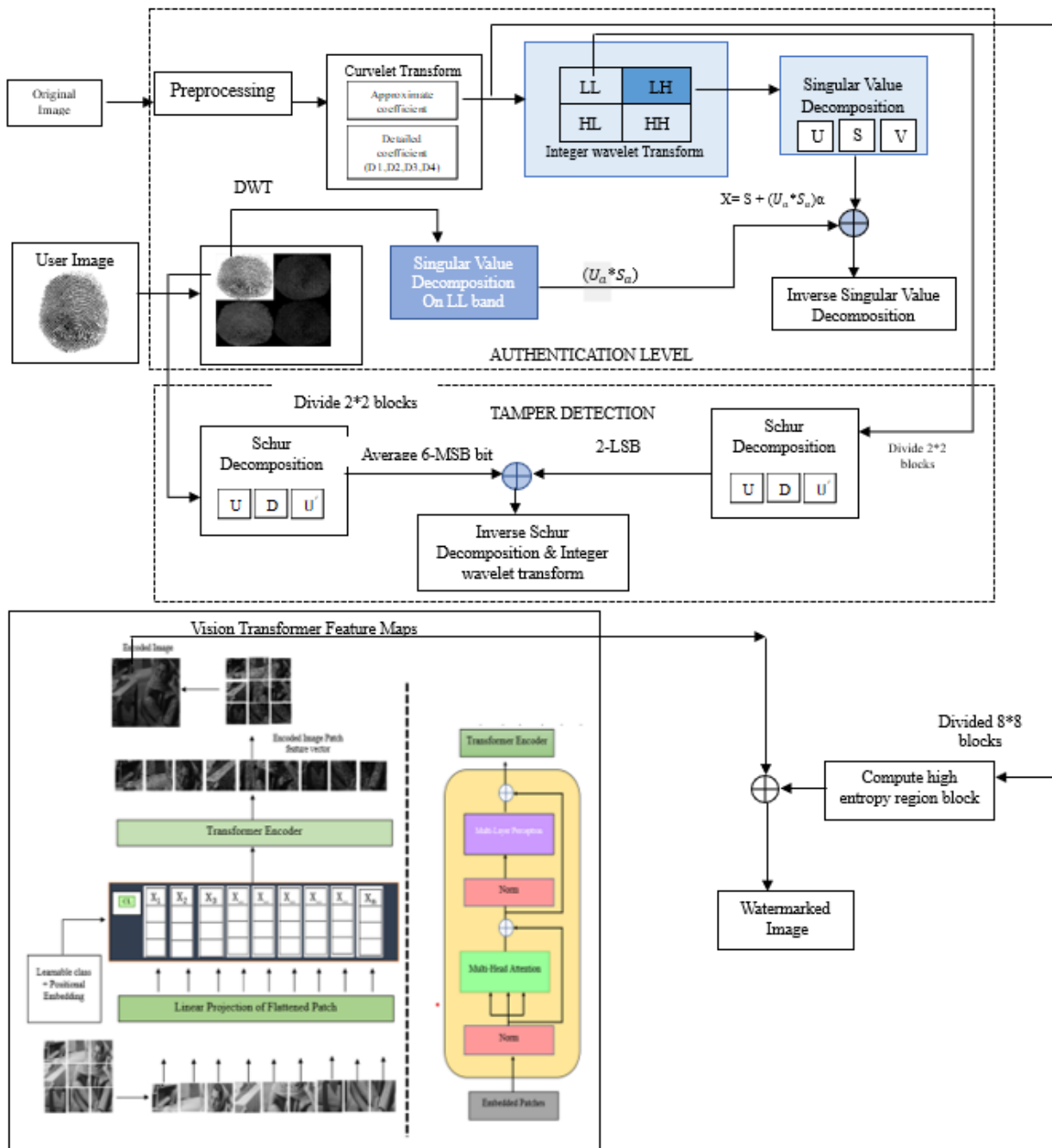


Figure 3: Proposed Embedding Process

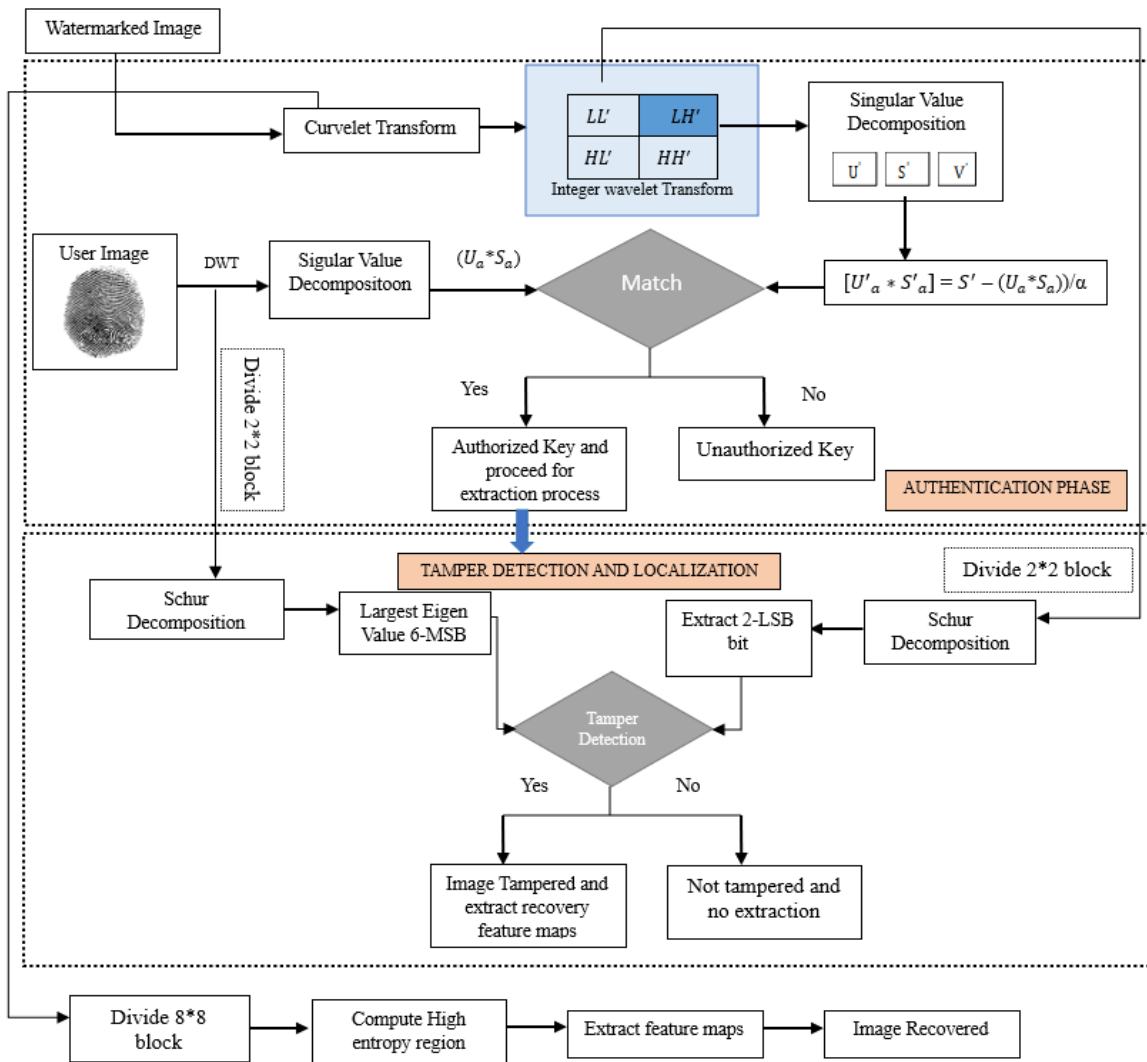


Figure 4: Proposed Extraction Process

IV. Proposed Model

Generating and embedding a watermark for each image slice using the conventional watermarking technique is difficult. So, deep learning models are exposed to be a highly efficient technique to learn the image features from the dataset and embed the appropriate watermark in an appropriate original image slice through the class label. Compared to deep learning techniques such as CNN, vision transformer has shown state-of-the-results in image classification, which attracted the researcher’s intention to extend it to computer vision applications. In this paper, we have attempted a vision transformer model for the first time in a tamper detection application to generate encoded features as the watermark. In vision transformer (ViT), the attention mechanism plays a prominent role to learn the global important intrinsic feature in the image, which gives out encoded features. Section 3.1, clearly shows the workflow of vision transformer-based feature map generation. In the proposed model, vision transformer feature maps were employed as the tamper localization watermark data that made the system attain high robustness against various attacks than the existing traditional

watermarking techniques. Figure 3, shown above depicts the embedding process of the proposed model. To verify the ownership of the image shared by the sender, at the receiver the input image has to be preprocessed first to denoise it and restore the image quality. On the preprocessed image, the applied curvelet transform was to generate fine and detailed coefficient values. Further, on the fine coefficient values AC_c , applied integer wavelet transform to obtain integer-to-integer LH_c sub-band coefficient value. On the mid-frequency LH_c sub-band, singular value matrix decomposition is applied to embed the authentication key K_a in the S matrix. Then inverse SVD is performed which gives out a modified LH_c sub-band. To detect watermark data generated by partitioning owner biometric image into 2×2 blocks and performing the Schur decomposition method on the spatial domain of each block. From the upper triangular matrix D average value are computed and the six most significant bits of each block are selected as watermark bits Y_i . On the LL_c sub-band of the cover image, divide into 2×2 blocks and apply the Schur decomposition method on each block. The watermark bit Y_i is embedded in the 2LSB of the diagonal matrix of the cover image. Inverse Schur decomposition is performed and all the blocks are combined and inverse integer wavelet transform is performed which

produces a modified approximate coefficient. On the approximate coefficient, high entropy region is computed on every 8×8 block to embed the recovery watermark. At last, the vision feature is embedded in the high entropy region and the inverse curvelet transform is applied which gives out the final watermarked image.

Figure 4, depicts the watermark extraction process which is the inverse of the embedding process. On the watermarked image curvelet transform is applied and further integer wavelet transform is applied on the approximate coefficient AC_w . From the mid-frequency LH'_w , the primary authentication key EK_a is extracted from the singular value decomposition diagonal matrix S'_w . To validate the authentication key owner biometric image 'b' is chosen from the owners' database and the low-frequency LL_a sub-band is generated by applying a 1-level discrete wavelet transform and eigenvalue or principal component K_a generated from the SVD method. Match both the extracted original authentication key and watermarked authentication key. If the key matches proceed further with the extraction process otherwise no extraction process. To verify tamper detection, divide the LL'_w into 2×2 blocks and apply Schur decomposition on each block. From the diagonal matrix, extract the 6MSB from the 2LSB bit of each upper triangular pixel value. Using the owners' biometric image generate the average value of each 2×2 block of the Schur decomposition diagonal matrix. Compare both the MSB bit to verify whether the image is tampered with or not. If it has been tampered with, apply inverse Schur and integer wavelet transform. Then extract the feature maps from the high entropy region of 8×8 blocks of the approximate coefficient to locate the tamper region and recover it.

Algorithm

Process 1: Authentication watermark generation

Eigenvalues are generated as authentication keys in order to authenticate the proof of ownership of the watermarked image W'_i . Authentication primary key is required to proceed further for extraction purposes. Only if the primary authentication key matches the extraction process done, else no extraction process. Some of the sample owner's database images are mentioned in Figure 10.

1. In order to improve image quality and eliminate noise, the owner's biometric image 'b' is preprocessed using an adaptive median filter.
2. Apply 1-level Discrete wavelet transform to the preprocessed image to extract important data.

$$[LL_a, LH_a, HL_a, HH_a] = 1 - DWT(b) \quad (12)$$

3. At last, the principal component K_a generated from the LL band,

$$[U_a, S_a, V_a] = SVD(LL_a) \quad (13)$$

$$K_a = U_a \times S_a \quad (14)$$

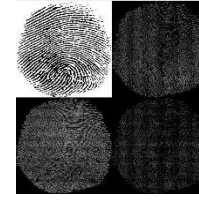


Figure 5: Discrete wavelet transform (Owner's database)

Process 2: Tamper detection data generation

1. From the owner biometric image 'b', the tamper detection watermark is generated by splitting the image into 2×2 blocks.

$$b_i = b_{n \times m} / \text{block size} \quad (15)$$

where $n \times m$ represents a dimension of the image.

2. For each block, Schur decomposition technique is applied

$$[U_T, D_T, U_T'] = \text{schur}(b_i), i=1,2,\dots,n,m \quad (16)$$

3. Compute average value Y_i of each block b_i and 6MSB are chosen as tamper detection watermark data.

$$Y_i = \sum_{j=1}^4 D_T / 4 \quad (17)$$

Process 3: Recovery Watermark generation

1. Fine-tune the training data I_i with a trained ViT transformer.
2. Split the image into n number of patches of fixed size
3. Flattened the image patches into a sequence.
4. Perform linear dimensional embeddings from the flattened image patches
5. Add positional embeddings and class tokens
6. Feed the sequence as an input to the transformer encoder
7. Compute multi-head attention weight values in the transformer block.
8. Combining all the attention heads output and fed as input to MLP classifier to attain encode feature maps Z_n^i of the image

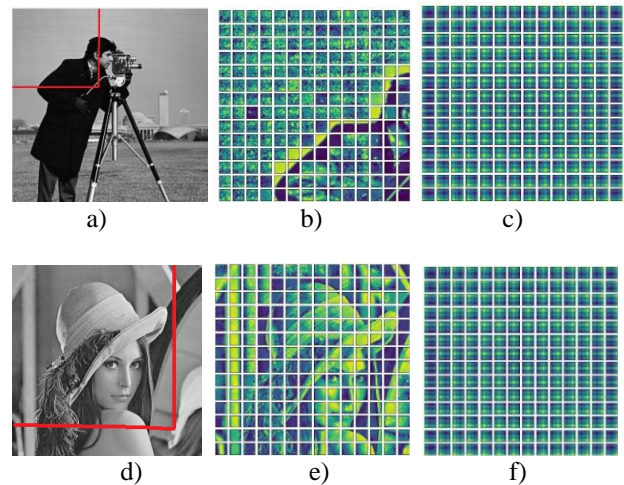


Figure 6: a) cameraman image, (red color marked portion patches shown in b), b) visualization of the patch, c) visualization positional embedding d) Lena image (red color marked portion patches shown in e), e) visualization of patch f) visualization positional embedding

Process 4: Watermark Embedding

1. Preprocess the cover image 'C' to remove noise and enhance the visual quality.
2. On the preprocessed image, the approximate coefficient LL_c and detailed coefficient LH_c, HL_c, HH_c obtained by applying curvelet transform

$$[AC_c, (D1_c, D2_c, D3_c, D4_c)] = CT(C) \quad (18)$$

3. On the approximate coefficient AC_c , high and low-frequency sub-band are obtained by applying Integer Wavelet Transform.

$$[LL_c, LH_c, HL_c, HH_c] = IWT(AC_c) \quad (19)$$

4. Embedding matrix S_c generated by applying the singular value decomposition method on the mid-frequency LH_c sub-band.

$$[U_c, S_c, V_c] = SVD(LH_c) \quad (20)$$

5. The authentication key K_a is embedded in the singular value of the cover matrix S_c and the inverse singular value decomposition method is applied to get back modified LH'_c .

$$S'_c = S_c + \alpha K_a \quad (21)$$

$$LH'_c = ISVD[U_c, S'_c, V_c] \quad (22)$$

where $\alpha=0.04$ is the robust strength factor in our case.

6. Split low-frequency sub-band coefficient LL_c into 2×2 blocks b_i , where, i represents the number of blocks, $i=1,2,\dots,n$

7. Apply Schur matrix decomposition on the blocks b_i

$$[U_{cb_i}, D_{cb_i}, U_{cb_i}^T] = Schur(b_i) \quad (23)$$

8. Replace the 2-LSB bit of each pixel in the cover image block D_c by a 6MSB bit of average eigenvalue Y_i .

9. At last, apply inverse Schur decomposition and integer wavelet transform

$$b'_c = Ischur(U_c, D'_c, U_c^T) \quad (24)$$

$$IWT[LL_c, LH'_c, HL_c, HH_c] = AC'_c \quad (25)$$

10. All the blocks are combined to form modified LL'_c sub-band

$$LL'_c = \sum_{b_i}^n b'_c \quad (26)$$

11. To embed the recovery watermark Z'_n , approximate coefficient values AC'_c of the curvelet transform are divided into 8×8 blocks.

12. The entropy region of each block ER^{b_i} is computed and high entropy regions are selected for embedding,

$$ER^{b_i} = -\sum_{k=1}^K P(z_k) \log P(z_k) \quad (27)$$

13. Embed the Vision transformer feature maps Z'_n in high entropy regions based on the adaptive strength factor.

14. All the blocks are combined which produces a modified curvelet transform

$$WM = ER^{b_i} + Z'_n * SF \quad (28)$$

15. Finally, inverse curvelet transform is performed to produce a watermarked image.

$$WM' = ICT(WM) \quad (29)$$

Process 5: Extraction Process

1. On the watermarked image WM' , primary authentication key has to be verified K_a ,

If key matches

Then proceed with the extraction process

Else

unauthentic and no extraction process.

2. Apply curvelet transform on the watermarked image WM' .

$$[AC_w, (D1_w, D2_w, D3_w, D4_w)] = CT(WM') \quad (30)$$

3. On the approximate coefficient values, apply 1st-level IWT

$$IWT(AC_w) = [LL'_w, LH'_w, HL'_w, HH'_w] \quad (31)$$

4. Apply SVD on the mid-frequency sub-band LH'_w

$$[U'_w, S'_w, V'_w] = SVD(LH'_w) \quad (32)$$

$$[U'_a * S'_a] = EK_a \quad (33)$$

5. Extract the principal component from LH'_w band

$$[U'_a, S'_a] = (S'_w - K_a) / \alpha \quad (34)$$

$$EK_a = U'_a \times S'_a \quad (35)$$

6. Verify image authentication as

$$A = EK_a - K_a \quad (36)$$

If A=0

Authentic

Else

Unauthentic then detect the tampered region as shown in steps 8-14

7. When the authentication process is valid, verify the tamper detection field from the low-frequency LL'_w sub-band by partitioning it into 4×4 blocks Wb_i

8. On each block b_i , apply Schur decomposition in order to extract the tamper detection watermark bit

$$[U'_w \times D'_w \times U'^T_w] = Schur(Wb_i^{LH_w}) \quad (37)$$

9. Extract 2LSB bit to verify every block by comparing it with the corresponding original 6MSB of the biometric image

10. Detect the tampered region, by matching the original 6MSB and the extracted 6MSB.

If matches

Declare as not tampered

Else

Tampered and extract feature maps

11. Apply the inverse schur decomposition method on each block $Wb_i^{LL_w}$, then combine all the blocks to obtain LL'_w .
12. To get back the approximate coefficient values

$$AC_w = \text{IIWT} [LL'_w, LH'_w, HL'_w, HH'_w] \quad (38)$$

13. To locate and recover the tampered region, the high entropy region is calculated using Eq.27 and repeat steps 10 & 11 as given in (Process 4) for each block $Wb_i^{LL_w}$ in AC_w block.
14. Extract the vision feature map watermark EZ'_n from the selected high entropy region $ER^{Wb_i^{LL_w}}$,

$$\text{EWM} = Wb_i^{LL_w} - EZ'_n / \text{SF} \quad (39)$$

15. Using the extracted feature maps localization of tampered region and recovery is performed.

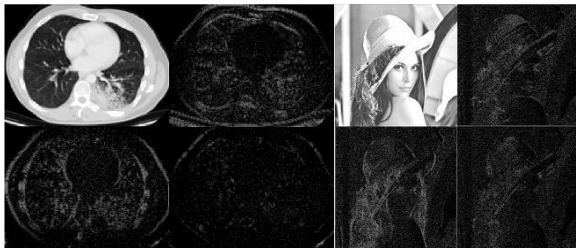


Fig. 7. 1-level Integer wavelet transform (chest CT-Scan & Lena image)

V. EXPERIMENTAL RESULT

The proposed model is evaluated against various attacks to verify the robustness and fidelity of the watermarked image. The fidelity of the image is checked by a quality metric named Peak-signal-to-noise-ratio (PSNR) and structural similarity index (SSIM). Whereas robustness is verified by the Normalized correlation coefficient (NCC) and Bit error rate (BER). The proposed model is validated for various noise attacks, median filter attacks, and geometric attacks.

A. Dataset Description

For generic images, we have used a benchmark dataset where some of the sample images are shown below in Figure 9. For medical images, we used the Kaggle chest CT-scan Dataset [41] to demonstrate the efficiency of the proposed watermarking scheme. In order to train the vision pre-trained transformer by combining our benchmark and Kaggle dataset and fine-tuning the model. The Chest CT-scan dataset has 1000 images and due to the minimal number of benchmark dataset images, we have augmented the image with various manipulation like rotation, splicing, and flipping. The additional advantage of augmenting the dataset is the feature maps will be invariant against those augmented attacks. We performed the experiments on images in terms of PSNRs, and NCCs metrics with our proposed scheme. Figure 8 and Figure 9 represents a few

sample images from the dataset, and Figure 10 represents a few owner sample database images.

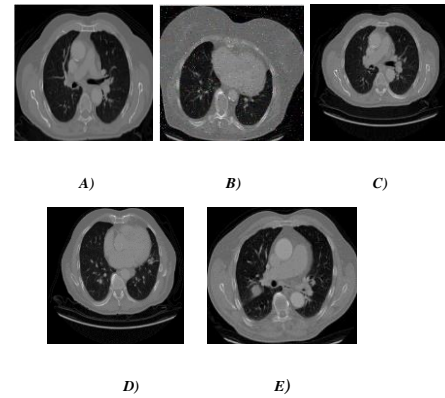


Fig. 8. Sample images of the Chest CT-Scan dataset

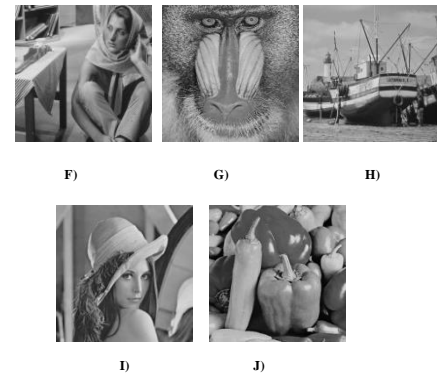


Fig. 9. Sample benchmark dataset

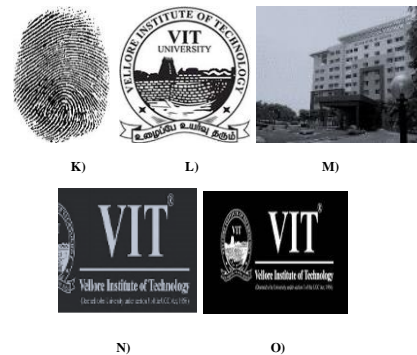


Fig. 10. Owner database K) fingerprint L) VIT Logo1 M) Silver Jubilee Tower N) VIT Logo2 O) VIT Logo3

B. Performance Analysis

The performance of the proposed model is evaluated using various metrics in order to measure the imperceptibility and robustness.

1) Imperceptibility

PSNR and SSIM metrics are generally employed to measure the similarity between original and watermarked images in terms of Imperceptibility.

PSNR (Peak Signal-to-Noise Ratio) [42]: PSNR metric which compares the similarity or distortion rate of the extracted watermarked image with the original watermarked image. The watermarked image is said to be of acceptable quality if the scoring rate is more than 25 to 30 dB. The PSNR is determined by using the following formula:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\sum_{M,N} (I_1(m,n) - I_2(m,n))^2} \right) \quad (39)$$

where M and N are the numbers of rows and columns in the input images.

Structural similarity index (SSIM): The SSIM measures the similarity index between watermarked image and the original image using the below Eq.4. and also, the distortion rate will be calculated. If the value is near 1, SSIM becomes effective.

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_1)} \quad (40)$$

where x and y are the original images and the watermarked image, μ_x and μ_y are, respectively, the local means of x and y , σ_x is the variance of x whereas σ_y is the variance of y , c_1 , and c_2 are two variables used to stabilize the division with weak denominator.

2) Robustness

The robustness of the extracted watermark against various attacks is measured using the Normalized Correlation Coefficient (NCC) and Bit Error Rate (BER). If the value is between 0 and 1, if it is nearer to 0 it is in the acceptable range.

Normalized Correlation Coefficient [10]: NCC measures the robustness between the original watermark and extracted watermark. NCC can be calculated using below shown equation:

$$\text{NCC} = \frac{\sum_{i=1}^{n_L} \sum_{j=1}^{n_K} (|W(i,j) + W'(i,j)|/2)}{n_L \times n_K} \quad (41)$$

where W and W' are the binary original and extracted watermark images, and n_L and n_K are the width and length of the host image, respectively.

Bit error rate (BER) [10]: BER is employed to measure the number of watermark bits extracted during watermark extraction divided by the total number of bits embedded.

$$\text{BER} = \frac{N_{\text{Err}}}{N_{\text{Bit}}} = 100 \frac{\text{Number of bit error}}{\text{Total no. of bit embedded watermark}} \quad (42)$$

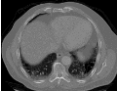
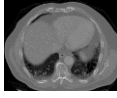



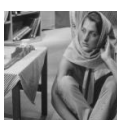




C. Attacks analysis

Attacks on a multimedia image may happen either intentionally or unintentionally. The efficiency of the vision transformer model is validated against various attacks on a watermarked image and determines the image's robustness and imperceptibility. There are several types of attacks: 1) No attacks, 2) Image processing attacks (salt and pepper noise, gaussian noise, speckle noise, poisson noise), 3) Geometric attacks (cropping, filtering (median filter, average filter), rotation, scaling, translation).

1) No attacks: The performance of embedding on the cover image is evaluated in terms of sensitivity and

robustness if there are no attacks on the watermarked image. Using the sample benchmark image from Table 1, evaluate the performance of the PSNR, and NCC. We have shown the imperceptibility of original and watermarked images for five sample images. The PSNR value ranges from 61 dB for the minimum to 50 dB for the maximum.

Table 1: Performance measured using PSNR and NCC for no attacks

Attacks	Original Image	Watermarked Image	Performance Measures
No Attacks			PSNR=61.24 dB NCC=0.999
			PSNR=59 dB NCC=0.99
			PSNR=58.5 dB NCC=0.998
			PSNR=58 dB NCC=0.99
			PSNR=59.2 dB NCC=0.999

2) Image Processing attacks: Various image processing unintentional attacks such as Salt and pepper noise (SP) with various densities, Gaussian noise (G) with a variance value, Speckle noise, and Poisson noise (P) are considered to evaluate the performance in terms of imperceptibility and robustness. Intentional attacks such as Median filtering (M) of sizes 3×3, and 5×5 and average filtering of 3×3 & 5×5, Content removal with tampering rates of 10%, and 30% respectively, splicing attacks, and Rotation attacks with various degrees. All the above attacks are considered to evaluate the performance of the proposed approach.

Table 2. Performance evaluated against Intentional and unintentional attacks (image processing attacks)

Unintentional Attacks				
Attacks	Attacked Image	Tamper Detection Accuracy	Recovered Image	Performance Measure
SN (density=0.02)		 TDA=99.9		PSNR=6.19 dB NCC=0.999
Salt and Pepper Noise (density=0.05)		 TDA=99		PSNR=6.002 dB NCC=0.993
Salt and Pepper Noise (density=0.09)		 TDA=98.2		PSNR=5.996 dB NCC=0.999
Salt and Pepper Noise (density=0.1)		 TDA=99		PSNR=5.801 dB NCC=0.991
Salt and Pepper Noise (density=0.5)		 TDA=99.01		PSNR=5.601 dB NCC=0.992
Gaussian Noise ($\alpha = 0.01$)		 TDA=100		PSNR=59 dB NCC=0.999
Gaussian Noise ($\alpha = 0.05$)		 TDA=99.7		PSNR=58.05 dB NCC=0.991
Gaussian Noise ($\alpha = 0.09$)		 TDA=99.01		PSNR=57 dB NCC=0.999
Gaussian Noise ($\alpha = 0.5$)		 TDA=99.8		PSNR=54.93 dB NCC=0.999
Speckle Noise ($\alpha = 0.02$)		 TDA=99.40		PSNR=58 dB NCC=0.999
Speckle Noise ($\alpha = 0.05$)		 TDA=98		PSNR=57.97 dB NCC=0.986
Speckle Noise ($\alpha = 0.09$)		 TDA=98		PSNR=57.5 dB NCC=0.972
Poisson		 TDA=99.99		PSNR=55 dB NCC=0.98

Intentional Attacks				
Attacks	Attacked Image	Tamper Detection Accuracy	Recovered Watermark Image	Performance Measure
Median Filter (3 × 3)		 TDA=98.24		PSNR=52 dB NCC=0.97
Median Filter (5 × 5)		 TDA=96		PSNR=54.2 dB NCC=0.973
Average Filter (3 × 3)		 TDA=96		PSNR=52.5 dB NCC=0.98
Average Filter (5 × 5)		 TDA=96		PSNR=52.02 dB NCC=0.98
Content removal (10%)		 TDA=99.9		PSNR=58.5 dB NCC=0.99
Content removal (30%)		 TDA=99.9		PSNR=57.5 dB NCC=0.991
Splicing/Cropping Attack 10%		 TDA=98.95		PSNR=56.7 dB NCC=0.999
Splicing Attack 50%		 TDA=99		PSNR=52 dB NCC=0.999
Rotation 25°		 TDA=99		PSNR=58.7 dB NCC=0.98
Rotation 45°		 TDA=99		PSNR=58.01 dB NCC=0.999
Rotation 65°		 TDA=99		PSNR=57 dB NCC=0.98

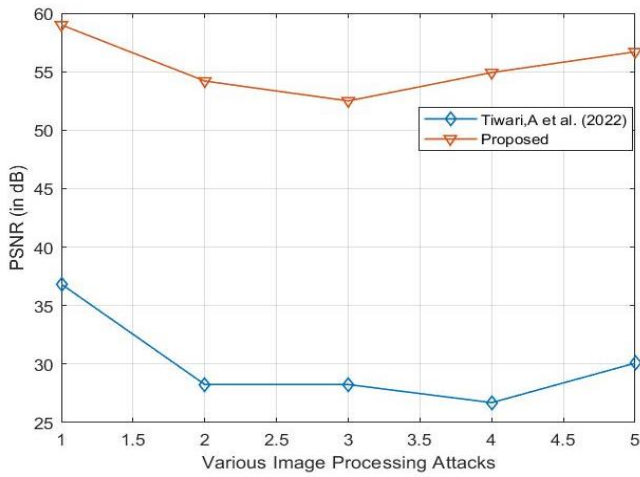


Fig. 11. Comparison of the PSNR metric of the proposed algorithm with the existing algorithm [27]

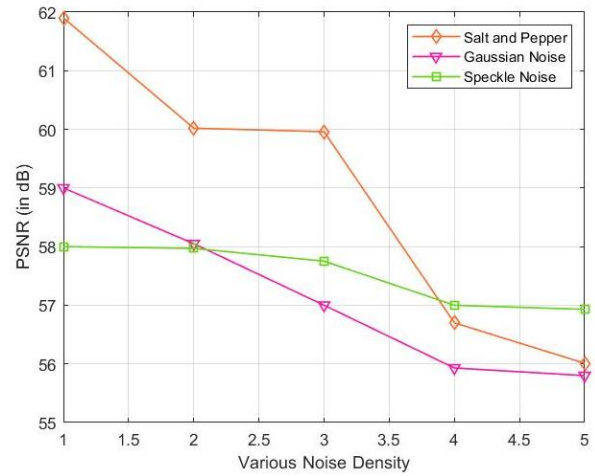


Fig. 14. Performance of the proposed algorithm for various noise - PSNR values

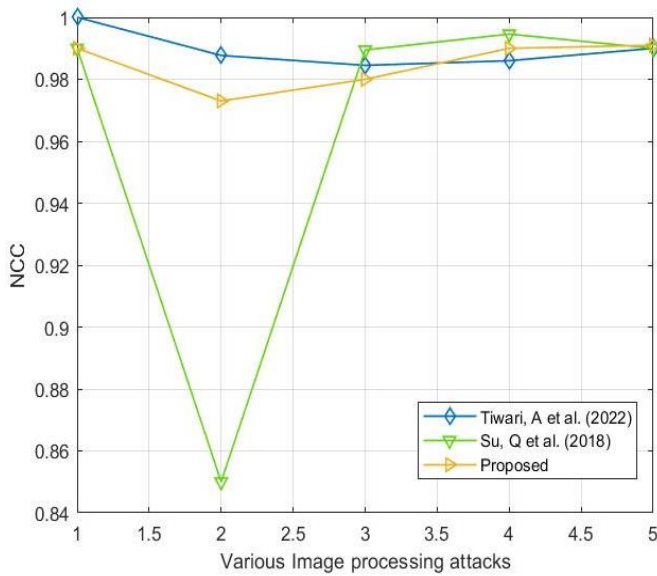


Fig. 12. Comparison of NCC metric of the proposed algorithm with the existing algorithm Tiwari, A et al. [27] & Su, Q [19]

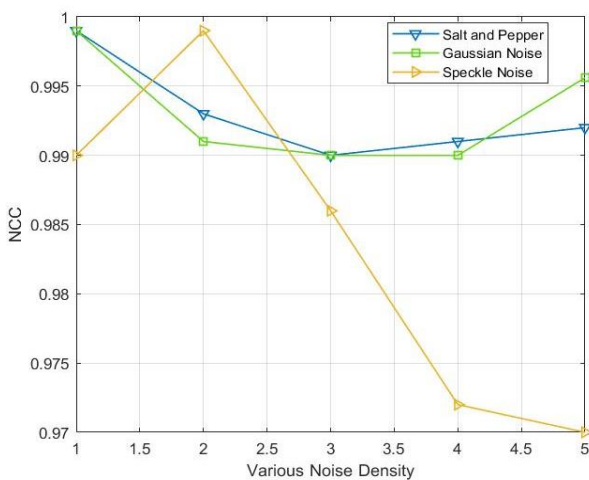


Fig. 13. Performance of the proposed algorithm for various noise - NCC values

We also evaluate the performance of the proposed hybrid algorithm with the existing lifting wavelet transform [27] and the result is shown in figure 11. Compared with the existing model hybrid watermarking algorithm has highly improved the imperceptibility as well as efficiency of the proposed algorithm. The efficiency of the proposed system is illustrated in figure 12 which compared the existing lifting Tiwari, A et al. [27] and schur decomposition method Qingtang Su et al. [19]. The efficiency of the proposed system is nearly closer to the existing Tiwari, A et al. [27] but the imperceptibility is higher than the Tiwari, A et al. [27] system. The result shown in table 2 and figure 14 illustrate salt & pepper noise attains high imperceptibility than speckle noise where speckle and salt & pepper noise is a little higher imperceptibility compared with Gaussian and Poisson noise. Figure 13 describes the efficiency of the algorithm against Salt & Pepper, Gaussian, and speckle noise from which it is observed that salt & pepper and gaussian noise are nearly close to 1 than speckle noise. In table 2 intentional attacks, content removal, median filter, average filter, copy-paste, and rotation attacks have achieved high tamper detection accuracy and recovery of the tampered image has shown better result with high imperceptibility and efficiency of about 0.999. From these results, it is inferred that the proposed algorithm can resist unintentional attacks as well as intentional attacks based on the tampering ratio.

Table 3. PSNRs, and NCCs comparison between proposed and existing algorithm

Images	Su, G. D [2]		Han, B [43]		Proposed	
	PSNR	NCC	PSNR	NCC	PSNR	NCC
Medical Image A	42.01	0.9999	12.7889	0.87595	59	0.991
Medical Image B	39.15	0.9998	16.637	0.87595	58.01	0.999
Benchmark Image F	44.25	0.9996	21.8267	0.8136	58.5	0.998
Benchmark Image G	45.08	0.999	21.3456	0.87509	56.01	0.999
Benchmark Image H	43.57	0.9998	28.1822	0.87823	57.97	0.986

Table 4. Comparison with Proposed Vs state-of-the-art watermarking techniques

Algorithm	Objective	Watermark generation method	Tamper detection	Network Training
Fragile Watermark [44]	Medical Image Tamper	Block average values	Turtle shell	No
Zero Watermarking [11]	Medical image security	VGG16-DFT	XOR	Yes
CNN + Attention [45]	Tamper detection on normal images	No	Local Interpretable Model-agnostic Explanations (LIME)	Yes
Proposed	Image security	Vision Transformer	Entropy, CT-IWT	Yes

Table 5. Comparison of performance of the dataset

Model	Dataset	Algorithm	Accuracy	Recall
Proposed	Medical Images + Benchmark	Vision Transformer	88%	0.94
Siteforge [45]	CASIA 2.0	DCNN + Attention	94.7%	0.98

Table 3, depicts the comparison of performance evaluation in terms of imperceptibility and robustness of the existing and the proposed algorithm. Table 4 describes the state-of-art algorithms of the proposed and existing systems. Table 5 shows the performance of the vision model on the benchmark and medical dataset was vision model has attained less accuracy compared with the real CASIA dataset. To improve the accuracy of the vision model, need to be trained with more features.

VI. CONCLUSIONS AND FUTURE WORK

The proposed vision transformer-based feature is embedded in the hybrid watermarking domain and attained high robustness with less distortion rate efficiently. A hybrid combination of curvelet integer wavelet transform has given a multi-directional invariant domain where the watermark is embedded in the singular coefficient values. The advantage of the matrix decomposition method is: when the watermark is embedded in the singular values matrix not vulnerable to any attack which increases the robustness of the proposed system. We fine-tuned the pre-trained vision transformer model on the medical and benchmark dataset which attained an accuracy of 88%. The feature maps are embedded in the high entropy region of the curvelet transform which added a value to improve the imperceptibility of the image. The PSNR and NCC value of the proposed model has attained a maximum of 61.9 dB and 0.999. Vision features maps have achieved better performance in tamper localization and recovery. In the future, this work can be extended using Multiple attentions for feature map generation and it can be embedded in the adaptive location using hybrid transforms.

References

- [1] S. A. Parah, J. A. Sheikh, F. Ahad, N. A. Loan and G. M. Bhat, "Information hiding in medical images: a robust medical image watermarking system for E-healthcare," *Multimedia Tools and Applications*, vol. 76(8), pp. 10599-10633, 2017.
- [2] G. D. Su, C. C. Chang and C. C. Lin, "Effective self-recovery and tampering localization fragile watermarking for medical images," *IEEE Access*, vol. 8, pp. 160840-160857, 2020.
- [3] Y. Huang, W. Lu, W. Sun and D. Long, "Improved DCT-based detection of copy-move forgery in images," *Forensic science international*, Vols. 206(1-3), pp. 178-184, 2011.
- [4] I. Cox, M. Miller, J. Bloom, J. Fridrich and T. Kalker, "Digital watermarking and steganography," *Morgan Kaufmann*, 2007.
- [5] M. Hamidi, M. El Haziti, H. Cherifi, and M. El Hassouni, "Hybrid blind robust image watermarking technique based on DFTDCT and Arnold transform," *Multimedia Tools and Applications*, vol. 77(20), pp. 27181-27214, 2018.
- [6] Q. Dai, J. Li, U. A. Bhatti, Y. W. Chen and J. Liu, "SWT-DCT-based robust watermarking for medical image," *Innovation in Medicine and Healthcare Systems, and Multimedia*, Springer, Singapore, pp. 93-103, 2019.
- [7] F. Tohidi, M. Paul and M. R. Hooshmandasl, "Detection and recovery of higher tampered images using novel feature and compression strategy," *IEEE Access*, vol. vol. 9, pp. 57510-57528, 2021.
- [8] A. Anand and A. K. Singh, "An improved DWT-SVD domain watermarking for medical information security," *Computer Communications*, vol. vol. 152, pp. 72-80, 2020.
- [9] L. Agilandeewari and K. Ganesan, "A robust color video watermarking scheme based on hybrid embedding techniques," *Multimedia Tools and Applications*, vol. 75(14), pp. 8745-8780, 2016.
- [10] L. Agilandeewari and K. Ganesan, "RST invariant robust video watermarking algorithm using quaternion curvelet transform," *Multimedia Tools and Applications*, vol. 77(19), pp. 25431-25474, 2018.
- [11] L. Agilandeewari, K. Ganesan and K. Muralibabu, "A side view based video in video watermarking using DWT and Hilbert Transform," *In Security in Computing and Communications: International Symposium,SSCC*, pp.

- [12] S. P. Vaidya, "Fingerprint-based robust medical image watermarking in hybrid transform," *The Visual Computer*, pp. 1-16, 2022.
- [13] L. Agilandeewari and K. Ganesan, "An efficient hilbert and integer wavelet transform based video watermarking," *Journal of Engineering Science and Technology*, vol. 11(3), pp. 327-345, 2016.
- [14] B. I. Hongbo, L. I. Xueming and Y. Zhang, "A novel hvs-based watermarking scheme in contourlet transform domain," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11(12), pp. 7516-7524, 2013.
- [15] H. S. Alshanbari, "Medical image watermarking for ownership & tamper detection," *Multimedia Tools and Applications*, vol. 80(11), pp. 16549-16564, 2021.
- [16] J. Dafni Rose, K. Jaspin and K. Vijayakumar, "Lung cancer diagnosis based on image fusion and prediction using CT and PET image," *Signal and Image Processing Techniques for the Development of Intelligent Healthcare Systems*, Springer, vol. 2021, pp. 67-86, 2021.
- [17] J. Mo, Z. F. Ma and Q. L. Huang, "An adaptive watermarking scheme using SVD in Contourlet domain," *Advances in Information Sciences and Service Sciences*, vol. 4(15), pp. 221-232, 2012.
- [18] N. Zermi, A. Khaldi, R. Kafi, F. Kahlessenane and S. Euschi, "A DWT-SVD based robust digital watermarking for medical image security," *Forensic Science International*, vol. 320, pp. 110691, 2021.
- [19] Q. Su, Z. Yuan and D. Liu, "An approximate Schur decomposition-based spatial domain color image watermarking method," *IEEE Access*, vol. 7, pp. 4358-4370, 2018.
- [20] H. Luo, F. X. Yu and Z. L. & L. Z. M. Huang, "Blind image watermarking based on discrete fractional random transform and subsampling," *Optik*, vol. 1, pp. 311-316, 2011.
- [21] T. K. Tsui, X. P. Zhang and D. Androustos, "Color image watermarking using multidimensional Fourier transforms," *IEEE Transactions on Information Forensics and security*, vol. 3(1), pp. 16-28, 2008.
- [22] J. Molina, V. Ponomaryov, R. Reyes and C. Cruz, "Watermarking-based self-recovery and authentication framework for colour images," *2019 7th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1-6, 2019, May.
- [23] A. Abdelhakim, H. I. Saleh and M. Abdelhakim, "Fragile watermarking for image tamper detection and localization with effective recovery capability using K-means clustering," *Multimedia Tools and Applications*, vol. 78(22), pp. 32523-32563, 2019.
- [24] X. Yu, C. Wang and X. Zhou, "Review on semi-fragile watermarking algorithms for content authentication of digital images," *Future Internet*, vol. 9(4), p. 56, 2017.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [26] E. Campos-Ponce and M. .. Cedillo-Hernandez, "Tamper detection and localization in color images using secure block-based watermarking," *In 2022 IEEE Mexican International Conference on Computer Science (ENC).IEEE*, pp. 1-7, 2022.
- [27] A. Tiwari and V. K. Srivastava, "Integer Wavelet Transform and Dual Decomposition Based Image Watermarking scheme for Reliability of DICOM Medical Image," *In 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON). IEEE*, pp. 1-6, 2022.
- [28] L. Zhang and D. Wei, "Image watermarking based on matrix decomposition and gyator transform in invariant integer wavelet domain," *Signal Processing*, vol. 169, p. 107421, 2020.
- [29] Y. Luo, L. Li, J. Liu, S. Tang, L. Cao, S. .. Zhang and Y. Cao, "A multi-scale image watermarking based on integer wavelet transform and singular value decomposition," *Expert Systems with Applications*, vol. 114272, p. 168, 2021.
- [30] M. Rezaei and H. Taheri, "Digital image self-recovery using CNN network," *Optik*, vol. 264, p. 169345, 2022.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ... and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [32] S. N. Mohanrajan and A. Loganathan, "Novel vision transformer-based bi-LSTM model for LU/LC prediction—Javadi Hills, India," *Applied Sciences*, vol. 12(13), p. 6387, 2022.
- [33] L. Agilandeewari and K. Ganesan, "A robust color video watermarking scheme based on hybrid embedding techniques," *Multimedia Tools and Applications*, vol. 75(14), pp. 8745-8780, 2016.
- [34] "DWT Process," <https://www.researchgate.net/publication/331168576/figure/fig1/AS:727678764199949@1550503545624/Sub-bands-separated-by-a-three-level-dyadic-discrete-wavelet-transform-DWT.png>.
- [35] L. Agilandeewari and K. Muralibabu, "A robust video watermarking algorithm for content authentication using discrete wavelet transform (DWT) and singular value decomposition (SVD)," *International Journal of Security and Its Applications*, vol. 7(4), pp. 145-158, 2013.
- [36] A. M. Hammouche and H. M. El-Bakry, "A New FDCT-USFFT and FDCT-Wrap Algorithms for Image Contrast Enhancement," *International Journal of Artificial Intelligence and Mechatronics*, vol. 18, p. 89, 2017.
- [37] E. Candes, L. Demanet, D. Donoho and L. Ying, "Fast discrete curvelet transforms," *multiscale modeling & simulation*, vol. 5(3), pp. 861-899, 2006.
- [38] A. Khaldi, M. R. Kafi and M. S. Moad, "Wrapping based curvelet transform approach for ECG watermarking in telemedicine application," *Biomedical Signal Processing and Control*, vol. 75, p. 103540, 2022.
- [39] L. Agilandeewari and K. Ganesan, "An efficient hilbert and integer wavelet transform based video watermarking," *Journal of Engineering Science and Technology*, vol. 11(3), pp. 327-345, 2016.
- [40] "Kaggle-Chest CT scan Images," <https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>.
- [41] A. Soualmi, A. Alti and L. Laouamer, "An Imperceptible Watermarking Scheme for Medical Image Tamper Detection," *International Journal of Information Security and Privacy (IJISP)*, vol. 16(1), pp. 1-18, 2022.
- [42] B. Han, J. Du, Y. Jia and H. Zhu, "Zero-watermarking algorithm for medical image based on VGG19 deep convolution neural network," *Journal of Healthcare Engineering*, 2021.
- [43] G. D. Su, Chang, C. C. and C. C. Lin, "Effective self-recovery and tampering localization fragile watermarking for medical images," *IEEE Access*, vol. 8, pp. 160840-160857, 2020.
- [44] B. Singh and D. K. Sharma, "SiteForge: Detecting and localizing forged images on microblogging platforms using deep convolutional neural network," *Computers & Industrial Engineering*, vol. 162, p. 107733, 2021.

- [45] Agilandeewari, L., Manoharan. P & Alenizi, F.A. (2023). A robust semi-fragile watermarking system using Pseudo-Zernike moments and dual tree complex wavelet transform for social media content authentication. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-023-15177-4>
- [46] Agilandeewari, L., & Meena, S. D. (2023). SWIN transformer-based contrastive self-supervised learning for animal detection and classification. *Multimedia Tools and Applications*, 82, 10445 – 10470, <https://doi.org/10.1007/s11042-022-13629-x>
- [47] Mohanrajan, S.N., Loganathan, A. Novel Vision Transformer–Based Bi-LSTM Model for LU/LC Prediction—Javadi Hills, India. *Appl. Sci.* 2022, 12, 6387. <https://doi.org/10.3390/app12136387>

Author Biographies



Aberna P is pursuing her Ph.D. at the School of Information Technology and Engineering, Vellore Institute of Technology Vellore. She completed her M.Tech in Software Engineering from Vellore Insitute of Technology in the year 2018. Her research interest includes Digital Image Watermarking, Image and Video Processing, Multimedia Security, Digital forensics, Machine Learning, and Deep Learning.



Agilandeewari L completed her Ph.D. and working as a Professor in the School of Information Technology & Engineering (SITE), VIT Vellore. She received her Bachelor's degree in Information Technology and Master's in Computer Science and Engineering from Anna University in 2005 and 2009 respectively. She is having around 20 years of teaching experience and published 70+ papers in peer-reviewed reputed journals. Her reputed publications include research articles in peer-reviewed journals namely *Expert Systems with Applications*, *IEEE Access*, *Journal of Ambient Intelligence and Humanized Computing*, *Multimedia Tools and Applications*, and *Journal of Applied Remote Sensing* indexing at Thomson Reuters with an average impact factor of 5. She is a peer reviewer in journals including *IEEE Access*, *Pattern Recognition*, *International Journal of Remote Sensing*, *Array*, *Artificial Intelligence Review*, *Informatics in Medicine Unlocked*, *Neurocomputing*, *Computers*, and *Electrical Engineering*, *Journal of King Saud University–Computer and Information Sciences*, *IET ReView*, *Journal of Engineering Science and Technology (JESTEC)*, etc. She also published about 13 engineering books as per Anna University Syllabus. Her areas of interest include Image and video watermarking, Image processing, Neural networks, Cryptography Fuzzy Logic, Machine Learning, IoT, Information-Centric Networks, and Remote Sensing.



Aashish Bansal is pursuing his B.Tech. at the School of Information Technology and Engineering, Vellore Institute of Technology, Vellore. He is working as Academic Intern Tech at Bank of America Continuum India since January 2023. His research interest includes Digital Image and Video Processing, Artificial Intelligence, Machine Learning, Deep Learning, Digital Forensics, and Network and Information Security.