Dilesh Khairwar, 10:5
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

# A Survey on Image Water Marking Techniques and Attacks

**Dilesh Khairwar, Asst. Prof. Sumit Sharma**
Department of Computer Science & Technology,
Vaishnavi Institute of Technology and Science,
Bhopal, MP, India

**Abstract- Digital information may be transferred from one location to another with the least amount of difficulty than any other media. Text, music, video, and image data, among other types of data, can all be transferred using the same media and the same methods. However, certain precautions have been made by the owner of the data by embedding some signature or validating information at the receiver end. The security of these data is greatly dependent on the protocols. This article does a comprehensive review of several approaches to the protection of digital image data that have been offered by researchers. In the study, signature embedding techniques and their attributes were broken down in detail in order to better the reader's grasp of the subject field. In this research, different network assaults that may have an effect on the data that was received were also elaborated upon. The study also provided an explanation of the various aspects that researchers make use of to secure digital data. This is because each feature has its own significance and area of use that varies according to the type of image and the attacks that are being made.**

**Keywords- Image Processing, Stenography, Feature Extraction, Data hiding.**

## I. INTRODUCTION

The acronym "DS" stands for "digital signature," which is an electronic signature that can be used to safeguard digital records or digital contracts. Its goal is to authenticate the document, and by extension, the participants to an agreement, much in the same way that a traditional signature would.

During the process of moving the data from the sender to the receiver, it is used to check and make sure that the original data has not been changed in any way. It is also now required to authenticate users on a regular basis in order to maintain safety and prevent fraud. Since DS cannot be reproduced by anyone else, it offers protection to its users. Essentially, it ensures the legitimacy of the message and gives the recipient the peace of mind that it came from a sender who is already known to them. Important information can be preserved through the usage of digital photographs.

However, ensuring the authenticity and integrity of these photographs is a difficult undertaking because they are frequently sent across unsecure networks like the internet. The contents of these digital photos can be easily modified in this day and age thanks to the rapid and advanced technology that is available. Therefore, it is necessary to protect these images against the various attempts that are made to manipulate them, and it is important to make an effective method to solve the problem of image authentication, which is ensuring the integrity of an image, in particular for document images such as significant certificates, scanned cheques, art drawings, signed documents, circuit diagrams, design draughts, and so on.

An increased interest in digital watermarking can be attributed to the growing concern for the protection of intellectual property. People are interested in downloading photographs, music, and videos since the internet is, for the most part, a user-friendly

location where content can be easily accessed. The internet offers a distribution strategy that is both effective and economical to use. The amount of time necessary to acquire a variety of media via the internet is a minute fraction of the amount of time necessary to acquire the same media by going to a physical store. In addition, if one purchases media via the internet, the purchaser only needs virtual space to store the media in question, as opposed to the physical space required to put the media on a shelf or in any other location suitable for such material. On the other hand, such rapid availability makes it possible for individuals to violate copyrights.

## II. RELATED WORK

In [4] In this research, an unique technique for protecting the copyright of colour images via watermarking in the spatial domain is described. The goal of the technique is to do so in a quick and efficient manner. First, the direct current (DC) coefficient of the 2D-DFT that was obtained in the spatial domain will be discussed, and then the relationship between the change of each pixel in the spatial domain and the change in the DC coefficient in the Fourier transform will be proven.

This will be done by proving that the DC coefficient is proportional to the change in each pixel. After that, the DC coefficient is what is utilized by the suggested quantization technique in order to embed and extract the watermark in the spatial domain. There are three main contributions that are novel about this paper: 1) the DC coefficient of 2D-DFT is obtained in the spatial domain without of the true 2D-DFT; 2) the relationship between the change of each pixel in the image block and the change of the DC coefficient of 2D-DFT is found; and 3) the proposed method has the short running time and strong robustness. 1) The DC coefficient of 2D-DFT is obtained in the spatial domain without of the true 2D-DFT. 2) The relationship between the change of each pixel in the image block and It is necessary to replace the hashing information since the value of the hash must be preserved in order to locate the embedded blocks.

A approach for watermarking has been proposed by Mohammed A. M. Abdullahet. al. in [5]. This technique involves embedding binary data in the DCT middle band frequency area. In this study, the image was cropped to a fixed size, and the co-ordinate values for the fixed middle band of the DCT were swapped according to the watermark bit. The swapping of this was dependent on a set of criteria, and it was necessary to keep those parameters exactly the same at the receiver side in order to extract the watermark. This work suffers from low watermark absorption and has a low level of resilience against spatial attacks.

For the purpose of embedding an image in the edge region of the image, Kazuki Yamato and colleagues have developed a between-class variance notion in [6]. The author began by applying the discriminate analysis method for the purpose of transforming the image into a binary format. Next, the author applied the BCV method, which categorizes pixels into edge and non-edge regions. A modification in the image's spatial region was carried out in order to conceal data in the image edge region. In this case, there is a limited amount of space available for the watermark, whereas the hidden information can be easily uncovered if it is embedded in the edge region.

Watermarks are generated by Angela Piper et al. in [7] using simply the input image, and they are embedded in the low frequency region of the image. The technique of fragile watermarking, which was proposed in this research, is intended to protect photos from JPEG compression attacks. This paper has not covered any other kinds of attacks, and the amount of time it took to do this work was also rather lengthy.

**Hanieh Khalilian et al. [8],** A fractal code-based self-reconstruction technique was proposed, with the idea being that the input image would be sent to an extremely noisy area. Therefore, it was considered that there was a loss of information, which was remedied by adding additional packets of fractal code. Hashing the hash key protected the tempering of the image as well as the secret information it contained. This work was able to improve robustness in a lossy environment, albeit at the cost of additional bandwidth and increased computational complexity.

A technique called Singular value Decomposition was proposed by the author in [9] in order to locate data that is similar to the original image. The authors of this article take an image and cut it up into patches of a fixed size, and then they replace those patches with KSVD patches. This not only encrypts the

watermark before it is embedded, but it also increases the image's level of security while it is being transmitted over a network. The time that was taken here was spent searching for the correct patch in the KSVD library.

The storage of dictionaries on either the transmitter or the recipient side proved cumbersome. CNN made use of in order to integrate the data for the watermark into the original image. It was discovered, with the assistance of a few pieces of supplementary data, that the watermarking had been removed from the image. Here, it was discovered that the watermarking as well as the image were inverted once they reached their destination.

Back Propagation neural network in wavelet domain is used in the unique blind watermarking technology that was proposed by Huang et al. [10]. In order to achieve greater imperceptibility and robustness, a scrambled watermark has been inserted in this paper by taking advantage of the Human Vision System (HVS). A neural network is utilized in order to memories the connection that exists between the embedded watermark and the image that has also been watermarked.

**Peng et al. [11]** have come up with an original method for image watermarking that is based on SVM and takes place in the multi-wavelet domain. For the purpose of watermarking, the algorithm has made use of a specific frequency band as well as the property of the image. Although the scheme is relatively resistant against a variety of assaults, it is not strong enough to reach the desired level of robustness when faced with average filtering, median filtering, JPEG attacks, or scaling attacks.

**Yang et al. [12]** have also presented a reliable method for correcting geometric distortion by utilizing fuzzy SVM in the domain of undecimated discrete wavelet transform (UDWT). The method, despite the fact that it offers sufficient resilience, necessitates an excessive amount of processing time and, in addition, it is not robust to local geometric distortions.

In [13], the Lifting Fourier transform (LFT) at the third level was utilized to insert the watermark. In a feed-forward neural network, the feature set that was generated from the blocks in which the reference watermark RW was contained was used as the input feature vector. The target vector is determined by using the bits of RW that correspond to it.

The method offers an adequate degree of robustness when confronted with a variety of attacks, including noise generation and removal attacks, some geometry attacks, and others.

In paper [14], the adaptive scaling factor was proposed. This factor is based on selected DWT-DCT coefficients of the image's content. The adaptive scaling factor was calculated by analyzing the relationship between a few carefully chosen DWT-DCT coefficients and the overall average value of all of the DWT-DCT coefficients. A suggested set of criteria that take into account the adaptive scaling factor was utilized in the process of embedding the watermark image.

In the paper [15], the authors suggest a technique for numerical relational databases called Genetic Algorithm and Histogram Shifting Watermarking (GAHSW), which is a watermarking method that is both resilient and reversible. The genetic algorithm is used to pick the best possible secret key for the grouping database, so that the watermarking can be implanted with a level of distortion that is balanced with the capacity of the database. In order to embed the watermark in a manner that is both reliable and secure, the histogram of the prediction error has been modified. The resilience of the job is diminished when histogram shifting is used.

The authors of [16] provide a digital watermarking technique designed specifically for neural networks. The author poses a whole new problem, which is the incorporation of watermarks into neural networks using an approach that is based on the discrete cosine transform (DCT).

Combining DWT with DCT could result in significant performance improvements for algorithms that are based on the discrete wavelet transform (DWT). These improvements could be used for digital picture watermarking. The paper also describes the requirements, embedding conditions, and attack forms that are used for watermarking throughout the neural networks. Because the watermark is incorporated into the host network at the same time that it is being trained, the method that is provided here does not have any impact on the performance of a network that has a watermark.

## III. DATA SECURITY ALGORITHM PROPERTIES

Creating universal digital image security involves taking a number of critical elements into consideration, including the following:

Fidelity refers to the notion that people should not be able to detect the presence of a watermark on a picture, and that images should not be altered either before or after the watermarking process.

The ability of an image to remain intact despite being subjected to a number of different processing techniques is referred to as its robustness. These kinds of assaults are typically carried out with the purpose of disrupting the watermark in order to carry out the behaviour that was intended. Attacks of this type include cryptographic assaults, removal assaults, geometric assaults, and protocol assaults, to name a few examples [21]. The watermarking algorithms are vulnerable to a variety of different types of assaults. Although comprehensive watermarking is not necessary in all applications, it is required in certain of the applications that are now available.

The concept of data payload, often known as capacity, refers to the capacity of an image to conceal a certain amount of bits without degrading the image's overall quality. Another thing to think about is how much information may be saved in a picture and then easily extracted if it ever becomes necessary. Embedding capabilities are subject to change based on the programme being used.

The capacity of an image to endure pressure from the outside world is referred to as its security. The watermarking system needs to have a high enough level of security such that unauthorised users who are unable to decipher the algorithm cannot access the information they are trying to access. The watermark ought to be able to be removed only by a person who can be relied upon. [20]

The amount of time necessary to both extract and embed the watermark is referred to as the computational complexity of the process. Some real-time applications are quite speedy, but in situations where a high level of security must be maintained, it might be time-consuming to apply complex algorithms.

This metric, known as "perceptibility," reflects the degree to which the quality of an image is diminished as a result of the watermark being implanted. When using a technique for invisible watermarking, it is optimal to keep this parameter as low as is practically possible. [13]

Imperceptibility is a concept that describes the invisibility that is required in watermarking systems. This feature requires that the original image and the watermarked image have the same appearance [20]. This can be achieved by reducing either the capacity or the robustness of the system, or both. The Peak Signal to Noise Ratio (PSNR) and the Structural Similarity (SSIM) index [13] are the fundamental benchmarks that are used to measure the perceptuality of every image.

## IV. FEATURE FOR DIGITAL IMAGE

The image is a matrix of light intensity values, and these intensity values indicate a variety of colours. The colour feature is represented by these intensity values. Therefore, the ability to identify an object's colour is an important feature, and a cheap computation cost is a key aspect of this feature's properties.

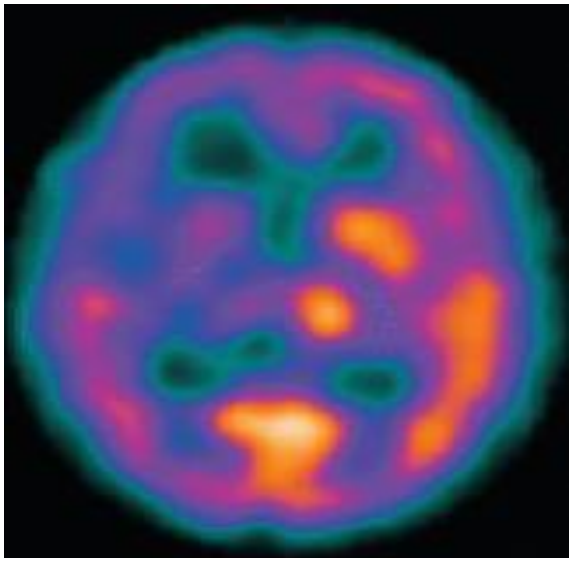

Fig 1. Represent the HSV (Hue Saturation value) format of an image.

There are many distinct picture files accessible, each with its own unique colour format. For example, photographs can be saved in a variety of colour formats, such as RGB, which stands for red, green, and blue. This is a depiction of a single image in

three dimensions, where each hue is represented by a two-dimensional matrix, and the collection of those matrices tries to give the impression of a third dimension. In order to calculate the intensity of each pixel, the grayscale format that is used is a two-dimensional space with values that range from 0 to 255.

In the case of binary format, which is a colour matrix with just the values 0 and 1, binary matrices are black and white. Faces have been successfully identified in [8] with the assistance of this colour characteristic.

**1. Edge Feature:**

An picture is a collection of intensity values, and when there is a dramatic change in those values, an essential new feature known as the Edge appears in the image, as depicted in figure 4. This property can be utilized for the detection of various types of visual objects, such as highways, buildings, and other similar elements [5]. There are a number of algorithms that have been created that successfully point out all of the images within an image or frames that are Sobel, perwitt, canny, and other similar images. Canny edge detection is one of the best algorithms out of them, and it is one of the best methods to locate all of the possible boundaries of a picture.

**2. Texture Feature**:

The term "texture" refers to the degree to which one surface differs from another in terms of its intensity, and it enumerates qualities such as smoothness and regularity [1]. In contrast to the colour space model, the texture model calls for an additional processing step. The aspects of the texture that are based on colour are less susceptible to changes in illumination, and this is also true of the edge features.

**3. Histogram Feature:**

This phase uses the image vector that was created after the pre-processing step. The histogram of the image is located in one of the bins in this step. This can be understood by assuming that the colour scale in figure 4.2 ranges from 1 to 10, and then doing a count on each pixel value in the image.

Therefore, based on the information presented above, the vector Hi equals [0, 0, 0, 4, 3, 5, 2, 1 2, 0], where H represents the count of colour pixel values and I represents the position in the H matrix that corresponds to a colour value.

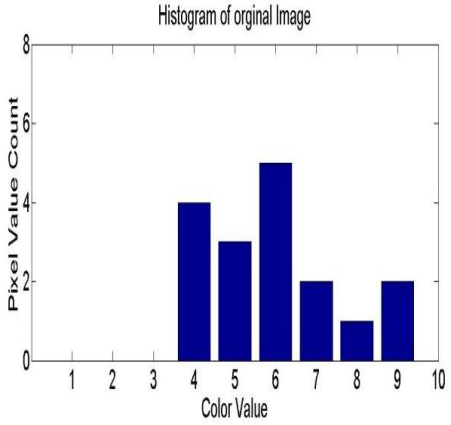


Fig 3. Histogram of the original image.

**4. Corner Feature**:

In the event that the camera is moving, it is necessary to determine the difference between the two frames, and this distinction is indicated by the corner feature in the image or frame. This allows the video frames to remain stable. Therefore, in order to detect resizing of the window in the original view, one must first locate the corner location of the two frames. This function can also be used to determine the angles between the object in the two separate frames, in addition to the distance between them. Because each of them represents a point in the image, it may be used to track the thing that is the focus of attention.

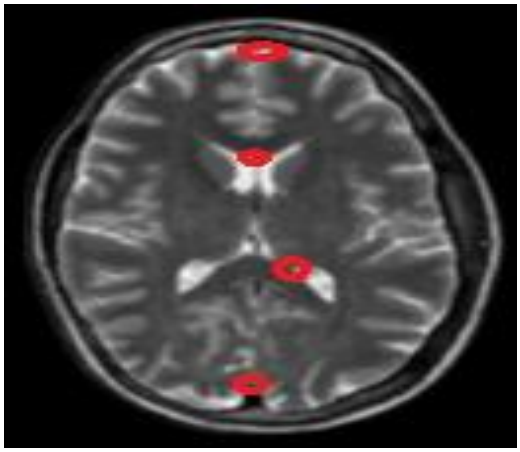

Fig 3 Represent the corner feature of an image with green point.

**5. DWT (Discrete Wavelet Transform):**

LL: Figure 3's upper left section is referred to as the LL block. After filtering the picture rows using a low pass filter, this block of the image is obtained by passing the same data through the low pass filter again, but this time the columns are filtered for the analysis. Because this version of the image does not contain any edge information, it is referred to as an

approximate version of the image. This block contains the flat region of the image.

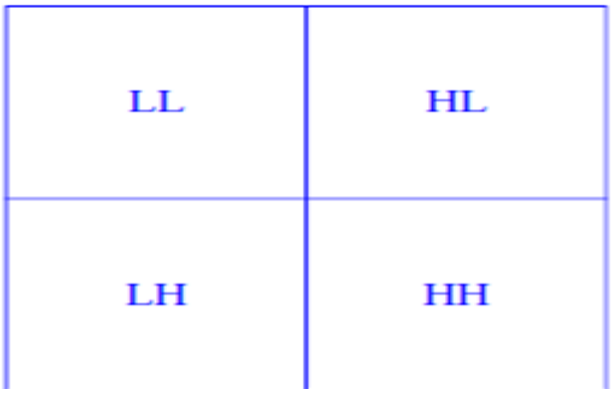


Fig 3.11 DWT of image from [8].

HL: The area labelled as the HL block can be found in figure 3's upper right corner. This section of the image was obtained by first filtering the image rows with a high pass filter and then doing the same filtering process on the image using a low pass filter, with the columns being filtered for analysis. The horizontal edge region of the image that is contained within this block does not have any flat information.

LH: The lower left section of figure 3 is referred to as the LH block. This section of the image was obtained by first filtering the image rows using a low pass filter and then passing those filtered rows through a high pass filter. Finally, the image columns were filtered so that they could be analysed. This block contains the portion of the image along the vertical edge that does not contain any flat information.

HH: The area labelled as the HH block can be found in figure 3's lower right corner. After filtering the image rows with a high pass filter, this block of the image is obtained by passing the same data through the high pass filter again, but this time the columns are filtered for the analysis. This block contains the image's diagonal edge region, which is devoid of any flat information and contains that region.

## VI. ATTACK ON IMAGE

When compared to wired networks, wireless sensor networks (WSNs) that are deployed in severe environments face a greater number of risks. In addition, the public communication protocol that is chosen by WSNs makes the risk of physical tampering much greater [22, 23]. The fact that the detected node has limited processing capabilities as well as energy resources makes it more challenging to design security mechanisms. The primary forms of assault are broken down into the following five categories:

- Packet tampering is when a malicious node is added to a WSN, and then that node tampers with the value of the packets and then transmits the tampered packets, which can result in extremely serious repercussions in certain specific circumstances.
- Packet forging occurs when a malicious node is joined to WSNs and it continues to send phoney packets to other nodes. This results in a significant increase in the amount of network traffic and causes the WSNs as a whole to waste energy.
- Selective forwarding occurs when a malicious node is added to WSNs. This node deletes partial packets and selectively transmits other packets to their destinations. The loss of data could lead to a disastrous circumstance in which the sink node is unable to provide the appropriate reaction.
- Packet replay: a rogue node added to WSNs forwards the packets that have already been sent, one more or repeatedly to other nodes, which will create the traffic congestion and waste of energy.
- Transfer delay: a malicious node that has been added to WSNs will forward the packets later than the predetermined time, which will cause the sink node to drop the packets because of the timestamp.

## VII. CONCLUSIONS

This research presents evidence to support the contention that the process of data concealing, which consists of message embedding and extraction, intuitively transfers onto the encoder-decoder network design. This architecture divides the learning model into two networks. These models involve the training of an encoder network to incorporate input messages into picture data.

After the images have been put through various forms of assault using distortion layers, the decoder network is tasked with recovering the initial message from the corrupted version of the image. Some examples of these distortions are blurring, cropping, compression, and many others. In the near future,

academics will work in this research field with the goal of improving network training in order to reduce distortion. This objective takes into account the differences that exist between the cover image and the encoded image, in addition to the differences that exist between the embedded and extracted input messages.

## REFERENCES

[1] X. Sun, J. Su, B. Wang, and Q. Liu, "Digital watermarking method for data integrity protection in wireless sensor networks," International Journal of Security and Its Applications, vol. 7, no. 4, pp. 407–416, 2013.

[2] David Ifeoluwa Adelani, Haotian Mai, Fuming Fang, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. 2019. Generating Sentiment-Preserving Fake Online Reviews Using Neural Language Models and Their Human- and Machine-based Detection. (2019).

[3] Beijing Chen, Jiaxin Wang, Yingyue Chen, Zilong Jin, Hiuk Jae Shim, and Yun-Qing Shi. 2020. High-Capacity Robust Image Steganography via Adversarial Network. KSII Transactions on Internet & Information Systems 14, 1 (2020).

[4] Qingtang Su, Decheng Liu, Zihan Yuan, Gang Wang, Xiaofeng Zhang, Beijing Chen, And Tao Yao. "New Rapid and Robust Color Image Watermarking Technique in Spatial Domain". IEEE Access March 25, 2019.

[5] Mohammed A. M. Abdullah, Satnam S. Dlay, Wai L. Woo, and Jonathon A. Chambers. "A Framework for Iris Biometrics Protection: A Marriage between Watermarking and Visual Cryptography". IEEE Access Year: 2016, Volume: 4 Pages: 10180 – 10193.

[6] Kazuki Yamato, Madoka Hasegawa, Yuichi Tanaka‡ and Shigeo Kato. "Digital Image Watermarking Method Using Between-Class Variance". 978-1-4673-2533-2/12/$26.00 ©2012 IEEE.

[7] Angela Piper1, Reihaneh Safavi-Naini. "Scalable Fragile Watermarking For Image Authentication". Published In IET Information Security, On 31st December 2012.

[8] Paweł Korus, Student Member, IEEE, and Andrzej Dziech. "Efficient Method for Content Reconstruction with Self-Embedding". IEEE Transactions On Image Processing, Vol. 22, No. 3, March 2013.

[9] Hanieh Khalilian, Student Member, IEEE, And Ivan V. Bajic Video "Watermarking With Empirical PCA-Based Decoding" IEEE Transactions On Image Processing, Vol. 22, No. 12, December 2013.

[10] S. Huang, W. Zhang, W. Feng and H. Yang, Blind watermarking scheme based on neural network, Proceedings of the 7th IEEE World Congress on Intelligent Control and Automation (2008), 5985–5989.

[11] H. Peng, J. Wang and W. Wang, Image watermarking method in multiwavelet domain based on support vector machines, Journal of Systems and Software 83(8) (2010), 1470–1477.

[12] H.Y. Yang, X.Y. Wang and C.P. Wang, A robust digital watermarking algorithm in undecimated discrete wavelet transform domain, Computers and Electrical Engineering 39(3) (2013), 893–906.

[13] MohiulIslama,∗, Amarjit Roy b and Rabul Hussain Laskar. "Neural network based robust image watermarking technique in LWT domain". Journal of Intelligent & Fuzzy Systems 34 (2018) 1691–1700.

[14] Ferda Ernawan, Dhani Ariatmanto, and Ahmad Firdaus. "An Improved Image Watermarking by Modifying Selected DWT-DCT Coefficients", March 29, 2021..

[15] Donghui Hu, Dan Zhao, ShuliZheng. "A New Robust Approach for Reversible Database Watermarking With Distortion Control". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2019.

[16] R. S. Kavitha, U. Eranna, M. N. Giriprasad. "An Image Compression Based Technique to Watermark a Neural Network". (IJITEE) ISSN: 2278-3075, Volume-9 Issue-4, February 2020

[17] Suresh Kuri, Gururaj Kulkarni. "Robust Digital Image Watermarking using DWT, DCT and Probabilistic Neural Network". International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017.

[18] Linfeng Geng, Weiming Zhang, Haozhe Chen, Han Fang, and Nenghai Yu. 2020. Real-time attacks on robust watermarking tools in the wild by CNN. Journal of Real-Time Image Processing 17 (6 2020).

[19]] Seung-Min Mun, Seung-Hun Nam, Haneol Jang, Dongkyu Kim, and Heung-Kyu Lee. 2019. Finding robust domain from attacks: A learning framework for blind watermarking. Neuro computing 337 (2019), 191–202.

International Journal of Science, Engineering and Technology

An Open Access Journal

[20] Mahdi Ahmadi, Alireza Norouzi, S. M. Reza Soroushmehr, Nader Karimi, Kayvan Najarian, Shadrokh Samavi, and Ali Emami. 2020. ReDMark: Framework for Residual Diffusion Watermarking on Deep Networks. Expert Systems with Applications 146 (2020).

[21] Srivastava Kumar Sumit, Pandey Harikesh. "Medical Image Watermarking with Patient Details as Watermark". International Journal of Advance research, Ideas and Innovations in Technology, Volume2, Issue6, 2016.

[22] A. Gutub and M. Al-Ghamdi, "Hiding shares by multimedia image steganography for optimized counting-based secret sharing," Multimedia Tools and Applications, vol. 79, no. 11-12, pp. 7951–7985, 2020.

[23] K.-H. Jung, "Data hiding scheme improving embedding capacity using mixed PVD and LSB on bit plane," Journal of Real-Time Image Processing, vol. 14, no. 1, pp. 127–136, 2018.

[24] Jain M, Lenka SK (2016) Diagonal queue medical image steganography with rabin cryptosystem. Springer Brain Inform 3(1):39–51.