Machine Learning approach of Artificial Immune optimization for Security of Digital Data Pawan

Singh Rajput[1]
Research Scholar
Department of Computer Science & Engineering, LNCT University, Bhopal, Madhya Pradesh, India

Dr. Megha Kamble[2]
Associate Professor
Department of Computer Science & Engineering, LNCT University, Bhopal, Madhya Pradesh, India

Abstract:

Digital data ease the communication by reducing the cost, time, resources, labor etc. At the same time chance of attack is also increases as medium of communication is open to all and some time access as well. In order to increase the security of data various measures were adopt and proposed by researchers. This paper has also developed a model that provide security to digital image data b embedding it in other image. For selection of position of embedding work has adopt a clustering approach of data mining. Artificial immune genetic algorithm was adopted by the paper to cluster the embedding coefficient into two classes. As images have different pixel arrangement hence dynamic algorithm is highly desirable for selection of embedding position. Experiment was done on real image dataset and comparison of model was done on different evaluation parameters. Result shows that proposed model improved the PSN, SNR value of the image after embedding and extract information in different attack environment as well.

Keywords: Digital Image, Image Processing Frequency Feature, Data Embedding, Genetic Algorithm

DOI Number: 10.14704/nq.2022.20.9.NQ44451 **Neuro Quantology 2022; 20(9):3953-3961**

I. INTRODUCTION

The amount of data being transferred and stored on the Internet, as well as the breadth of its application of digital media, are both growing at an exponential rate. As a direct consequence of this, owners of digital data are able to rapidly and extensively distribute multimedia documents via the Internet. Because of this, there is a great deal of interest in the protection of multimedia copyrights and multimedia security [1]. As digital medical images are viewed, extracted, and transmitted, they are subject to possible manipulation both inside and outside of the protected environment of the medical system. Because of this, there is a risk of a breakdown in trust, which has the potential to affect the accuracy of image reading and diagnosis.

In the past, methods such as encryption and access control were utilized in order to ensure the security of the media. After the media have been successfully transmitted and decrypted, these techniques, however, do not protect against data copies that have been modified or newly created. The focus of this paper is on the application of digital techniques to the safeguarding of digital multimedia. The process of embedding digital data has been identified as a potentially useful method for ensuring the confidentiality and authenticity of images through the utilisation of various techniques. Data hiding is a method of efficient private communication that involves embedding secret messages into various forms of media, such as digital images, videos, and audios, in order to effectively protect the security of multimedia information [2]. [Citation needed] These methods ensure that the image is not tampered with, authenticated, that its content has been verified, and that it is integrated [3]. Displaying the data that has been watermarked or converting it into other file formats is not a very

straightforward method for removing a watermark. Because of this, it is possible to obtain information about the transformation after an attack by looking at the watermark. It is essential to be able to differentiate between digital watermarking and other technologies such as encryption [4]. Digital image watermarking techniques allow for the survivability of a variety of processes, including the conversion from digital to analogue, compression, changes in file format, re-encryption, and decryption. Because of these tasks, it can serve as an alternative to cryptography or as a complement to it. Because the information is hard-coded into the content, it is not possible to delete it through normal usage [3].

Data mining is the process of extracting useful information and knowledge from large amounts of imperfect, noisy, chaotic, and random application data [4, 5]. This information and knowledge may be hidden, unknown, or both. Data mining is capable of performing a wide variety of functions, including identification grouping, customization, prediction, cultural analysis, isolation analysis, and parallel analysis. The question of how to accurately extract buildings has always been at the forefront of contemporary architectural research. In this paper, we make the case for incorporating the invariant moment algorithm into the process of extracting features from images. Image processing technology, feature extraction methods, and segmentation techniques are the subjects of investigation and analysis in this paper.

II. Related Work

Zhang et al. [6] introduced a brand new technology that they called FFDNet. This system is capable of addressing a wide variety of noise that is present in excessive levels. Denoising is accomplished in a timely and appropriate manner by FFDNet through the utilisation of a CNN model that is analogous to that of DnCNN. However, this method does not permit the accurate identification of noises. A bidirectional downsampling operator and adjustable noise-level mapping are used in this design to convert the original image that serves as input into four different images that can then be used in the CNN layers. Invisibility and durability are two of the most significant obstacles that must be overcome when analysing images that have been watermarked.

The adaptive scaling factor was proposed by F. Ernawan and colleagues in [7], and it is based on selected DWT-DCT coefficients of the image's content. The adaptive scaling factor was calculated by analysing 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 rules that take into account the adaptive scaling factor was utilised in the process of embedding the watermark image.

A new blind and fragile method of watermarking was proposed by F. Tohidi and colleagues in [8]. This method can detect tampering and improve the recovery of tampered images. In order to enhance the quality of the watermarked images as well as the recovered ones, a new feature extraction method has been devised. This method will produce a recovery code that is concise yet thorough, and it will do so by utilising a new compression technique. In the event that a block in the image is altered, the proposed embedded feature will make it possible to recover the original data by extracting it. In order to eliminate the possibility of tamper coincidence, the data that is watermarked on each block includes not only the recovery code that is specific to the block itself, but also the data that belongs to its neighbours as a second layer of recovery. Different sized blocks were tested to see how well they performed and compared to one another in terms of how effectively they recovered an image after being tampered with at varying rates.

Karaboga and Akay [82] examined the performance of the ABC optimization method in comparison to other optimization approaches, such as GA and PSO, and came to the conclusion that the performance of ABC is superior to that of other optimization approaches.

In order to improve the robustness and quality of the DWT-SVD based watermarking method, Ali and Ahn [9] used a modified version of the DE (self-adaptive differential evolution) algorithm to optimise the scaling factors of the watermark data. This was done in order to achieve better results. In addition, Ali et al. [89] used a DE optimization method to search for

optimal scaling factors with the goal of improving the quality of the watermarked image as well as the watermark's robustness.

Aslantas [10] described a technique that embeds the watermark by manipulating the singular values (SVs) of the host image with the help of multiple scaling factors. This technique was reported by Aslantas. In addition, DE is utilised to optimise the scale factors in order to improve the performance of the proposed method. In a later section, Aslantas [94] uses the DE method to optimise the scaling factor parameters in order to achieve the highest possible level of robustness and transparency.

Lei et al. [11] proposed a method in which they used DE to optimally design the quantization steps (QSs) for calculating the strength of the watermark in order to achieve good robustness and quality. In this method, the DE was used to calculate the strength of the watermark. Method of Creating an Artificial Bee Colony One of the most recent developments in the field of swarm-based algorithms is the Artificial Bee Colony (ABC) optimization method. There have only been a very small handful of watermarking methods reported that make use of the potency of ABC for optimization in their processes to this day.

Akay and Karaboga [12] presented a survey of the current state of the art in the field of image processing as well as some of its applications. Sha et al. [13] improves the performance of their proposed image watermarking method by using the ABC method to optimise pixel-by-pixel embedding at various frequency sub-bands with DWT. This helps the method work more effectively.

III. Proposed Methodology

This section has brief the embedding methodology of the proposed model. In order to data embedding image was transform into frequency domain by applying discrete wavelet transformation. Whole explanation of model was done by the fig 1. For improving the robustness of data against various attack proposed model has choose random location for embedding in coefficient values. Selection of position was obtained by the use of artificial immune genetic algorithm, steps of genetic algorithm was shown in fig. 2.

Digital Image PreProcessing Input image was preprocessed by selecting the red region of image for transforming the image. As image has different dimension so conversion or transforming image from one to other was done by this step of work [2, 3]. Input digital data that need to be embedded is transformed into binary format. One more advantage of this conversion is that, input digital data is either image, voice, video it works for all.

Feature Extraction: Red region of image was used for the embedding of the image hence feature was extract form this portion of image. Embedding in a region portion increase the volume of the secret data storage, as same number of maximum binary information is stored in other region of image. One more advantage is changes in one and other remain unaffected maintains the PSNR value of image. Wavelet transformed features were extract from the image. As wavelet gives four equal matrix of different set of edges, so this work uses only low frequency region of image [14].

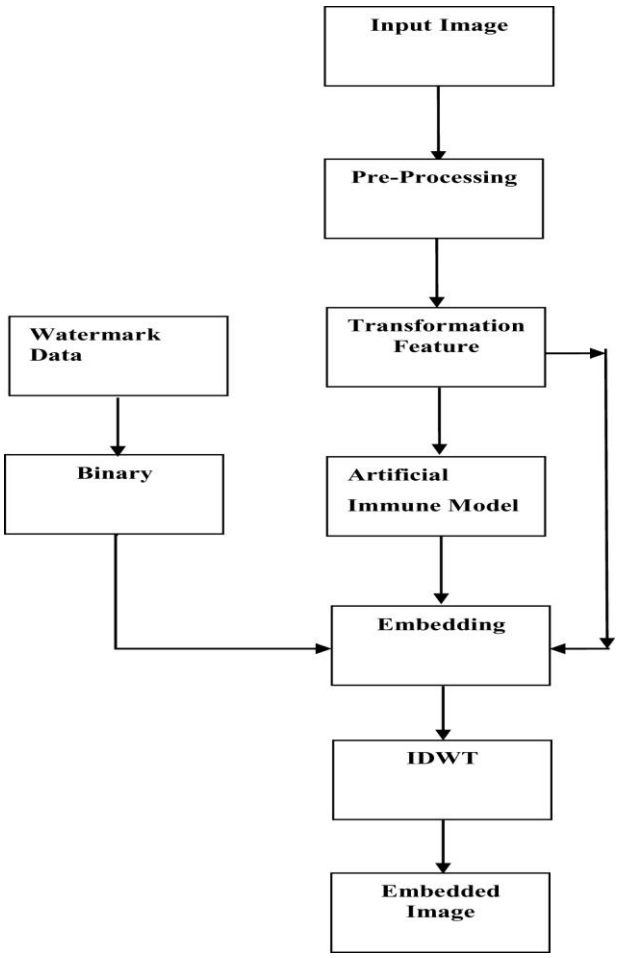

Fig 1 Proposed model embedding of watermark.

Artificial Immune Algorithm

This algorithm was based on the human body working against any disease. So antibodies are probable protein that may fight otherwise other set of antibodies were generate. So algorithm goal is to find quire the human body [15].

Generate Antibodies

Whole work finds good positions in the image for embedding data. So coefficient values present in the image were cluster into two group one is for embedding and is not use for embedding. To cluster coefficient cluster centers were used in the model. Coefficient range values were randomly selected as a antibody in the work. Collection of these random set of coefficient is population in the genetic algorithm For random coefficient value generation Gaussian distribution function was used.

PA←Generate_Coefficient_Antibody(c, m)

Above function is gives m number of antibody set, where each set has c number of coefficients. PA is population of antibodies used for getting the final cluster center in the work.

Affinity

Fitness of the antibody estimate by the affinity function. For this work has estimate the minimum coefficient difference summation. Each antibody coefficient value was compared with each coefficient of image and minimum difference value was summed to the affinity value of the antibody. Similarly all set of antibodies present in population were estimate by Eq. shown below.

$$A = \sum_{i=1}^{m} min(|PA[i,n] - I|)^n$$

Cloning

As per affinity value of each antibody in population, best solution PA_b is obtained. As per best antibody PA_b coefficient set few antibodies features were transform [16]. This cloning of coefficient values from one antibody to other is done in this step.

$$PA \leftarrow Cloning(PA_b, A)$$

In order to avoid over learning only single position cloning was applied in the work.

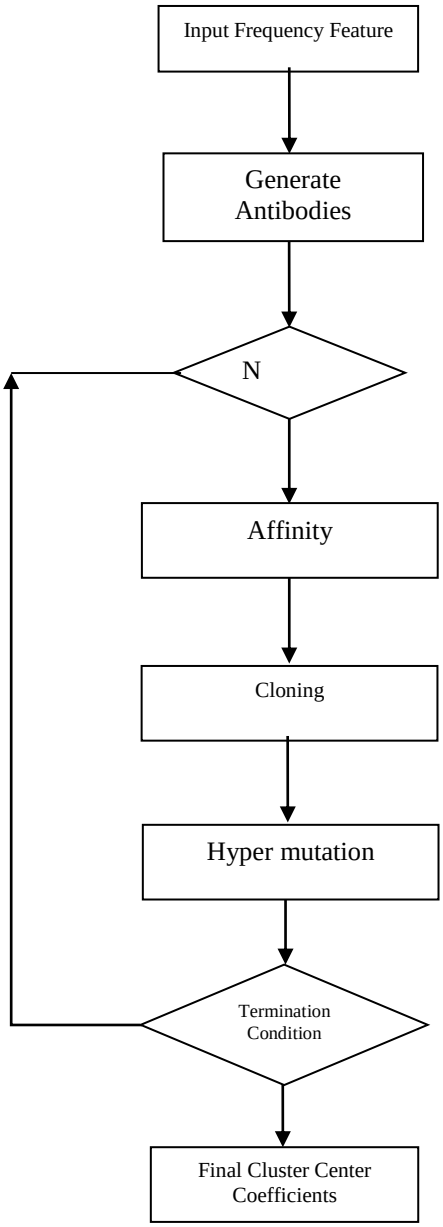

Fig. 2 Block diagram of proposed algorithm.

Hypermutation

Each antibody then undergoes self learning where random coefficient value of the antibody is change. This change was done by the hypermuation. All steps improved the work efficiency and chance of getting the best solution is also improved.

$$PA \leftarrow Hypermutation(PA)$$

Final cluster Center

Once algorithm reach to its maximum number of iterations artificial immune genetic algorithm gives the cluster center. Each antibody was finally pass in the affinity function that gives fitness. Based on the affinity value coefficient values were cluster into two group. Cluster having lower affinity value is used for the embedding of the digital data.

Binary Data Embedding

Embedding of watermark bit was done in fit cluster coefficient. Least significant bits were replace as per watermark information. Each fit coefficient value was transform into binary vector. In order to preserve the information against attacks paper has replace more than one bit LSB in single coefficient binary vector. This increases the capacity of watermark to embed more data. Each embedded coefficient binary vector is transform back to decimal number format. Once all data bits get embed then inverse wavelet transformation is perform to get embedded image.

Proposed Embedding Algorithm

Input: I // Image

Output: EI // Embedded Image

1. PI←Pre_Process(I) // PI: Preprocessed Image
2. [LL LH HL HH]←Extract_Feature(PI)
3. A←Generate_Antibody(LL, m)
4. Loop 1: its // Iterations
5. I←Immunity(A,LL)
6. A←Cloning(A, I)
7. A←Hypermutation(A)
8. EndLoop
9. I←Immunity(A)
10. CC←Best(A,LL) // Cluster Coefficients
11. Loop b:1:W // W: number of secret data bits
12. Ec ←Coefficient_Selection (CC, LL) // Ec: Embedding Coefficient
13. Fv ←Embedding(W[b], Ec)
14. EndLoop
15. EI←Inverse([LL LH HL HH])

IV. Experiments and Results

Implementation of proposed model was done on MATLAB software. Experimental was performed with real image set on the developed MATLAB code files. In order to compare proposed model results were compared with existing method proposed in [17].

Comparison of methods of embedding was done on different evaluation parameters shown below:

Peak Signal to Noise Ratio: The Peak Signal to Noise Ratio (PSNR) between the images OI and WI which are of size M x N is given by the following expression. Higher is the PSNR; higher is the similarity between the images. It is expressed in dB.

$$\text{PSNR} = 10\log_{10}\left(\frac{Max\_pixel\_value}{Mean\_Square\_error}\right)$$

Normalized Correlation: Normalized Correlation (NC) The Normalized Correlation (NC) between the images WM and EM which are of size m x n is given by the following expression. Its value ranges in the interval [0 1], closer the NC value to 1 indicates higher is the correlation between the two images.

$$NC = \frac{\sum_{i=1}^{s} W * W'}{\sqrt{\sum_{i=1}^{s} W^2 * \sum_{i=1}^{s} W'^2}}$$

In above formula s is number of pixels in the image, W is pixel value of the original watermark and W' is the pixel vale of extracted watermark. Value of NC range from 0 to 1. Correlation between W and W' is is high if NC moves towards 1.

Mean Square Error

$$MSE = \frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}$$

Where X_{obs} are original cover image pixel values and X_{model} was extracted the image.The smaller the means average error, the closer to the ground truth values.

Extraction Rate

$$\eta = \frac{n_c}{n_a} \times 100$$

Here N_C is true pixels count from extracted watermark and n_a is a total number of pixels present in the original watermark.

Results

Table 1. Image watermarking PSNR value Based Comparison under ideal condition.

Images	AIAIW	DWT-DCT coefficients [17]
Lena	55.6258	49.8413
Mandrilla	55.6258	44.8633
Tiger	55.7610	45.6843
Tree	55.7294	50.1257
Pepper	55.7829	50.8727

Table 2 Image watermarking MSE value Based Comparison under ideal condition.

Images	AIAIW	DWT-DCT coefficients [17]
Lena	0.178	0.6744
Mandrilla	0.178	2.122
Tiger	0.1726	1.8702
Tree	0.1738	0.6317
Pepper	0.1717	0.5319

Table 3. Image watermarking NC value Based Comparison under ideal condition.

Images	AIAIW	DWT-DCT coefficients [17]
Lena	1	1
Mandrilla	1	0.948
Tiger	1	0.963
Tree	1	0.984
Pepper	1	1

Table 1, 2 and 3 shows that proposed AIAIW model has maintain the image quality after embedding of watermark. It was obtained that use of genetic algorithm for embedding of watermark bit in selected coefficient of DWT region has increase the embedding quality.

Table 4 Image watermarking PSNR value Based Comparison under noise attack.

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	9.0635	5.2266
Mandrilla	7.86	5.7991
Tiger	9.4589	7.315
Tree	6.3302	4.7167
Pepper	8.9321	5.9902

Table 5. Image watermarking MSE value Based Comparison under noise attack

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	8067.3	19517
Mandrilla	10640	17107
Tiger	7365.2	12066
Tree	1513.8	21949
Pepper	8315.1	16371

Table 6. Image watermarking PSNR value Based Comparison under filter attack

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	32.303	30.28
Mandrilla	21.3111	20.8371
Tiger	34.5936	19.8711
Tree	29.5361	28.2013
Pepper	31.6947	30.1967

Table 7 Image watermarking MSE value Based Comparison under filter attack

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	38.2566	60.9499
Mandrilla	480.8037	536.2505
Tiger	22.5798	122.089
Tree	72.3559	98.3907
Pepper	44.0157	62.1451

Table 4, 6 of noise, filter attack values for different watermarking algorithms. Values shows that proposed AIAIW has increases evaluation parameter of PSNR by applying artificial immune algorithm. Tables 5, 7 shows that MSE error was also reduced by the work as compared to other existing algorithm.

Table 8 Image watermarking extraction rate value Based Comparison under filter attack

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	1	0.1260
Mandrilla	1	0.649
Tiger	1	0.5
Tree	1	0.807
Pepper	1	0.43

Table 9. Image watermarking NC value Based Comparison under noise attack

Images	AIAIW	DWT-DCT Coefficients [17]
Lena	1	0.6730
Mandrilla	1	0.633
Tiger	1	0.78
Tree	1	0.71
Pepper	1	0.677

Table 8, 9 noise, filter attack values for different watermarking algorithms. Values shows that proposed AIAIW has increases evaluation parameter of NC value by applying artificial immune algorithm.

Above result section shows that proposed AIAIW model has increases the various evaluation parameters under ideal and attack environment. Use of Artificial immune algorithm for coefficient selection of transformed features has increases the PSNR value in all set of images, while in [17] embedding was done in all coefficients. This selected embedding process reduces the information loss in the image and increases the robustness against various attacks.

V. CONCLUSIONS

Images have important role in different section of life like, social, medical diagnosis, environmental science, etc. Images are easy to modify by different tools and methods, hence reality check or authentication required. This work has developed a robust image watermarking approach by using artificial immune algorithm. Embedding was done on low

frequency region of image. Embedding algorithm requires minimum displacement of original low frequency coefficient. Proposed AIAIW model has maintain the image quality after embedding of watermark. It was obtained that use of genetic algorithm for embedding of watermark bit in selected coefficient of DWT region has increase the watermarking quality. Experiment was done on real dataset of images. Results shows that proposed model has increases the robustness of image under noise and filter attack. Work has maintain the PSNR value by 13.33% as compared to DWT-DCT Coefficients [17]. Work has maintain the PSNR value under filter attack by 13.41% as compared to DWT-DCT Coefficients [17], average watermark extraction was high by 0.02% as compared to algorithm proposed in [17].

References

1. Hurrah, Nasir N., et al. "Dual watermarking framework for privacy protection and content authentication of multimedia." Future generation computer Systems, 94 (2019): 654-673.
2. Hu, Runwen, and Shijun Xiang. "Cover-Lossless Robust Image Watermarking Against Geometric Deformations." IEEE Transactions on Image Processing, 30 (2020): 318-331.
3. Wenjie Lin, Xueke Zhu, Wujian Ye, Chin-Chen Chang, Yijun Liu, Chengmin Liu, "An Improved Image Steganography Framework Based on Y Channel Information for Neural Style Transfer", Security and Communication Networks, vol. 2022, Article ID 2641615, 12 pages, 2022
4. Singh, O. P., and A. K. Singh. "A robust information hiding algorithm based on lossless encryption and NSCT-HD-SVD." Machine Vision and Applications, 32.4 (2021): 1-13.
5. J. E. Lee, J. W. Kang, W. S. Kim, J. K. Kim, Y. H. Seo, and D. W. Kim, "igital Image Watermarking Processor Based on Deep Learning," Electronics, vol. 10, no. 10, p. 1183, 2021.
6. K. Zhang, W. Zuo, and L. Zhang, "FFDNet: toward a fast and flexible solution for CNN-based image denoising," IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608–4622, 2018.
7. F. Ernawan, D. Ariatmanto and A. Firdaus, "An Improved Image Watermarking by Modifying Selected DWT-DCT Coefficients," in IEEE Access, vol. 9, pp. 45474-45485, 2021
8. F. Tohidi, M. Paul and M. R. Hooshmandasl, "Detection and Recovery of Higher Tampered Images Using Novel Feature and Compression Strategy," in IEEE Access, vol. 9, pp. 57510-57528, 2021.
9. [82] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," Applied Mathematics and Computation, vol. 214, pp. 108–132, 2009.
10. [85] M. Ali and C. W. Ahn, "An optimized watermarking technique based on self-adaptive de in dwt–svd transform domain," Signal Processing, vol. 94, pp. 545–556, 2014.
11. [88] V. Aslantas, "Optimal svd based robust watermarking using differential evolution algorithm," in Proc. of World Congress of Engineering, London, UK, 2008.
12. [90] B. Lei, E. L. Tan, S. Chen, D. Ni, T. Wang, and H. Lei, "Reversible watermarking scheme for medical image based on differential evolution," Expert Systems with Applications, vol. 41, pp. 3178–3188, 2014.
13. [91] F. Sha, F. Lo, Y. Y. Chung, X. Chen, and W. C. Yeh, "A novel optimized watermark embedding scheme for digital images," in Proc. of MultiMedia Modeling. Springer, 2015.
14. Manish Kumar, Asst. Prof. Shivraj Singh. "Image Security of Water Marking Digital Image Based on based on SVD and DWT with Haar Transform". IJSRET, Volume 8 issue 1, 2022.
15. Dai Yongshou, Li Yuanyuan, Wei Lei, Wang Junling, Zheng Deling. "Adaptive immune-genetic algorithm for global optimization to multivariable function". Journal of Systems Engineering and Electronics, Volume 18, Issue 3, 2007.
16. Heder S. Bernardino & Helio J. C. Barbosa. "Artificial Immune Systems for Optimization". Part of the Studies in

Computational Intelligence book series (SCI,volume 193)

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