

Light Weight CNN based Robust Image Watermarking Scheme for Security

R. Dhaya,

Professor,

Department of Computer science and Engineering,

King Khalid University,

Kingdom of Saudi Arabia,

dhayavel2005@gmail.com

Abstract- In recent years, digital watermarking has improved the accuracy and resistance of watermarked images against many assaults, such as various noises and random dosage characteristics. Because, based on the most recent assault, all existing watermarking research techniques have an acceptable level of resistance. The deep learning approach is one of the most remarkable methods for guaranteeing maximal resistance in the watermarking system's digital image processing. In the digital watermarking technique, a smaller amount of calculation time with high robustness has recently become a difficult challenge. In this research study, the light weight convolution neural network (LW-CNN) technique is introduced and implemented for the digital watermarking scheme, which has more resilience than any other standard approaches. Because of the LW-CNN framework's feature selection, the calculation time has been reduced. Furthermore, we have demonstrated the robustness of two distinct assaults, collusion and geometric type. This research work has reduced the calculation time and made the system more resistant to current assaults.

Keywords: *digital water marking, convolution neural network*

1. INTRODUCTION

Digital watermarking refers to the process of sub repetitiously incorporating and extracting the information from a image cover. To produce a marked image, the data are hidden behind a portrait. The tagged image does not visually show the watermark and only authorized

recipients may properly extract the information from watermark [1]. Image watermarking techniques can be utilized for different purposes. In several formats, information required for watermarks may be provided according to various target circumstances, such as random bits or electronic signatures in order to secure and authenticate image or secret communications [2]. Figure 1 shows the visible and invisible watermarking picture.



Figure 1 visible & invisible watermarking

In addition, error corrections codes may be used in a cyber-attack to encrypt the watermark for different purposes, such as improving encryption security or refurbishing information reliability [3, 4]. Images are threatening to filter and denoise, because they restore their original value to each pixel of the image. Recent experiments have been conducted for watermarking the images with deep neural networks [5]. For example, it is challenging to properly utilize the fit of deep neural networks to learn and generalize the integration and extraction processes of watermarks [6].

Furthermore, it is possible to encode the watermark for various objectives, such as adding encryption security or restore information integrity using mistake correction codes in a cyber-attack [7]. Filtering and denoising images are hazardous since it restores their original value for each pixel of the image. There have been recent studies on the watermarking of images with deep neural networks, however, the challenges remain. For instance, the fit of deep neural networks is difficult to fully utilize it to learn and generalize the processes of embedding and extraction of watermarks [8, 9]. Figure 2 shows the simplified block diagram of watermarking scheme.

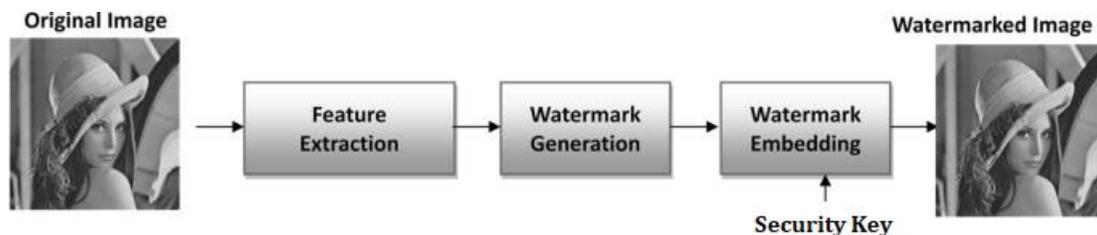


Figure 2 simple blind watermarking scheme

It might be unspecified or time-consuming to identify the reality of the ground for aquatic image work. Lastly, without knowing the bad cases, it is unknown to concurrently acquire resilience and blindness [10, 11]. Modern watermarking research mainly focuses on multi-bit scenarios that collect all the information from the communicative watermarks to allow many applications in the trends [12]. In general, a photo watermarking process should take into account various aspects such as the integrity of the picture and the detection of watermarks for computer analysis [13, 14].

A wide range of algorithms are being developed to watermark and not only address safety concerns but also achieves the force of the priorities: the watermark should remain even when the marked image gets damaged and distorted [15]. Ideally, a robust watermark system with no technical interference is kept intact under a defined distortion class. However, watermarks are erased and alternative encoding techniques may be used to restore them in various attacking scenarios. Robustness is a major problem in the blind picture watermarking system that needs extraction of the original pictures without any information [16]. Figure 3 shows block diagram of the traditional approach used in digital image watermarking.

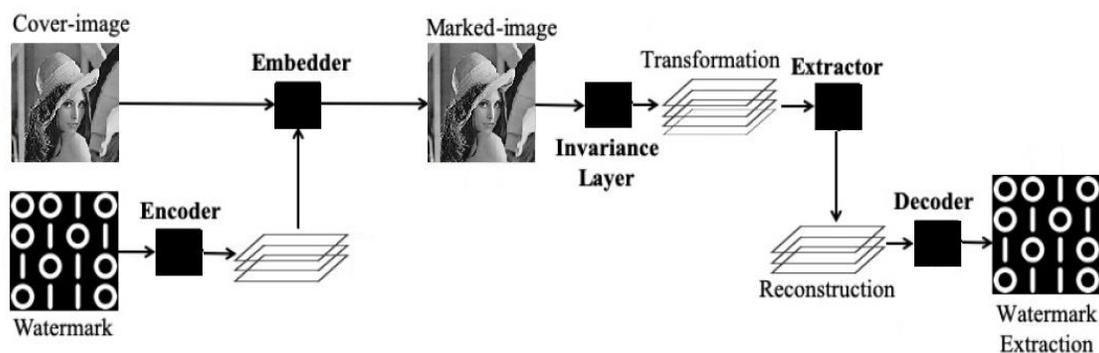


Figure 3 traditional approaches for digital watermarking

The inclusion of the watermark into the changed domain was employed in various digital watermarking systems. The inclusion in the low-frequency field might damage the quality of these images. In addition, the watermark is not incorporated into the frequency domain, which is set in case of JPEG compression due to the presence of quantization tables in frequency domain. Embedding in the middle frequency field might influence the image texture. It leads us to develop a novel method of integrating the watermark which would give the watermarked photos the least distortions in the medium frequency domain with acceptable locations [17].

2. ORGANIZATION OF THE RESEARCH

This research article contains the following section that consisting of; section 3 related work and motivation of the research, section 4 contains the proposed methodology for solving the existing algorithm problems. Section 5 delivers the obtained results with the proof table and graph. Section 6 concludes our proposed work and future task.

3. PRELIMINARIES

Kandi et al. used two convolution auto-encoders to generate a cover picture. A noted image demonstrates that the pixels of the first auto-encoder display bits with zero value as well as bits with one value in the pixels generated by the second auto-encoder [18]. Vukotic et al created a deep, one-bit watermarking method by employing imagers and unfavourable responses to include and find the first layer of a trained profound study model [19].

Li et al incorporated in the discrete cosine domain the watermark with standard techniques and utilized neonatal networks for extraction [20]. To erase the feature from the screen Fierro-Radilla et al have utilized neural networks to tie features to the watermark to create a major share [21].

Kim et al included a produced template and projected likely distortion parameters by using an additive technique and compared the extracted with the original with the use of convolution neural network approach. The existing systems for deep learning watermarking were

not entirely able to train and generalize integration and extraction methods using deep neural networks [22]. In the course of training to solve this difficulty, Mun et al supported proactive use of labelled photos of noise as opponents. However, the list of all kinds of attacks and their combinations cannot be logistically practical [23].

Recently, Uchida et al. have indicated that, the integration of watermarks into deep neural networks should include a framework. This is the first attempt at DNN digital watermarking for deep-neural network model protection. The technique suggested includes watermarks in deep neural network parameters utilizing the regularised parameter throughout the training period, resulting in limitations in the white box. All model parameters for extracting the watermark are necessary for the model owners. This significantly reduces its applicability as the stolen model parameters would not be revealed from the plagiarised service [24].

As far as we know, our proposed program is the first one that examined the ability of the deep neural networks to learn, generalize, and at the same time achieve resilience and blindness with a low computation time algorithm for higher accuracy [25].

Motivation of the research work

It is frequently desired for rapid watermarking with an easy-to-use function because of its low computation time. The proposed digital water marking approach should achieve an acceptable quality after the watermark is incorporated in perceptual distortion. The integrated watermark must be strong in terms of resilience for various types of attacks. For example, DWT provides a high level of imperceptibility and resilience for watermarked pictures, despite the fact that it has a high processing complexity owing to the wavelet changes. The challenges of achieving high imperceptibility, high strength and low computing time lead to a new watermarking technique. Therefore, this research work implements a light weight CNN model for digital watermarking scheme to reduce computation time in various geometric attacks without fault [26].

4. Proposed Method

Construction of LW-CNN is the solution used to provide less computation with quick response. This section has explained the implementation of LW-CNN in digital watermarking scheme and

it reaches higher robustness. The network structure of LW-CNN is concentrating on setting the minimum node set and set of edges for neural network.

Step 1:

Consider minimum node set of computing unit and set of edges connection between two nodes is defined as,

$$\text{LW-CNN} = (\mathbf{N}, \mathbf{S})$$

$$\text{Where, } \mathbf{N} = \{N_i \mid i = 1 \dots m\} \text{ and } \mathbf{S} = \{S_{ij} \mid 1 < i < j \leq m\}$$

Step 2:

Condition 1:

N node should be set of minimum computing unit participate in feature extracting.

Condition 2:

The edge set S should be the union mathematical function (M) of i and j.

Step 3:

Implementation

$$\mathbf{N} = \{N_{kb} \mid k = 1 \dots b; i = 1 \dots a\}, (M = a \times b)$$

$$\mathbf{S} = \{S_{ij} \mid i \in \{(1,1), (1,2) \dots (b, a)\}; j \in \{(i+1,1), (i+1,2) \dots (i+1, a)\} \cup \alpha * \text{drop}\{(i+2,1) \dots (b, a)\}\}$$

Step 4:

$$\text{Training set } \mathbf{D}_{train} = \{x_i y_i\}_{i=1}^S$$

$$\text{Secured key generation} = \mathbf{k} = \{y_s, y_d\} (\mathbf{D} \neq \mathbf{S})$$

Step 5:

Function watermark _ neural network

$$D_{LW} \leftarrow \theta$$

$$D_{tmp\ file} \leftarrow sample(S_{train}, y_s, \%)$$

For each $d \in D_{tmp\ file}$ do

$$x_{LW} = watermark\ original(s[x], embedded)$$

$$y_{LW} = y_d$$

$$D_{LW} = D_{LW} \cup \{x_{LW}, y_{LW}\}$$

End for

End function

Step 6:

Output $F_\theta = Train(D_{LW}, S_{train})$

This algorithm will be embedded with the preprocessing of input cover image through CNN network. Figure 4 shows the whole internal architecture of proposed framework.

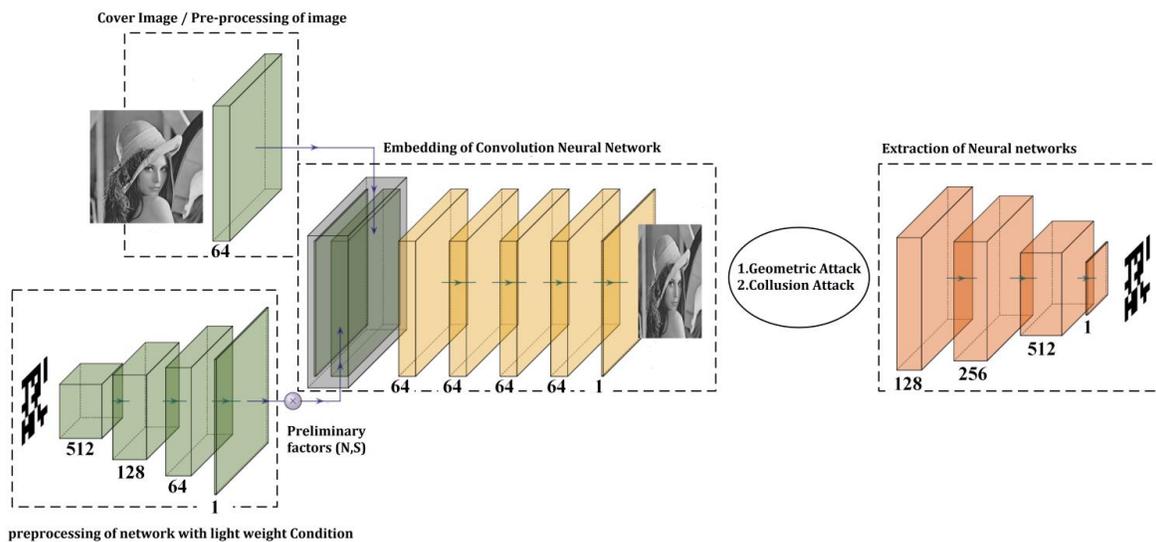


Figure 4 Proposed Architecture of Framework

5. RESULTS & DISCUSSION

In this research project, the suggested framework is tested by using a typical image dataset. Protection with various watermarks is successfully learnt by using the recommended architecture [27]. One of the objectives of the proposed study is to validate the ownership of proposed watermarking model in less time at a successful rate.



Figure 5 Results Obtained by Proposed Framework

Figure 5 shows the results obtained by proposed framework. After further examination, we confirm that our technique has a faster computation time and a higher success rate for proving ownership [28]. For various processes known as geometric and collusion type attacks, two separate attack types were explored. Segment wise analysis is considered to prove the robustness of the proposed LW-CNN framework. The robust segment is identified in the images part by part by including many pixels in various angle directions.

Table 1 Robust and Quality of Attacked Images by Proposed Framework

Attacks	Faulty bits	BER (SD)	SSIM (SD)	NCC (SD)
Noisy domain				
(i) Average	0%	0.066	0.89	0.796
(ii) Gaussian	0%	0.009	0.98	0.998
(iii) Median	0%	0.31	0.791	0.875

In the deep learning process, the costly computing can be avoided by reducing the model construction parameters of the CNN technique. Table 1 shows the robust and quality of attacked images by using the proposed LW-CNN framework.

To distinguish the unique early forecasting, the display of training watermarks is employed. The suggested method LW-CNN is a verification and comparison with many other classifiers named support vector machine, decision tree classifiers, which are pre-trained [29]. In comparison to other current algorithms throughout various attacks, the proposed model has showed a better resilience. Because of fewer network parameters, it is possible to train LW-CNN, which indicates that the high-speed neural network will provide less computing time. Table 2 shows the overall performance of proposed framework.

Table 2 Overall Performance of Proposed Framework

Method		Robustness	Computation time (Sec)
DT	Segment1	87%	62
	Segment2	79%	79
	Segment3	69%	98
SVM	Segment1	89%	50
	Segment2	80%	60
	Segment3	70%	69
Proposed LW-CNN	Segment1	100%	12
	Segment2	99%	19
	Segment3	99%	28

Figure 6 shows the overall performance of the proposed framework. From the graph, the LW-CNN provides stable robust and less computation time in all three segments, which contains various areas from the input image. Another conventional technique requires huge datasets or several data sets to improve the efficiency and system robustness.

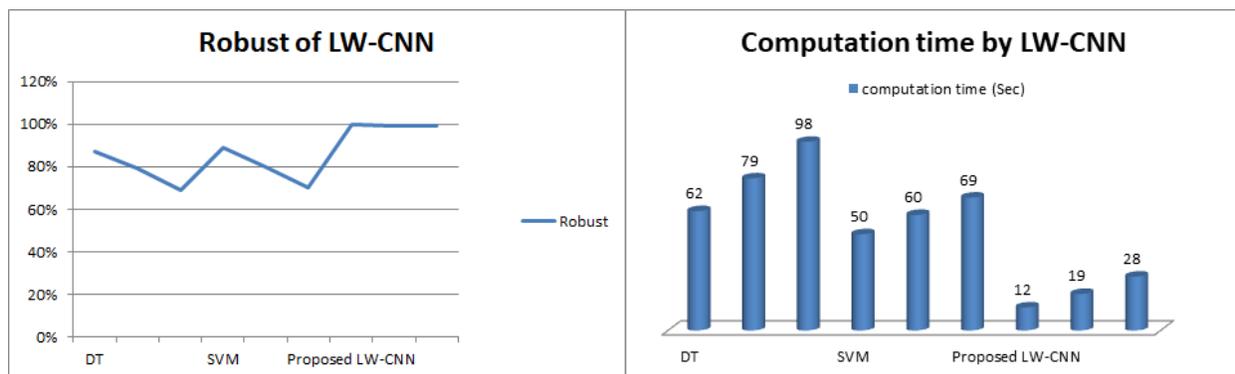


Figure 6 Overall Performance Metrics of Proposed Framework

However, the suggested approach is demonstrated to be cost-effective and it is most resilient than other traditional methods that are already included in Table 2 and further it is displayed in the graph. The proposed model has achieved a good computation time with the proposed algorithm. This approach makes us to adjust the finest tuning with the current state-of-the-art learning model to make the model more resilient.

6. CONCLUSION

Thus, an image watermarking system has been developed and tested by using lightweight and convolution neural networks (LW-CNN). The proposed digital watermarking framework takes advantage of the deep neural network's adaptability to generalize the watermarking picture algorithms and it uses an unattended watermarking architecture to accomplish its robust characteristics without the need for previous awareness on any distortions in the marked image.. This research work has demonstrated that the proposed scheme is feasible and capable of supporting combination, sophisticated, and demanding cameras for demonstrating the superiority of the proposed scheme and provide good results for every common attack. In our future studies, the deep neural networks present inside the system will be addressed and the architecture of

geometric, objective, and loss will improve, for instance, through various techniques such as ablation studies. Furthermore, as a future research project, we intend to focus on the maximum robustness of the algorithm in terms of more rotational and translational type attacks.

Thus, an image watermarking system is developed and tested by using lightweight and convolution neural networks (LW-CNN). The proposed blind image watermarking system takes advantage on the adaptability deep neural networks to generalize watermarking picture algorithms, which displays a watermarking architecture that remains unattended to achieve its robust properties and without requiring prior knowledge on the potential distortions present in the marked picture [30, 31]. We have demonstrated that, the proposed scheme is feasible and gains the ability to support combination, sophisticated, and demanding cameras, showing the supremacy of the offered scheme, and have shown the good outcomes for every common attack [32]. In our future study, the deep neural networks present inside the system will be addressed and the architecture of geometric, objective, and loss will improve, for instance through various techniques, such as ablation studies. Besides this research work will also concentrate on the greatest robustness of the algorithm in terms of more rotation and translational type attacks as future research work [33, 34].

REFERENCES

- [1] Dongyu Meng and Hao Chen. 2017. MagNet: a Two-Pronged Defense against Adversarial Examples.. In ACM Conference on Computer and Communications Security.
- [2] Manoharan, J. Samuel. "A Novel User Layer Cloud Security Model based on Chaotic Arnold Transformation using Fingerprint Biometric Traits." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 36-51.
- [3] Thakur, Shabnam, Rajesh Mehta, and Geeta Kasana. "Color Image Watermarking in DCT Domain Using SVR." In *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020*, pp. 301-312. Springer Singapore, 2021.
- [4] Erwan Le Merrer, Patrick Perez, and Gilles Trédan. 2017. Adversarial Frontier Stitching for Remote Neural Network Watermarking. In arXiv:1711.01894.

- [5] Shrestha, Sujana, and Subarna Shakya. "A Comparative Performance Analysis of Fog-Based Smart Surveillance System." *Journal of trends in Computer Science and Smart technology (TCSST)* 2 02 (2020): 78-88.
- [6] Kumar, Vikas, Prateek Muchhal, and V. Thanikasiselvan. "Information Security Through Encrypted Domain Data Hiding." In *International Conference on Innovative Data Communication Technologies and Application*, pp. 370-379. Springer, Cham, 2019.
- [7] Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, and Jan Kautz. 2017. Pruning Convolutional Neural Networks for Resource Efficient Inference. In *International Conference on Learning Representations (ICLR '17)*.
- [8] Raj, Jennifer S. "Improved Response Time and Energy Management for Mobile Cloud Computing Using Computational Offloading." *Journal of ISMAC* 2, no. 01 (2020): 38-49.
- [9] Vaneeta, M., V. Sangeetha, and S. Swapna Kumar. "Efficient Two-Layer Image Protection with Wavelet Transform Compression." In *Innovative Data Communication Technologies and Application*, pp. 433-448. Springer, Singapore, 2021.
- [10] Shelby Pereira and Thierry Pun. 2000. Robust template matching for affine resistant image watermarks. In *IEEE Transactions on Image Processing*.
- [11] Ananth, C., M. Karthikeyan, and N. Mohananthini. "Discrete Wavelet Transform Based Multiple Watermarking for Digital Images Using Back-Propagation Neural Network." In *International Conference on Inventive Computation Technologies*, pp. 441-449. Springer, Cham, 2019.
- [12] Sungheetha, Akey, and Rajesh Sharma. "3D Image Processing using Machine Learning based Input Processing for Man-Machine Interaction." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 1-6.
- [13] Nikiforos Pittaras, Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras. 2017. Comparison of Fine-Tuning and Extension Strategies for Deep Convolutional Neural Networks. In *International Conference on Multimedia Modeling*.
- [14] Ghosh, Atonu, Debashis De, and Koushik Majumder. "A Systematic Review of Log-Based Cloud Forensics." *Inventive Computation and Information Technologies* (2021): 333-347.
- [15] Ranganathan, G. "A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 66-74.

- [16] T. Yamada and M. Kamitani, "A method for detecting watermarks in print using smart phone: finding no mark," in *Proceedings of the 5th Workshop on Mobile Video*. ACM, 2013, pp. 49–54.
- [17] Adam, Edriss Eisa Babikir. "Survey on Medical Imaging of Electrical Impedance Tomography (EIT) by Variable Current Pattern Methods." *Journal of ISMAC* 3, no. 02 (2021): 82-95.
- [18] H. Kandi, D. Mishra, and S. R. S. Gorthi, "Exploring the learning capabilities of convolutional neural networks for robust image watermarking," *Computers & Security*, vol. 65, pp. 247–268, 2017.
- [19] V. Vukotic, V. Chappelier, and T. Furon, "Are deep neural networks good for blind image watermarking?" in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2018, pp. 1–7.
- [20] D. Li, L. Deng, B. B. Gupta, H. Wang, and C. Choi, "A novel cnn based security guaranteed image watermarking generation scenario for smart city applications," *Information Sciences*, vol. 479, pp. 432–447, 2019.
- [21] A. Fierro-Radilla, M. Nakano-Miyatake, M. Cedillo-Hernandez, L. Cleofas-Sanchez, and H. Perez-Meana, "A robust image zero watermarking using convolutional neural networks," in *2019 7th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2019, pp. 1–5.
- [22] W.-H. Kim, J.-U. Hou, S.-M. Mun, and H.-K. Lee, "Convolutional neural network architecture for recovering watermark synchronization," *arXiv preprint arXiv:1805.06199*, 2018.
- [23] S.-M. Mun, S.-H. Nam, H. Jang, D. Kim, and H.-K. Lee, "Finding robust domain from attacks: A learning framework for blind watermarking," *Neurocomputing*, vol. 337, pp. 191–202, 2019.
- [24] Yusuke Uchida, Yuki Nagai, Shigeyuki Sakazawa, and Shin'ichi Satoh. 2017. Embedding Watermarks into Deep Neural Networks. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval (ICMR '17)*.
- [25] L. A. Delgado-Guillen, J. J. Garcia-Hernandez, and C. Torres-Huitzil, "Digital watermarking of color images utilizing mobile platforms," in *2013 IEEE 56th*

- International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2013, pp. 1363–1366.
- [26] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A Systematic Approach." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 04 (2020): 226-235.
- [27] M. Zhao, Y. Wu, S. Pan, F. Zhou, B. An, and A. Kaup, "Automatic registration of images with inconsistent content through line-support region segmentation and geometrical outlier removal," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2731–2746, 2018.
- [28] Sivaganesan, D. "A Data Driven Trust Mechanism Based on Blockchain in IoT Sensor Networks for Detection and Mitigation of Attacks." *Journal of trends in Computer Science and Smart technology (TCSST)* 3, no. 01 (2021): 59-69.
- [29] Samuel Manoharan et al "Analysis of Data Embedding Techniques for Medical Images" published in *International Journal of Computer and Electrical Engineering* January 2012 , DOI: [10.7763/IJCEE.2012.V4.523](https://doi.org/10.7763/IJCEE.2012.V4.523)
- [30] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation–A Systematic Approach." *Journal of Artificial Intelligence* 3, no. 01 (2021): 1-16.
- [31] A. Fierro-Radilla, M. Nakano-Miyatake, M. Cedillo-Hernandez, L. Cleofas-Sanchez, and H. Perez-Meana, "A robust image zero water marking using convolutional neural networks," in *2019 7th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2019, pp. 1–5.
- [32] Hariharakrishnan, Jayaram, and N. Bhalaji. "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things." *Journal of ISMAC* 3, no. 02 (2021): 69-81.
- [33] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2016, pp. 372–387.
- [34] Y. Huang, B. Niu, H. Guan, and S. Zhang, "Enhancing image watermarking with adaptive embedding parameter and psnr guarantee," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2447–2460, 2019.

Author's Biography

R. Dhaya is currently a Professor, in the Department of Computer science and Engineering at King Khalid University, in the Kingdom of Saudi Arabia. His major area of research includes Image and Video Processing Algorithms, Computer Vision, Motion Analysis, Stereo Vision, Object Recognition, computer graphics, photo interpretation, image retrieval, Embedded Image Processing and Real-time image and video processing applications.