INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

CNN approach for medical image authentication

B Madhu[1], **Ganga Holi**[2]*

1 Assistant Professor, Department of Computer science and Engineering, Dr. Ambedkar Institute of Technology, Outer ring road, Bengaluru, India. Tel.: +919945187060
2 Professor and Head , Department of Information Science and Engineering, AMC Engineering College, Bengaluru, 560083

OPEN ACCESS

Received: 29.10.2020

Accepted: 13.01.2021

Published: 02.02.2021

Citation: Madhu B, Holi G (2021) CNN approach for medical image authentication. Indian Journal of Science and Technology 14(4): 351-360. https://doi.org/10.17485/IJST/v14i4.1963

* **Corresponding author**.

gangaholi@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To propose a Non-blind watermarking based on a Convolutional Neural Network (CNN). **Methodology:** An iterative learning model is proposed to ensure robustness and imperceptibility of watermarking process. In the first step, Stationary Wavelet Transformation (SWT) and Singular Value Decomposition (SVD) are used for the initial transformation and for embedding. The neural network is used to determine the relationship between host and watermarked image to extract the watermark. **Findings:** We have implemented our algorithm using Magnetic Resonant Imaging (MRI) and Computerized Tomography (CT), Mammogram and Retinal Images with different attacks and proved to have good robustness with Normalized Correlation coefficient (NC) value of 0.99 and invisibility feature with Peak Signal to Noise Ratio (PSNR) of 43.77 DB. We have compared our method with that of others and it proves to be good in terms of PSNR and NC values. **Novelty/Application:** This study provides a novel method to train CNN with both watermarked , attacked images and to classify them.

Keywords: Medical image authentication; stationary wavelet transform; convolutional neural networks; singular value decomposition

1 Introduction

Digitization of smart healthcare has led to the development and growth of communication and multimedia information systems. Nowadays, we find an essential requirement in telehealth applications, such as telediagnosis, teleconsulting, and telesurgery, which are being utilized for providing effective health care facilities. This comprises the exchange of important health data among physicians, hospital personnel, and patients, which need to be secured against any form of malicious and unintentional attacks. Digital watermarks have been majorly used for ownership protection to multimedia data. Machine learning is a part of artificial intelligence (AI) that enables the systems to automatically learn and improve from the experience. Machine learning emphasizes on the development of programs that can access the data and use it learn from them.

Artificial Intelligence (AI) is the best tool to assist doctors to diagnose, analyze with a prediction of the disease, so that faster actions can be taken. These techniques facilitate doctors and researchers to understand and analyze the diseases. Bilal et.al[(1)]

had suggested an image watermarking method of security using a fundus scan images with the help of a hybrid technique using Curvelet Transform and Singular Value Decomposition. The authenticity of medical images to verify patient information is a major challenge in an e-Health application. The technology and software services use the internet to manage, record, and transmit medical images in digital-driven applications that are known as e-Health, which stores electronic health records. Mehdi et al.[(2)] suggested quantization models for feature maps for JPEG images. Steganalysis Residual Network (SRNet) was suggested to design methods for forensics and Steganalysis using heuristic and constrained kernels. Assem[(3)] et al. suggested a time-efficient optimization process based on machine learning. Discrete Cosine Transformation (DCT) was used to embed the watermark to guarantee robustness against common watermarking attacks. K-Nearest Neighborhood regression method was used by the author for training and testing purpose. The results indicate that the method will take minimum time for the evaluation process.

Parmalik[(4)] suggested a watermarking method using Cascaded Neural Network (CNN) with feature optimization. CNN works on two different neural network models. CNN models generate the dynamic pattern for embedding process.M Saiful Islam[(5)] suggested a watermarking technique based on discrete wavelet transformation in three levels. The dynamic pattern is the basic unpredictable pattern of the watermark.Ramamurthy[(6)] et al. suggested a robust watermarking technique based on Propagation Neural Network (PNN). PNN is an example of a back propagation neural network, used to train the errors and to provide robustness for the method. Rikiya[(7)] suggested an overview of Convolutional Neural Networks and its applications in radiology. IP protection using Deep Neural Networks was suggested by Bita[(8)] to provide an efficient security framework for end to end system using convolutional neural network approach. Various image classification methods such as; Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest neighbor (KNN), Convolutional Neural Network (CNN), etc. are well-known concepts for classifying the images accurately.

This paper proposes a Neural Network model to perform an invisible, robust non-blind watermarking system for Medical images. It is a Convolutional Neural Network (CNN)-based method consists of pre-processing networks for watermarked image, an attacked images for training, and a watermark extraction process to extract the watermark. It has three peculiarities for the application aspect: The first is the imperceptibility of the Medical image. Imperceptibility refers to the condition that the embedded watermark should not produce distortion to the original image quality. That is, the watermarked version of the image must be indistinguishable from the original image. The second peculiarity is the robustness of the watermark information. Robustness is a crucial property of a watermark should be readable from images that go through different image processing operations, like addition of noise, lossy compression, filtering, histogram manipulation, and various geometrical transformations.

The novelty of our method is the use of CNN on watermarked to classify the images as attacked and non attacked image for efficient and securely transmit on over the public networks.

The organization of the paper is done as mentioned below: Section II of this paper contributes to the related work. The proposed method is explained in Section III. Section IV illustrates the details of the experiment and analysis. The proposed method is concluded in Section V.

2 Related Work

2.1 Stationary Wavelet Transformation (SWT)

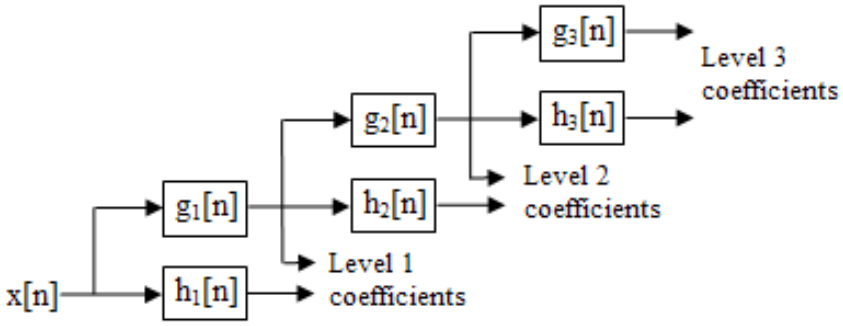

Fig 1. Decomposition in SWT

SWT is a transformation in wavelets considered overcoming the lack of translation-invariance of the discrete wavelet transform (DWT). This can be achieved by removing the down samplers and up samplers in the DWT and up-sampling the filter coefficients by a factor of $2^{(j-1)}$jth level in the algorithm. SWT provides an inherent redundant scheme since the output of each level in SWT have the same number of samples as the input – so the decomposition of N levels there is a redundancy of N in the wavelet coefficients. Figure 1 depicts the decomposition in SWT.

2.2 Singular Value Decomposition (SVD)

SVD [9] [10] is an effective mathematical tool used in the analysis of matrices. The SVD transformation gives three matrices with identical matrix size as the original. Consider n X n matrix of A, the output of SVD can be divided into three components, U, D, and V as in Eq. (1).

$$[\mathrm{UDV}] = \mathrm{SVD}(\mathrm{A}) \text{ and } \mathrm{A}^{-1} = \mathrm{UDV}^{\mathrm{T}} \tag{1}$$

Eq. (1) Represent components of U and V unitary matrices of n X n real, and the D is an n X n diagonal matrix. The diagonal values of 'D' are called the singular value of the given matrix A. Singular values are less affected, in image processing operations and image can be reconstructed with 70 percent of diagonal elements.

2.3 Convolutional Neural Networks (CNN)

CNN is a type of Deep Neural Network (DNN). CNN represents vast applications in image recognition and is used mainly to analyze visual imagery in image classification. CNN [11] can be applied for a lot of applications like image and video classification, recognition, medical image analysis, and recommendation systems. CNN requires three layers, namely input, target, and hidden layers. CNN has hidden layers has as Convolutional, ReLU, Pooling, and Fully Connected (FC) layers.

2.3.1 Convolution layer

Convolution (Conv) [12–14] is a special kind of linear function used for feature extraction with the help of a small number of arrays known as the kernel. This kernel is applied to the input which an array of numbers known as tensor. The feature map is calculated as a product of every element of the kernel and the every element in the tensor input at every location of the tensor and calculates summation. Similar procedure is applied to multiple kernels for an arbitrary number of feature maps. The feature maps with different parameters of the input tensors and different kernels will be taken as different feature extractors. Mainly two hyper key parameters define the convolution operation are the size and number of kernels. For an image recognition problem, based on the size of the image pixels larger filters (11x11 or 9x9) are used. In case of local and small features we can use small filters (3x3 or 5x5). Each filter captures different statistics of the image which ultimately helps in recognizing the words. The convolution layer was used with a filter size of 5X5 in our algorithm.

2.3.2 Nonlinear activation function

The rectified linear unit (ReLU) is a nonlinear activation function used in our algorithm. ReLU eliminates the negative values of the matrix. With simple computation it can be denoted in Eq. (2)

$$\mathrm{f}(\mathrm{x}) = \max(0, \mathrm{x}) \tag{2}$$

2.3.3 Pooling layer

Pooling offers usual down sampling operation with reduction of dimensionality for the feature maps. The reduction in the dimension will introduce translation invariance to small shifts and distortions. The pooling also reduced the number of successive learnable parameters. The hyper parameters of the pooling are filter size, number of strides, and padding. The output of this layer changes to the selection of filter size.

There are three types of pooling namely:

Max pooling: The maximum value of the pixel in the group will be selected.

Min pooling: The minimum value of the pixel in the group will be selected.

Average pooling: Average values of whole pixels for the group will be selected.

2.3.4 Fully connected layer

In the last convolution or pooling layer, the output feature maps are commonly smoothed. Smoothening converts the previous array of feature maps into a single-dimensional array of vectors. This stage may have one or more fully connected layers termed

as dense layers. The dense layer connects each input to every output with a learnable weight. The features are extracted by Convolution and downsampled by pooling are passed to the subset of fully connected layers to lead to the final outputs. A fully connected layer will be having the same number of output nodes as the number of categories and this will be trailed by a nonlinear function.

A convolution operation is applied to the input by the convolutional layer and the output is known as a feature map. The next layer after convolution is the Pooling layer. The function of the Pooling is combining the previous layer outputs in clusters of neurons into a single neuron in the next layer. Fully connected layers are used to connect each neuron from the previous layer of all the neurons in the next layer. The convolutional layer can receive neurons only from the previous layer. But fully connected layer will receive each neuron input from every element of the previous layer. CNN [15] majorly works on the concept of extracting features from the set of images. This property excludes the necessity of feature extraction manually. This feature of DNN makes the tasks in computer vision extremely more accurate. CNN's can learn about the features with the extension of tens or hundreds of hidden layers. Each layer increases the complexity of the gained features. The architecture of CNN takes advantage of combining the neural network training and the convolution operation for the classification of images in an accurate way. The CNN's can also be extended for the large data sets classification. The main advantage of CNN is the exclusion of feature engineering and reduces a lot of time for feature selection. CNN adopts less pre-processing steps, in contrast with other image classification methods. The prior knowledge and human effort are not required since the network filters learn on its own. CNN has used to establish with relationship between watermarked and attacked pixels. As a testing step watermarked image is altered with some attacks. The observed results demonstrate that the suggested technique has good performance with the identification of image.

This proposed method is for watermarking technique in medical image using CNN. In the initial stage, the input medical image is taken as the dataset and performed watermarking using SWT and SVD. The watermarked image is fed into to the CNN to classify the watermarked or attacked images.

3 Proposed method

The proposed method is used with the hybrid combination of SWT and SVD. SWT splits the host image as four equal-sized frequency bands LL, HL, LH, and HH. LL band indicates the low frequency of SWT, HH denotes high-frequency band, HL and LH denote mid-band frequencies respectively. The approximation details of the image will be present in the LL band, HL, and LH band gives horizontal and vertical details. HH band represents diagonal details of the image. The method selects HL and LH sub-band frequencies because the changes in these images will have less impact on imperceptibility. Watermark inserted in these bands sustains for image processing operations like changes in intensity, the addition of noise, and human visual system limitations. The visual system cannot differentiate modifications made in the mid-bands. The suggested system is constructed with the concept of substituting singular values of the host image with singular values of the watermark.

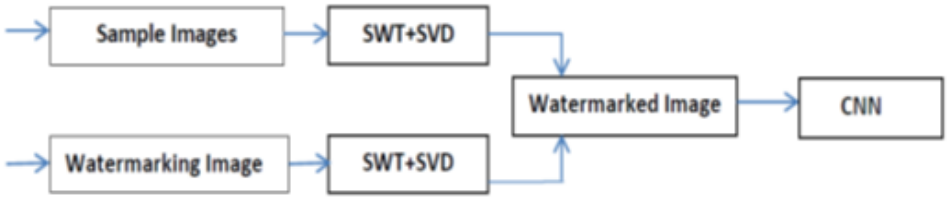

Fig 2. CNN training

Fig 3. CNN testing

The process of CNN Training is shown in Figure 2 . The watermarking process generates a watermarked image. The watermarked image is then fed as input to CNN by resizing it into 28X28X1 gray scale images. These image patches are given as input to the CNN for the classification. The watermarked images with Salt and Pepper, Gaussian and histogram attacks are also given as an input to the CNN.CNN classifies the image as watermarked and attacked image.

The process of CNN Testing is shown in Figure 3 . Watermarked image at the other end is sent to the CNN test procedure to verify the originality of the image. In either watermarked or attacked cases the extraction of a watermark is performed.

The parameters used for CNN construction are mentioned below.

1. The input layer of size (28X28X1).
2. Convolution layer 9X9 with 10 filters
3. Mean pooling with 2x2
4. Convolution layer 3X3 with 20 filters
5. Mean pooling with 2x2
6. Fully connected layer with 2 outputs

4 Experimental Results

The experiment was conducted with a dataset of 884 different medical images. The medical images and watermark logo are of 512 × 512 in size as in Figure 4. CT and MRI images were collected from Padmashree diagnostics and Retinal[16], and Mammogram[17] images were used for the experiment purpose. The training process is conducted on 884 images with different attacks and testing is performed with 200 images. The model is trained to recognize two classes; one is a watermarked and the other image is attacked image.

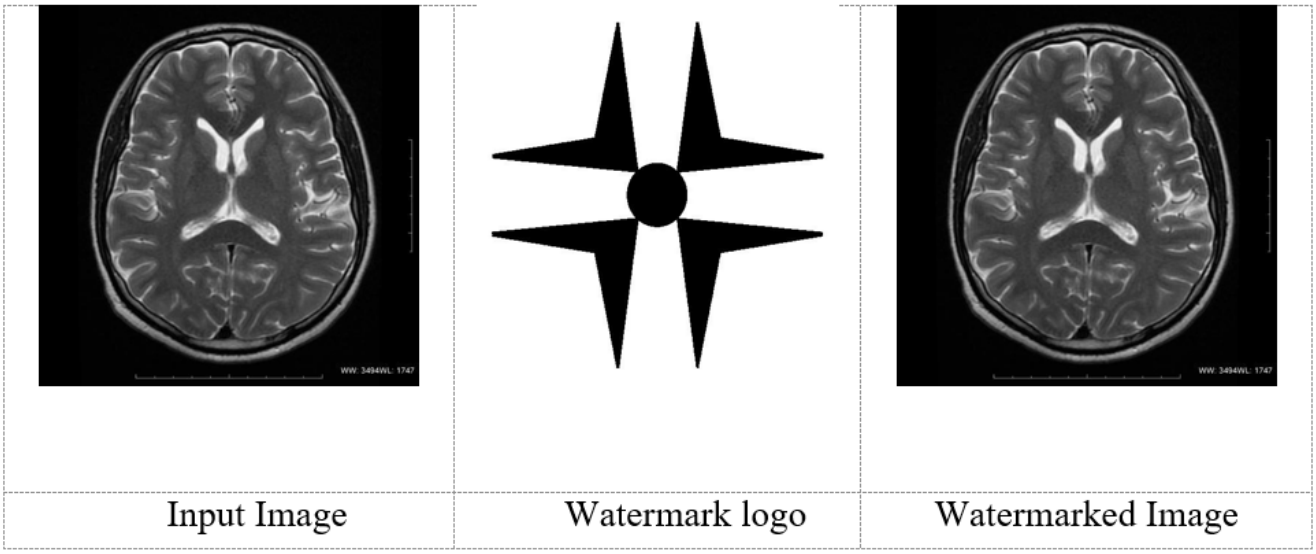

Fig 4. Input watermark and watermarked images without attacks

4.1 Performance parameters

The quality of an image can distort due to variations during transmission of the image and these distortions are called errors. The robustness of the said method are tested with the simulation of different attacks viz. Gaussian, Salt and pepper noise, sharpening attacks.

4.1.1 Peak Signal to Noise Ratio (PSNR)

PSNR is the ratio of the maximum power of a signal to the corrupting noise power. The PSNR can be conveyed through logarithmic decibels. PSNR was calculated between the original and watermarked images. The attacks are applied to the

watermarked images to check the robustness and imperceptibility. The testing images include a combination of attack and watermarked images. Figure 5 represents PSNR values for testing images. PSNR is calculated for all the test images. Figure 6 indicates the average PSNR value for attack free watermarked images as 43.77 DB. Similarly, the average PSNR attacked images as 34 DB is shown in Figure 7.

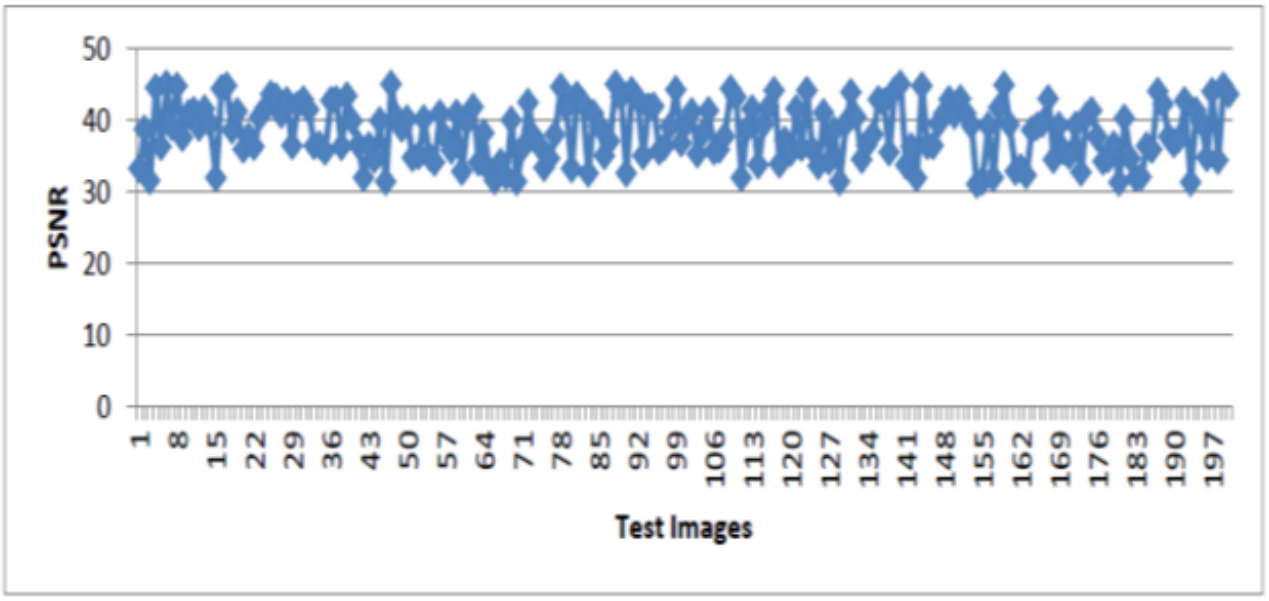

Fig 5. PSNR values for testing images

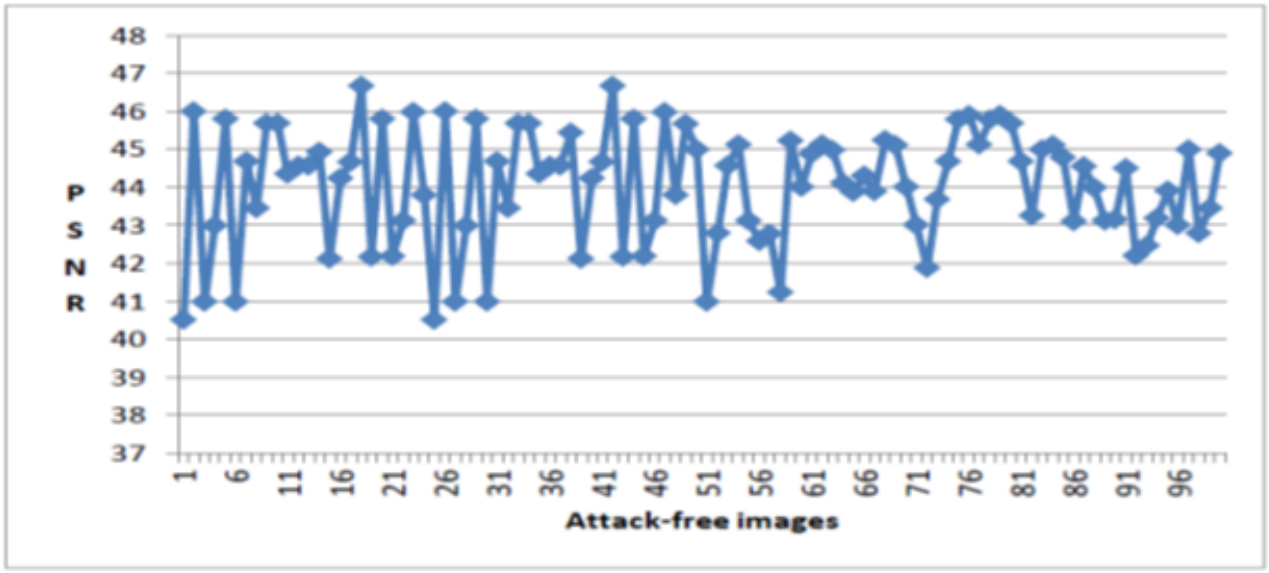

Fig 6. PSNR values for attack free watermarked images

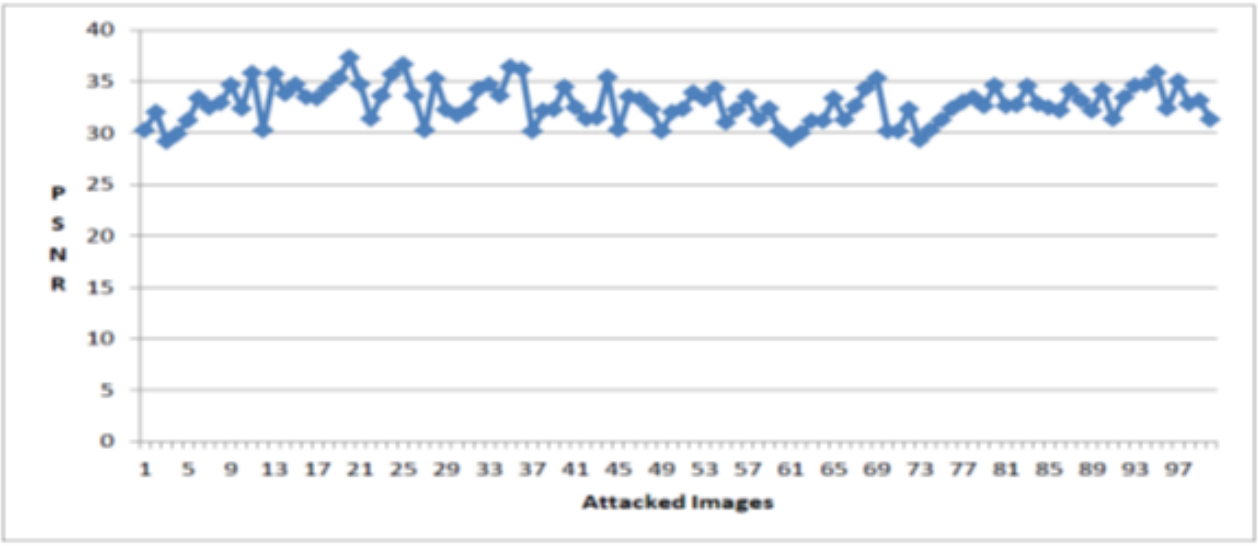

Fig 7. PSNR values for Attacked watermarked images

In[10] used spatial domain LSB based substitution method for image watermarking. In[18] has tested DWT and SVD watermarking with different scaling factors. In[14] has given DWT and SVD technique with a CNN model for embedding and extraction process. All these methods used Sipi database for the implementation. Figure 8 indicates the comparison results of above methods with our method.

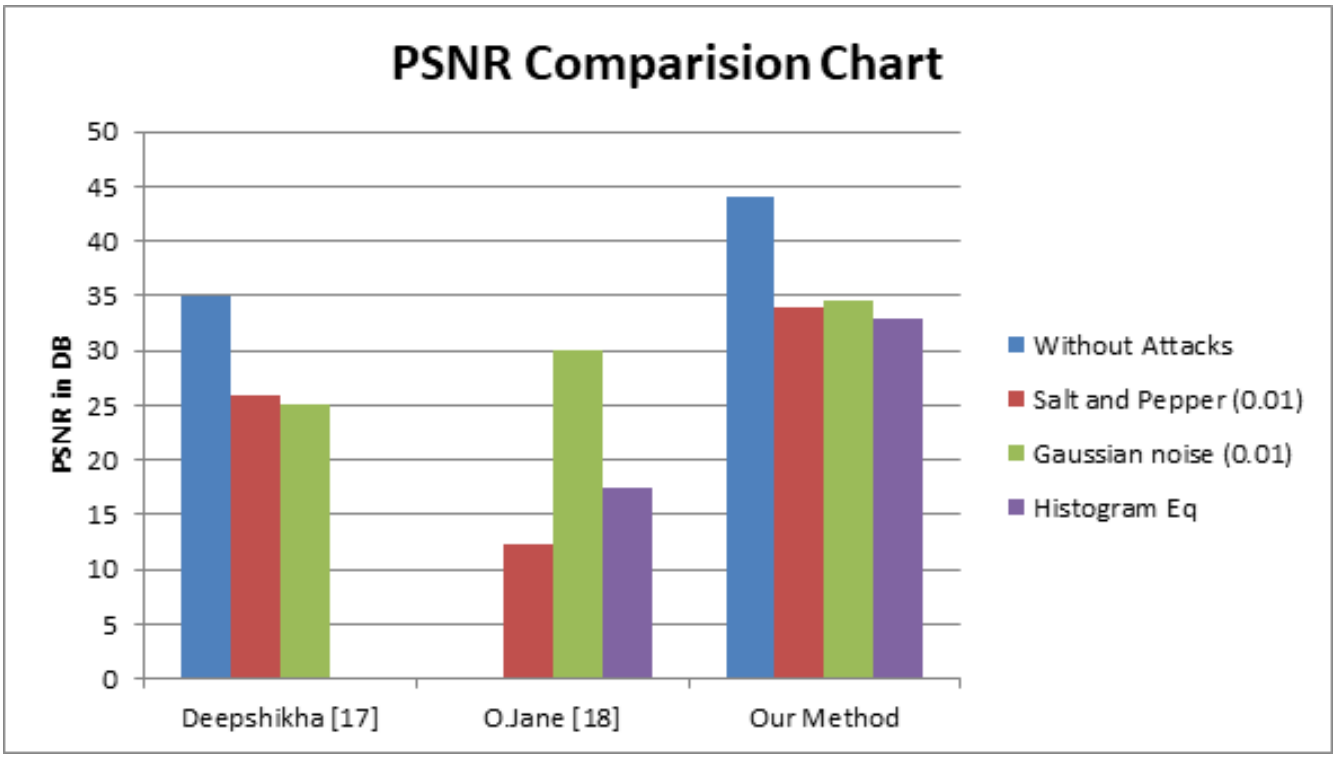

Fig 8. PSNR value comparision with state of art

4.1.2 Normalized correlation coefficients

The Normalized Correlation coefficients are used to measure the relationship between two images. The correlation between two signals or simply cross-correlation is a standard tool for evaluating the degree of similarity between two signals. NC is taken between original and extracted watermarks. Figure 9 represents NC values for testing images. The value of NC is greater than 0.9 indicating good robustness parameter.

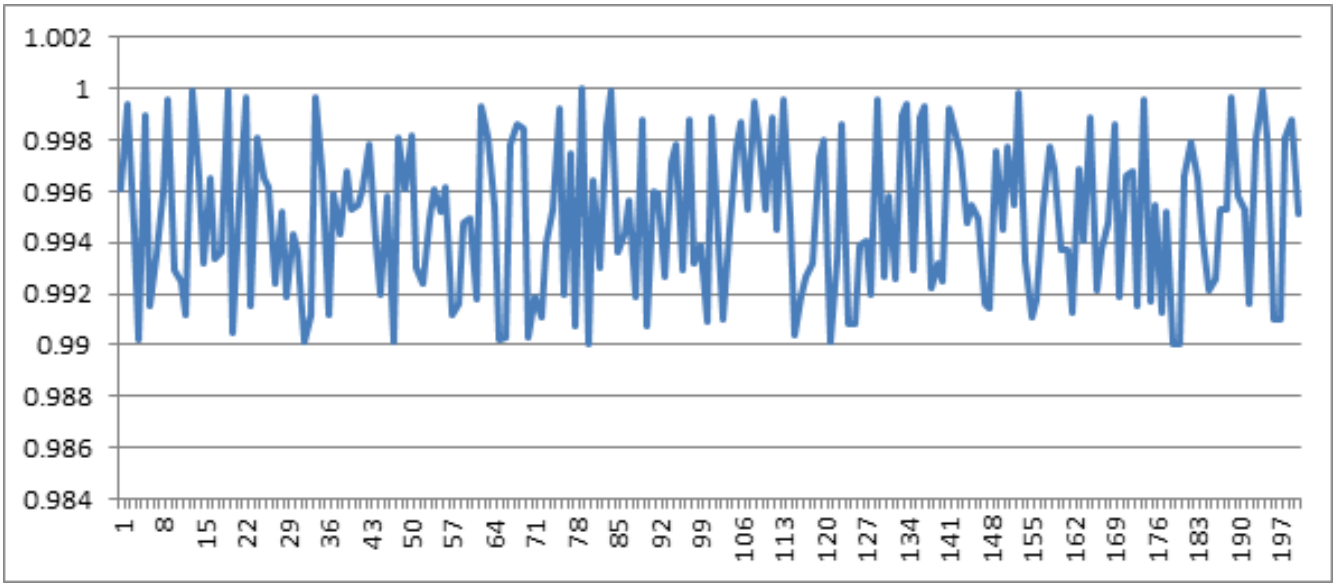

Fig 9. NC values for testing images

In [14] used Electronic Patient Record[EPR] as a watermark and the watermark was encoded first using turbo code. This watermark is embedded the redundant discrete wavelet transform (RDWT) and Singular value decomposition (RSVD) coefficient of the cover image. Figure 10 represents the comparison of robustness with Singh [14] method in Table 1. The comparison clearly indicated that our method performs better.

Madhu [19] suggested a medical image authentication model using SWT and SVD. In this method, watermark was encrypted with chaotic encryption and then embedded with wavelet coefficients as in Table 2. The bold value indicates that for Gaussian noise our previous method provides good PSNR compared to the current work.

Table 1. Comparision of Singh [14] and our method

Attack	Singh [14] et.al		Proposed work	
	NC	SSIM	NC	SSIM
S&P Noise (0.01)	0.8451	0.708825	0.9820	0.9807
Gaussian noise (0.05)	0.4987	0.045102	1.0000	0.9820
JPEG Compression(QF = 90)	0.9995	0.898515	1.0000	0.9978
Histogram equalization	0.5007	0.260528	1.0000	0.9818
Median filter (2X2)	0.9759	0.820554	1.0000	0.9830

Table 2. Comparision of previous method with proposed work

Attack	Madhu [19] et.al		Proposed work	
	PSNR	NC	PSNR	NC
S&P(0.05)	30.48	0.9353	33.9832	0.9826
GN(0.05)	**31.16**	0.9328	30.5579	1.0000
Median filter (3X3)	30.71	0.9176	45.4994	1.0000

4.1.3 Accuracy

Accuracy[20] is the measurement between a standard that is expected to be correct with an unknown quality image classification. The classification accuracy for watermarked and attacked images is 95.34 %.

The Imperceptibility of the proposed method is calculated using PSNR, for the proposed method, the PSNR is 43.77 DB. NC of an original and retrieved watermark is near to 1. It indicates that our method is more robust to various attacks .

5 Conclusion

This study discusses the hybrid method for watermarking using SWT, SVD and CNN. SWT and SVD are used to extract the dominant features of the both original image and watermark logo. The reconstruction of the watermarked image is done by adding both dominant features of both the images. The watermarked images are trained and tested using CNN with an accuracy of 95.34% of the medical images. The Experimentation has been conducted against different attacks and proved to be better in terms of the PSNR values with 43.77 DB and NC with 0.99. In the future, we can extend our implementation work for color images, audio and video watermarking.

Acknowledgment

The author acknowledges her sincere gratitude to Dr. Gopinath R G M.D., DNB, and FRCR, Consultant Radiologist at Padmashree Advanced Imaging Services, Bangalore, India, for providing CT and MRI Medical images and a logo for the implementation purpose.

References

1) Hassan B, Ahmed R, Li B, Hassan O. An Imperceptible Medical Image Watermarking Framework for Automated Diagnosis of Retinal Pathologies in an eHealth Arrangement. *IEEE Access*. 2019;7:69758–69775. Available from: https://dx.doi.org/10.1109/access.2019.2919381.
2) Boroumand M, Chen M, Fridrich J. Deep Residual Network for Steganalysis of Digital Images. *IEEE Transactions on Information Forensics and Security*. 2019;14(5):1181–1193. Available from: https://dx.doi.org/10.1109/tifs.2018.2871749.
3) Assem M. A "A time-efficient optimization for robust image watermarking using Machine Learning. *International Journal Expert Systems with Applications*. 2018;p. 197–210. Available from: https://doi.org/10.1016/j.eswa.2018.02.002.
4) Kumar P, Sharma AK. A Robust Image Watermarking Technique using feature Optimization & Cascaded Neural Network. *International Journal of Computer Science and Information Security*. 2019;17(8):36–45. Available from: https://sites.google.com/site/ijcsis/.
5) Islam MS, Ullah MA, Dhar JP. An imperceptible & robust digital image watermarking scheme based on DWT, entropy and neural network. *Karbala International Journal of Modern Science*. 2019;5(1):36–44. Available from: https://dx.doi.org/10.33640/2405-609x.1068.
6) Ramamurthy N. The Robust Digital Image Watermarking scheme with Back Propagation Neural Network in DWT Domain. In: and others, editor. International Conference on Modeling, Optimization and Computing Procedia Engineering (ICMOC-2012) proceedings in Science Direct. ;p. 3769–3778. Available from: https://doi.org/10.1016/j.proeng.2012.06.432.
7) Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*. 2018;9(4):611–629. Available from: https://dx.doi.org/10.1007/s13244-018-0639-9.
8) Darvish HB, Chen F. Deep-Signs: An End-to-End Watermarking Framework for Ownership Protection of Deep Neural Networks. In: and others, editor. ASPLOS Proceedings ASPLOS '19. ;p. 485–497. Available from: https://doi.org/10.1145/3297858.3304051.
9) Zhen W, Mo S, Jin X, Qu Y, Dengx F, Shuaix J, et al. Robust and High Capacity Watermarking for Image-Based on DWT-SVD and CNN. In: and others, editor. IEEE Conference Industrial Electronics and Applications (ICIEA). 2018;p. 1233–1237. Available from: https://doi.org/10.1109/ICIEA.2018.8397898.
10) Chopra D, Gupta P, Sanjay BCG, Gupta A. Lsb Based Digital Image Watermarking For Gray Scale Image. *IOSR Journal of Computer Engineering*. 2012;6(1):36–41. Available from: https://dx.doi.org/10.9790/0661-0613641.
11) Zubov IG, Lysenko NV, Labkov GM, Labkov. Detection of the Information hidden in the image by Convolutional Neural Networks. In: and others, editor. IEEE Conference. ;p. 393–394. Available from: https//doi.org/10.1109/EIConRus.22019.8656886.
12) Convolutional Neural Network. 2020. Available from: https://en.wikipedia.org/wiki/Convolutional_neural_networkaccessedon11th.
13) AtoanyFierro-Radilla, Nakano-Miyatake M, Cedillo-Hernandez M, Cleofas-Sanchez L, Perez-Meana H. A Robust Image Zero-watermarking using Convolutional Neural Networks. *International Workshop on Biometrics and Forensics (IWBF)*;p. 1–6. Available from: https://doi.org/10.1109/IWBF.2019.8739245.
14) Anand A, Singh AK, Zhihan V, Bhatnagar G. Compression-then-Encryption based Secure Watermarking Technique for Smart Healthcare System. IEEE Computer Society. . Available from: https://doi.org/10.1109/mmul.2020.2993269.
15) Holi G, Jain DK. Convolutional Neural Network Approach for Extraction and Recognition of Digits from Bank Cheque Images", Springer Nature Singapore. In: and others, editor. Lecture Notes in Electrical Engineering;vol. 545. ;p. 331–341. Available from: https://doi.org/10.1007/978-981-13-5802-9_31.
16) . 2020. Available from: http://www.isi.uu.nl/Research/Databases/DRIVE/.
17) Mammoimage database. 2020. Available from: https://www.mammoimage.org/databases/.
18) Jane O, Elbaşi E, İlk HG. Hybrid Non-Blind Watermarking Based on DWT and SVD. vol. 12. Universidad Nacional Autonoma de Mexico. 2014;p. 750–761. Available from: https://dx.doi.org/10.1016/s1665-6423(14)70091-4.
19) Vathsala MK, Holi G. RNN based machine translation and transliteration for Twitter data. *International Journal of Speech Technology*. 2020;23(3):499–504. Available from: https://dx.doi.org/10.1007/s10772-020-09724-9.

20) Perez H, Tah JHM. Improving the Accuracy of Convolutional Neural Networks by Identifying and Removing Outlier Images in Datasets Using t-SNE. *Mathematics*. 2020;8. Available from: https://dx.doi.org/10.3390/math8050662.