Deep Learning & Hybrid Model – The Future of Medical Image Watermarking?

Yew Lee Wong[1*], Jia Cheng Loh [1], Chen Zhen Li [1], Chi Wee Tan [1]

[1] Faculty of Computing and Information Technology, Tunku Abdul Rahman University College, Kampus Utama, Jalan Genting Kelang, 53300 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia
*Corresponding author: wongyewlee-wm19@student.tarc.edu.my

ABSTRACT

The frequent usage of medical records in electronic form has made Medical Image Watermarking (MIW) relatively more significant than it used to be. MIW is very significant to preserve the completeness and integrity of the medical images. For the time being, with the trade-offs between visibility and robustness, there are no perfect algorithms for invisible watermarking. In many novels, Deep-Learning-Based Approach has been proposed to solve the trade-offs. In this study, multiple implementations of invisible watermarking techniques such as Deep-Learning-Based Approach and Non-Deep-Learning-Based-Approach are being compared. This comparative study measures the limitations and robustness on a dataset of breast ultrasound images. Eighteen extreme attacking methods were carried out on the encoded images, performance was then evaluated using peak signal-to-noise ratio (PSNR) and normalized cross correlation (NCC). Encoded images were then tested against a digital transmission channel to test its robustness. To conclude, The Deep-Learning-Based-Approach of RivaGAN showed the best robustness against multiple extreme attacks. The Non-Deep-Learning-Based-Approach of discrete wavelet transform – discrete cosine transform – singular value decomposition (DWT-DCT-SVD) has the best imperceptibility. Therefore, we confirm the feasibility of Deep-Learning-Based-Approach in Medical Image Watermarking, however more work is needed to be done to achieve perfect Deep-Learning-Based-Approach in terms of imperceptibility.

Keywords: Invisible Watermarking, DCT, DWT, SVD, RivaGAN, Deep-Learning-Based Invisible Watermarking

1.0 INTRODUCTION

Medical image watermarking has been significant in this digital era. Watermarking on medical images which is equivalent to digital signatures, is required not to compromise the quality of image. Digital image watermarking can be understood as a process of embedding and extracting signatures such as names into images that are to be distributed through digital transmission. Conventional visible image watermarking is not applicable in the use case of medical images as integrity and completeness of the photo is the utmost priority.

Invisible watermarking can be traced back to the work by Yeung et al which proposed a method of image verification (Yeung & Mintzer, 1997). Watermarking can be done on either spatial domain or the transform domain.

In the scenario of medical image watermarking, digital medical images transmitted over any channel may raise data integrity problems, therefore, invisible watermarking could be the

solution. However, there is no perfect algorithms or solutions for invisible watermarking as trade-offs can happen between visibility and robustness when doing watermarking (Mousavi et al., 2014). For a watermarking technique to reach the optimum state in the use case of medical image watermarking, the techniques shall take into account robustness, imperceptibility and security. Robustness can be simply understood as the resilience of the watermarking towards any attacks while imperceptibility focuses on the quality of watermarked image after the embedment process.

With the need for a perfect algorithm that can satisfy the need of robustness and imperceptibility, research has been incorporating deep-learning-based techniques into the field of medical image watermarking. Embedded watermarks can be extracted using convolutional neural networks. However, the robustness challenge has always been hard to satisfy due to the fragility of the deep neural networks (Papemot et. al., 2016).

In our study, we explored the implementations of deep-learning-based invisible watermarking techniques with hybrid-based techniques. Hybrid invisible watermarking techniques has proven good performances in the past. Through our study, we hope:

- To identify and verify the robustness of deep-learning-based invisible watermarking algorithm
- To measure the watermarking effects on the medical images using the metrics of PSNR and NCC of deep-learning-based techniques and non-deep-learning-based-techniques.
- To investigate the limitations and resistance of the algorithms towards extreme attacks.
- To verify the completeness of embedded messages after digital transmission.

2.0 LITERATURE REVIEW

2.1 Digital Image Watermarking & Invisible Watermarking

Due to the rapid expansion of the internet, the distribution of digital photographs has become increasingly popular; as a result, data protection has become increasingly crucial. -(Abdulrahman, 2019). In the context of digital watermarking, the process of embedding or hiding data in another digital data, and then extracting the hidden information can be defined as invisible watermarking. (Tao et al., 2014) According to others, it has grown easier to tamper with medical photographs since modern picture editing software has become more widely available in the past few years (Coatrieux, 2006). In order to address these problems, invisible watermarking can be used for data concealing as well as to safeguard the integrity of data (Coatrieux, 2006). It is possible to divide the digital watermarking domain into two subdomains: the spatial domain and the frequency domain, respectively (EL-Shazly, 2004). Robustness and imperceptibility are two performance criteria that are commonly used to evaluate picture watermarking techniques; nevertheless, these two characteristics are diametrically opposed to one another (Usman et al., 2008). When measuring the imperceptibility of the watermark, peak signal to noise ratios (PSNR) are utilised. The image quality should not be distorted when there is a watermark present, as measured by the peak signal to noise ratios (PSNR) (Al-Haj, 2007). PSNR is commonly expressed in decibels (dB), and it is widely used in medical image watermarking (MIW) algorithms to compare their performance (Faragallah et al., 2021). A technique's robustness is measured by the watermark's resilience and immunity to removal attempts as well as degradation attempts (Voloshynovskiy et al., 2001).

2.2 Medical Image Watermarking

Telemedicine has grown in popularity over the last few decades as communication technology has advanced. Diagnostic procedures rely heavily on medical images. They can now be transmitted easily across the globe via communication channels (Pandey and Singh, 2016). However, transmission over public networks puts security, confidentiality, copyright, and integrity at risk. Medical data theft or tampering can result in incorrect diagnoses. Thus, during the transmission of medical images, security, confidentiality, and integrity are paramount concerns. In this situation, medical image watermarking (MIW) has emerged as a viable option (Hussain and Wageeh, 2013). Significant information is concealed within a cover medical image during the watermarking process, and that information should not be detected, retrieved, or modified by an unauthorized user. It is frequently used in one-to-many communication systems, whereas steganography is typically used in one-to-one communication systems (Sharma and Gupta, 2012). Watermarking medical images are classified as a reversible technique or an ROI (Region Of Interest) technique (Sonika and Inamdar, 2012). A robust and reversible watermark is required for diagnostic purposes in a health information system. The reversible watermarking technique maintains the integrity of the original medical image during recovery. If the extracted medical image is corrupted in any way, the result will be incorrect (Rohini and Bairagi, 2010).

2.3 Deep-Learning-Based Image Watermarking

Deep-Learning-Based Image Watermarking has been proved superior against other algorithsm in term of concealment and robustness. (Zhang et. al, 2021). Convolutional neural networks (CNN), autoencoders (AE), and generative adversarial networks (GAN), all of which are common in deep learning, have been the mainstays of research. This new deep learning architecture, known as RivaGAN, goes beyond the usual convolutional layers and algorithms. The encoder's robustness was tested and improved using two independent adversarial networks. A 32-bit watermark is embedded into a sequence of frames using this design. Any common video processing operations like cropping, scaling, and compression were shown to be robust to RivaGAN. (Zhang et al., 2019).

2.4 Hybrid Watermarking

The efficiency of the watermarking technique can be increased by the combination of different transformations (Assini et al., 2018). The hybrid watermarking of DWT-DCT-SVD was proven to be very robust because it does not embed all singular values and can be applied to create algorithms for loss image compression (Navas et al., 2008). The performance of the DWT-DCT hybrid watermarking was shown to be superior to the performance of the DWT method alone (Al-Haj, 2007). In comparison to DWT, the DWT-DCT significantly improved robustness especially to the linear and non-linear attacks (Abdulrahman & Ozturk, 2019). Additionally, it has been proved that the non-hybrid watermarking approach of DWT is resistant to any typical image processing processes (Lala, 2017).

3.0 RESEARCH METHODOLOGY

3.1 Dataset and Algorithms

The dataset used is a collection of breast ultrasound images among women between the ages of 25 and 75 years old which is available at Kaggle (Al-Dhabyani W et al., 2020). A total of 20 images were selected randomly from these 780 images with an average image size of 500×500 pixels. The chosen images were named alphabetically from "MRI_A" to "MRI_T". Four algorithms of invisible watermarking were chosen, namely DWT, DWT-DCT, DWT-DCT-SVD & RivaGAN.

3.2 General Framework

As illustrated in Figure 1, it shows the overall flow of our study. Firstly, we will encode a watermark in string format into the original MRI images. Then we will attack those encoded images using 18 different methods. Transmission of encoded images were also done on the attacking phase. After that we will try to decode the watermark from the attacked images and calculate the PSNR and NCC value. Lastly, the result will be visualize using some chart.

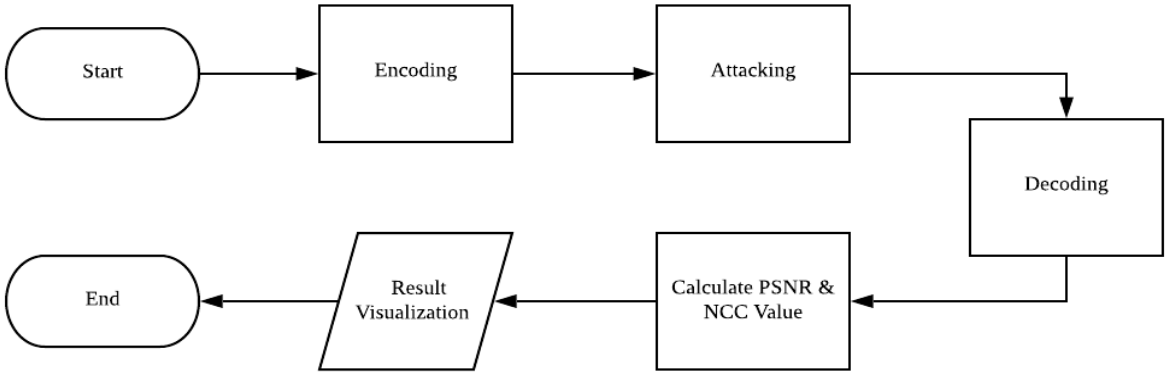

Figure 1. Medical Image Watermarking Framework.

3.3 Testing Criteria

On the pass rate, messages retrieved after being attacked is strictly being compared absolutely. Only if the output matches 100% with the initial input can be considered as passing the test. Decoding errors were counted as failure through exceptions caught by decoder. Partial success that the output matches the input was considered as failure. On the test of transmission, watermarked photos were transmitted through WhatsApp Image, WhatsApp Document, Google Drive, Facebook Messenger and Gmail. The images received on the receiving end were put into decoder to retrieve the embedded messages. Output that matches the initial input 100% will only be considered as pass the test. On testing the implementation of the selected library, average of decoding and encoding time were done using time library in Python. Elapsed time was recorded down over 1000 iterations of the operation and mean were calculated. On measuring the relationship between characters length of embedded message and file size, different randomly generated string of different length was encoded. File size was compared on before and after encoding. On measuring the performance of each watermarking algorithm, we evaluate the image of before using the value of Peak Signal-to-Noise Ratio (PSNR) and Normalized Cross Correlation (NCC).

3.4 Experiment Environment

The testing of the implementation was done on a desktop system of such specifications in Table 1.

Table 1. Testing system.

CPU	Intel Xeon E5-2650v2 @ 2.60Ghz, 8 Cores 16 Threads
RAM	16GB DDR3 1666Mhz
Operating System	Windows 10 Pro 64-bit (10.0, Build 19043
Python Version	3.8.10

4.0 RESULTS AND DISCUSSIONS

As shown in Table 2, there are 18 types of attacking methods that will be used to test the robustness of each watermarking algorithm.

Table 2. Attacking methods.

	Kernel Settings/ Ratio / Strength
Averaging	size = 5x5
Bilateral Filtering	D = 9, sigmaColor = 75, sigmaSpace = 75
Brightness Decrease	40 %
Brightness Increase	40 %
Crop Horizontal	50 %
Crop Vertical	50 %
Gaussian Blurring	Size = 5x5
Gaussian Noise	mean=0, variance=0.01
JPG	Convert to JPG
Masks	n = 5, ratio = 0.3
Median Blurring	Size = 7
Poisson Noise	Lambda = 20
Rotate	10 degrees
Salt & Pepper	10 %
Scale Down	25 %
Scale Up	25 %
Sharpen Filtering	[-1, -1, -1], [-1, 9, -1], [-1, -1, -1]
Speckle Noise	mean=0, variance=0.01

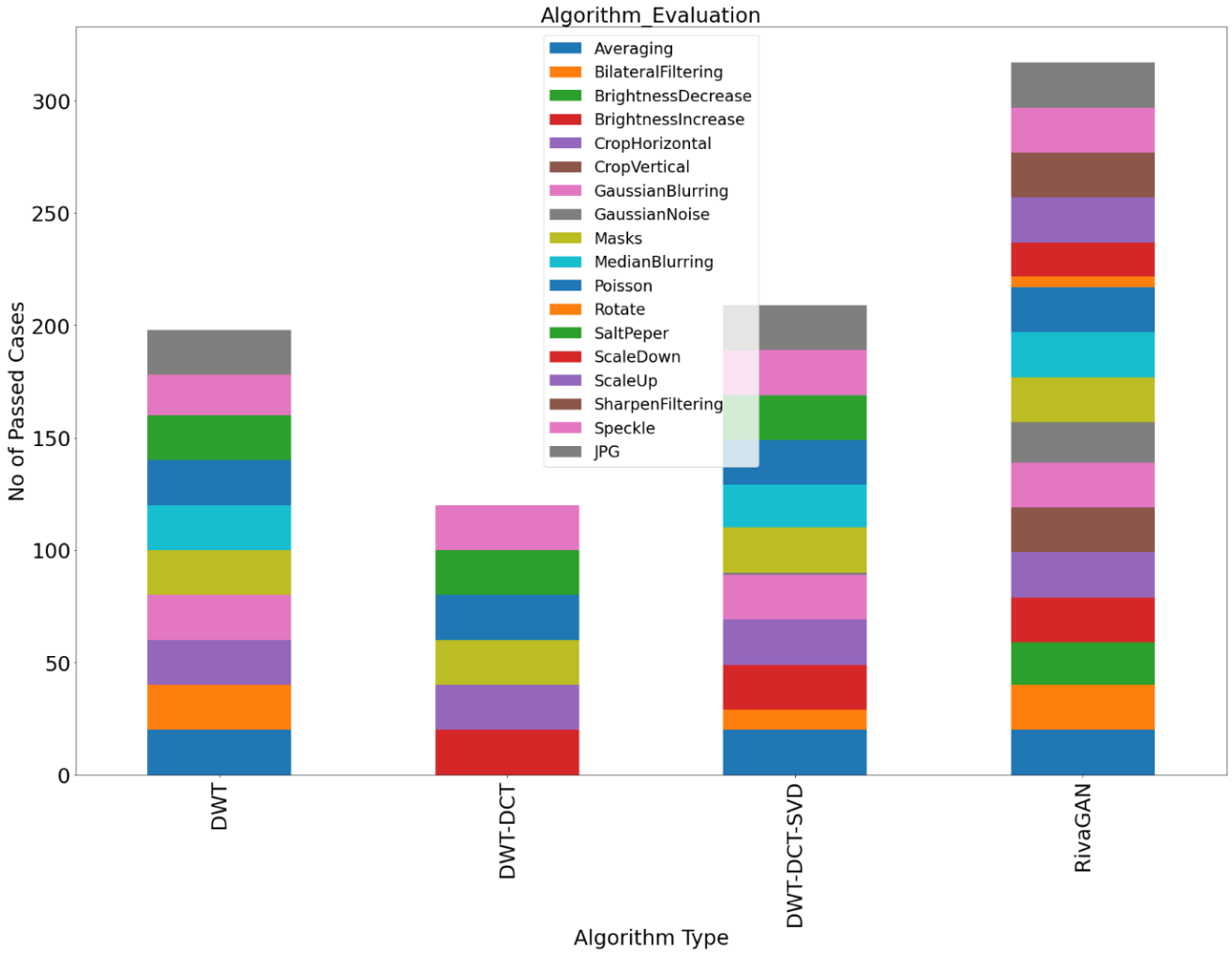

Figure 2. Algorithm Evaluation Based on Pass Rate.

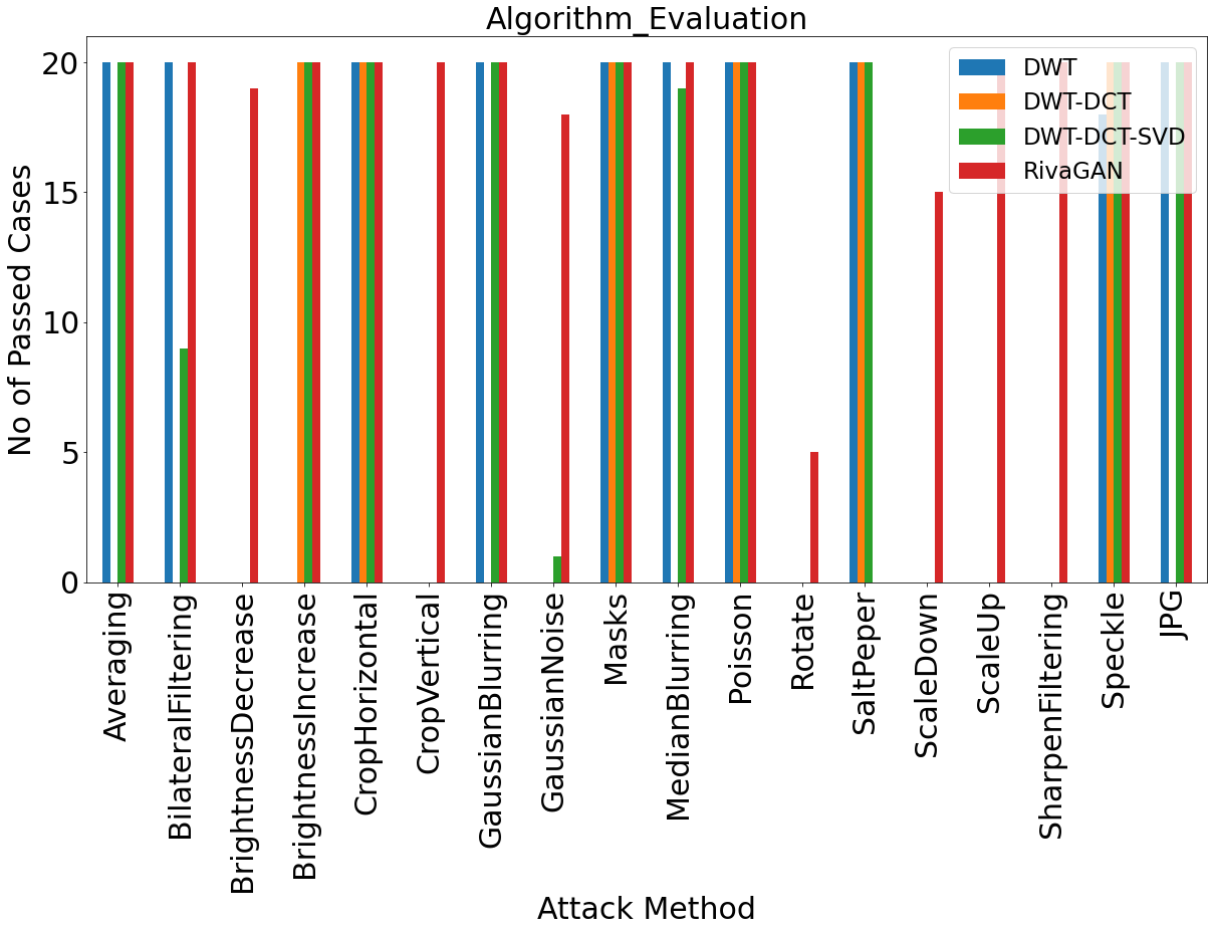

Figure 3. Attack Evaluation Based on Pass Rate.

As illustrated in Figure 2 & 3, RivaGAN has the highest passing rate among all the algorithms follow by DWT-DCT-SVD ranked at the second place. However, DWT-DCT has the worst performance with lowest passing cases.

Table 3. PSNR between Original Image and Encoded Image.

	DWT	**DWT-DCT**	**DWT-DCT-SVD**	**RivaGAN**
MRI_A	35.24	43.74	46.98	40.41
MRI_B	35.26	43.67	46.94	40.39
MRI_C	35.24	43.74	47.00	40.41
MRI_D	35.25	43.63	46.92	40.42
MRI_E	35.25	43.61	46.92	40.42
MRI_F	35.48	42.82	44.77	40.49
MRI_G	35.18	43.59	46.90	40.41
MRI_H	35.27	43.67	46.94	40.43
MRI_I	35.19	43.66	46.98	40.43
MRI_J	35.19	43.60	46.92	40.45
MRI_K	35.19	43.57	46.89	40.44
MRI_L	35.21	43.67	46.99	40.44
MRI_M	35.22	43.62	46.94	40.44

MRI_N	35.23	43.62	46.94	40.44
MRI_O	35.16	43.62	46.95	40.44
MRI_P	35.21	43.61	46.92	40.49
MRI_Q	35.36	43.74	46.96	40.45
MRI_R	35.28	43.65	46.95	40.42
MRI_S	35.23	43.60	46.93	40.41
MRI_T	35.20	43.59	46.90	40.43

The higher PSNR the better the quality of the compressed, or reconstructed image. Based on Table 3, DWT-DCT-SVD algorithm has the highest PSNR value with an average 46.83 dB that determine its criteria as best algorithm among all the algorithms.
re

Table 4. NCC between Original Image and Encoded Image.

	DWT	**DWT-DCT**	**DWT-DCT-SVD**	**RivaGAN**
MRI_A	0.9973	0.9992	0.9998	0.9992
MRI_B	0.9969	0.9990	0.9998	0.9990
MRI_C	0.9974	0.9992	0.9998	0.9992
MRI_D	0.9973	0.9992	0.9998	0.9992
MRI_E	0.9963	0.9989	0.9997	0.9986
MRI_F	0.9962	0.9983	0.9992	0.9988
MRI_G	0.9965	0.9989	0.9997	0.9989
MRI_H	0.9976	0.9993	0.9998	0.9993
MRI_I	0.9967	0.9990	0.9998	0.9993
MRI_J	0.9975	0.9993	0.9998	0.9992
MRI_K	0.9978	0.9993	0.9998	0.9993
MRI_L	0.9978	0.9993	0.9998	0.9993
MRI_M	0.9978	0.9993	0.9998	0.9993
MRI_N	0.9996	0.9990	0.9998	0.9990
MRI_O	0.9956	0.9987	0.9969	0.9987
MRI_P	0.9973	0.9992	0.9998	0.9992
MRI_Q	0.9968	0.9990	0.9998	0.9990
MRI_R	0.9969	0.9991	0.9998	0.9990
MRI_S	0.9962	0.9988	0.9997	0.9988
MRI_T	0.9972	0.9991	0.9998	0.9991

The higher NCC value the better the degree of similarity between two compared images. Based on Table 4, all the algorithms have similar performance on NCC value. However, DWT-DCT-SVD is the best performance with highest NCC value among all the algorithms.

Table 5. Encoded Algorithm vs Transmission Platform.

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
WhatsApp Image	✓	×	✓	✓
WhatsApp Document	✓	✓	✓	✓
Google Drive	✓	✓	✓	✓
Facebook Messenger	✓	✓	✓	✓
Gmail	✓	✓	✓	✓

As illustrated in Table 5, all algorithms managed to achieve full passes for every transmission method. However, DWT-DCT algorithm failed to achieve full passes as it failed to transmit across WhatsApp Image. It can be believed that this failure was caused by the compression of WhatsApp.

Table 6. Algorithm Implementation Benchmarking.

	DWT	DWT-DCT	DWT-DCT-SVD	RivaGAN
Character Length Limit	×	×	×	✓ (4)
Case Sensitive	✓	✓	✓	✓
Special Characters	✓	✓	✓	✓
Chinese Characters	×	×	×	×

Based on the Table 6, RivaGAN has the restrictions of 4 characters length, while other algorithms have no character length limit. Besides, every algorithm implementation exhibits perfect behaviours toward case sensitive and special characters. On the Chinese characters, all the algorithms fail to encode and decode.

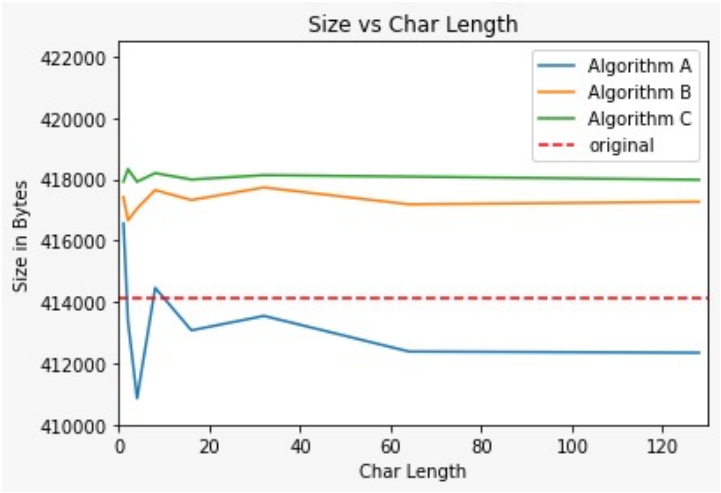

Figure 4. File Size vs Embedded Characters Length.

When comparing the length of character embedded into the image using Algorithm A (DWT), Algorithm B (DWT-DCT) and Algorithm C (DWT-DCT-SVD), as shown in Figure 4, the file size exhibited a big fluctuation for the first few 20 bytes. It can be observed that DWT went below the original file size when encoded with messages.

5.0 LIMITATIONS

As for the limitation of our study, RivaGAN was initially developed for video invisible watermarking. Then the implementation was redeveloped and ported to image watermarking. Therefore, the RivaGAN library we are using has limitations on the number of characters of 32bit allowed in the embedding process.

Other than that, the hybrid methods' implementations we are currently using are the publicly available open-source library from GitHub, as such, the implementation method might have a disparity with the original algorithm or research. Future studies can explore how the algorithm is implemented according to the formula to ensure consistency of the outcome.

6.0 CONCLUSIONS

Through our study, the deep-learning-based method of RivaGAN does exhibit the state-of-the-art robustness as claimed in the paper of RivaGAN authors (Zhang et al., 2019). Through our tests, we can confirm the feasibility of deep-learning-based invisible watermarking techniques as claimed by Vukotic (Vukotic et al., 2020).

Many extreme attacks were performed on RivaGAN's watermarked images and it was still able to pass all tests as it exhibited strong robustness as compared to other non-deep-learning-based watermarking techniques.

To satisfy the requirements of medical image watermarking, the algorithm shall have the high value of PSNR, NCC and exhibit a strong robustness. It was found that RivaGAN fails to surpass the PSNR and NCC value of DWT-DCT-SVD. This can be attributed to the nature of RivaGAN which is created specifically for video invisible watermarking.

The non-deep-learning-based hybrid algorithm of DWT-DCT-SVD showed the best criteria of Imperceptibility as it topped the PSNR value of 47.00 on comparing original image and encoded image. DWT-DCT-SVD also showed the best NCC value among the algorithms. With the feasibility of deep-learning-based invisible watermarking methods being confirmed through its strong robustness, more research is needed to refine the algorithm in terms of reaching higher PSNR and NCC as compared to the non-deep-learning-based methods.

7.0 ACKNOWLEDGEMENTS

The authors would like to thank Tunku Abdul Rahman University College (TAR UC) for providing financial support and technical support when completing this study.

REFERENCES

Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020 Feb;28:104863. DOI: 10.1016/j.dib.2019.104863.

Al-Haj, A. (2007). Combined DWT-DCT Digital Image Watermarking. Journal of Computer Science, 3(9), 740–746.

Abdulrahman, A. K., & Ozturk, S. (2019). A novel hybrid DCT and DWT based robust watermarking algorithm for color images. Multimedia Tools and Applications, 78(12), 17027–17049.

El-Shazly, E. H. M. (2004). Digital Image Watermarking in Transform Domains. Minufiya University.

Khare, P., & Srivastava, V. K. (2020). A Secured and Robust Medical Image Watermarking Approach for Protecting Integrity of Medical Images. Transactions on Emerging Telecommunications Technologies, 32(2).

Kuang, L.-Q., Zhang, Y. and Han, X. (2009). A Medical Image Authentication System Based on Reversible Digital Watermarking. 2009 First International Conference on Information Science and Engineering.

Lala, H. (2017). Digital image watermarking using discrete wavelet transform. International Research Journal of Engineering and Technology (IRJET), 4(01).

Sverdlov, A., Dexter, S., & Eskicioglu, A. M. (2005). Robust DCT-SVD domain image watermarking for copyright protection: embedding data in all frequencies. In 2005 13th European Signal Processing Conference (pp. 1-4). IEEE.

Tao H, Chongmin L, Zain JM, Abdalla AN (2014) Robust image watermarking theories and techniques: a review. J Appl Res Technol 12(1):122–138

Voloshynovskiy, S., S. Pereira and T. Pun, 2001. "Attacks on Digital Watermarks: C lassification, Estimation-Based Attacks, and Benchmarks," Comm. Magazine, 39(8): 118-126

Vukotić, V., Chappelier, V., & Furon, T. (2020). Are Classification Deep Neural Networks Good for Blind Image Watermarking? Entropy, 22(2), 198.

Yeung, M. M. (1998). Invisible watermarking for image verification. Journal of Electronic Imaging, 7(3), 578.

Zhang, K. A., Xu, L., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Robust invisible video watermarking with attention. arXiv preprint arXiv:1909.01285.

Zhang, L., Li, W., Ye, H. (2021). A blind watermarking system based on deep learning model. 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).