# Comparative Analysis of Lossless Image Compression Algorithms based on Different Types of Medical Images

1st Mona Alzahrani
*ICS Department*
*KFUPM*
Dhahran, Saudi Arabia
g201908310@kfupm.edu.sa

2nd Mona Albinali
*ICS Department*
*KFUPM*
Dhahran, Saudi Arabia
g201906770@kfupm.edu.sa

2021 International Conference of Women in Data Science at Taif University (WiDSTaif ) | 978-1-6654-4948-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/WIDSTAIF52235.2021.9430242

***Abstract*—In the medical field, there is a demand for high-speed transmission and efficient storage of medical images between healthcare organizations. Therefore, image compression techniques are essential in that field. In this study, we conducted an experimental comparison between two famous lossless algorithms: lossless Discrete Cosine Transform (DCT) and lossless Haar Wavelet Transform (HWT). Covering three different datasets that contain different types of medical images: MRI, CT, and gastrointestinal endoscopic images; with different image formats PNG, JPG and TIF. According to the conducted experiments, in terms of compressed image size and compression ratio, we found that DCT outperforms HWT regarding PNG and TIF format which represent CT-grey and MRI-color images. And regarding JPG format which represents the gastrointestinal endoscopic color images, DCT performs well when grey-scale images are used; where HWT outperforms DCT when color images are used. However, HWT outperforms DCT in compression time regarding all the image types and formats.**

***Index Terms*—Compression Algorithms, Medical Images, Lossless Image Compression, Discrete Cosine Transform, Haar Wavelet Transform.**

## I. Introduction

Image compression is a technique used widely to reduce the image size during storing and processing of the image. With the increasing quality and size of the images, compression has become essential in day-to-day life [1]. Moreover, due to the advent of technology, there exist different formats such as JPG, TIF, PNG, Gif ... etc; and high-resolution images are produced and require more storage memory. Furthermore, the issue arises through the internet connection when we want to transmit a huge number of images for some purposes which also affects the transmission time [2]. In some fields, the memory and transmission time is not a major issue due to the availability of the different type of storage or high-speed internet connection. However, these issues are essential in the medical field due to the development of technologies in digital medical imaging such as computed tomography (CT) and magnetic resonance imaging (MRI) which are produced as a series of image sequences. These sequence images require efficient storage of images and high-speed transmission between healthcare organizations [3]. To deal with this issue some compression techniques which are either lossless or lossy image compression are required. In this field, lossless algorithms are preferred for two reasons: i) ensuring no useful clinical information is lost, and ii) it is forbidden by law in some countries to lossy compress images used for medical diagnosis [4], [5].

Different studies [4]–[9] comparing different lossless image compression algorithms which have been used in the medical field. Those studies have been done remarkable and practical works. However, some of these studies were only a theoretical comparison by providing a review or survey [5], [6]. And all of these studies were focus in only one or two type of medical images and most of them grey CT or MRI. Moreover, they cover one data type of image either color or grey images.

So, that inspire us to implement a new comparative analysis for comparing different lossless image compression algorithms that have been used with the medical images. In our study, to achieve the primary goal which is conduct an experimental comparison between two famous lossless algorithms: lossless Discrete Cosine Transform (DCT) and lossless Haar Wavelet Transform (HWT). Covering three different datasets: COVID-19, Kvasir-SEG, and TCGA-LGG, that containing different type of medical images: CT, MRI, and gastrointestinal endoscopic images; with different image formats PNG, JPG and TIF. Moreover, these datasets cover the two type of color and grey images.

The structure of the rest of the paper is as follows: firstly, Section II highlights the comparative related works that compared different lossless image compression algorithms used in the medical field. Followed by Section III that illustrates our methodology which contains four phases: lossless compression algorithms selection, medical image datasets selection, image compression app development, experiment design. Section IV demonstrate the obtained results and tackles the discussion based on our observations. Lastly, Section V concludes this study and includes our future works.

978-1-6654-4948-9/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.

## II. Literature Review

Several research papers have been conducted theoretical [5], [6] or experimental [4], [7]–[9] comparison between image compression algorithms.

For the experimental comparison, Schaefer et al. [4] evaluated many famous lossless image compression algorithms with medical infrared images. Lossless JPEG, JPEG2000, JPEG-LS, CALIC, and PNG have been tested on a dataset of more than 380 thermal images. This resulted that JPEG-LS have the highest performance algorithm for compressing medical infrared images in terms of the largest compression ratio and the highest compression speed. Moreover, Amri et al. [7] offered a novel approach, which used to reduce the image and used the expansion techniques with digital watermarking and lossless compression standards such as JPEG-LS (JLS) and TIF formats. Their experiment has been carried on two medical image databases: one contains 30 MRI images while the other contains 30 CR Images. In their method, for the compression stage, they used the square-square decimation with "nearest neighbor interpolation", "zero-padding", "cubic interpolation" and "transformed B-Spline" expansion techniques in the decompression stage. Resulted that wREPro.JLS has better image quality than the JPEG compression for compression ratio. Even more, Ansari et al. [8] conducted a comparative study of different types of medical image compression techniques for the telemedicine application. Their experiment has been performed on CT brain and chest images. By using DCT and wavelet techniques, all these images were compressed. Their results show that the wavelet transform-based techniques (JPEG 2000) have shown the best results after comparing with DCT based techniques (JPEG). The evaluation of the result is based on compression ratio, MSE, SNR, and PSNR for medical images. In another study, Clunie et al. [9] evaluated the new JPEG-LS process (ISO/IEC1495-1), the lossless mode of the proposed JPEG20scheme(ISO/IECD154-1), and new standard schemes DICOM. They tested 3,679 single frame gray-scale images from different anatomical areas. Only single gray-scale component images have been tested. This resulted that State-of-the-art lossless compression methods perform significantly better than older lossless image compression methods.

However, these studies [4]–[9] have been done remarkable and practical works which contribute by comparing different lossless image compression algorithms. However, some of these studies were only a theoretical comparison by providing a review or survey [5], [6]. And all of these studies were focusing on only one or two types of medical images and most of them grey CT or MRI. Moreover, they cover one data type of image either color or grey images. So, in this study, we conducted an experimental comparison between two famous lossless algorithms DCT and HWT. Covering three different datasets that containing different types of medical images: CT, MRI, and gastrointestinal endoscopic; with different image formats including PNG, JPG, and TIF. Moreover, these datasets cover the two data types of color and grey images.

## III. Methodology

Our main goal of this study is to analyze the performance of two well-known lossless image compression algorithms DCT and HWT based on different types and different formats of medical images for comparison. To achieve that goal, a methodology of four basic phases was followed to conduct this study. In the first phase, we determine the used lossless image compression algorithms DCT and HWT as explain in Section III-A. While in the second phase, we collect the different types and formats of medical images from three well-known datasets as we will detail them in Section III-B. After that, in the third phase, we design and program our image compression application in Section III-C. Finally, in the fourth phase, we designed and conducted our comparative experiments in Section III-D.

### A. Image Compression Algorithms

In this section, we will describe briefly the lossless DCT and HWT algorithms that used in this paper for the comparison purpose.

*1) Lossless Discrete Cosine Transform (DCT) Algorithm:* The finding of DCT was in 1974 which is a significant achievement for the image compression scientific research [10]. The DCT helps to divide the image into bits of varying value regarding the visual quality of the image. By using the DCT algorithm, it is possible to compress both data types which are gray-scale and color images [11], [12]. It is commonly used in many fields and it is important for different implementations in different areas including industries, WWW, engineering and science ... etc [12]. Moreover, it has different advantages namely satisfactory performance, simplicity, and availability of special purpose hardware for implementation. It worth mention, that there are two types of it: lossless and lossy DCT. The lossless DCT has shown many advantages such as the compatibly with JPEG at high compression ratios and its shown better performance than LJPG in lossless [13]. In this work, we will focus on the lossless DCT algorithm.

*2) Lossless Haar Wavelet Transform (HWT) Algorithm:* Haar functions were first used in 1910 when the Hungarian mathematician Alfred Haar developed them [14]. HWT is one of the easiest and most essential transformations of a local frequency domain and a space domain. It divided each signal into two components. One component is referred to as the average and the other is referred to as the difference. It is used to decrease the requirements for memory and the amount of inefficient movement of the coefficients of Haar [12]. One of its advantages, that its basis functions have variable length and there is no need to block the input image. And because it is Wavelet-based coding, HWT is more robust under decoding errors and facilitates progressive transmission of images [11]. The main drawback is the number of subtraction and addition operations which can be balanced by reducing the division operations [12].

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.

### B. Medical Image Datasets

An image collection that represents the variety of types of medical images that are usually captured is needed to give a useful performance comparison of lossless compression algorithms. We have therefore used three datasets (see Table I) that represent different image types. These datasets are as follow:

1) COVID-19 Dataset [15]: an open-source available COVID-CT dataset, including 349 lung CT scans correspond to 216 patients that are positive for COVID-19 and 397 lung CT scans that are negative for COVID-19 which anticipate whether a person is affected with COVID-19 by assessing his/her CTs. In this work, we use a sample of 20 CT images where 10 of them are positive for COVID-19 and the rest are negative for COVID-19. Images are provided in PNG format for positive COVID-19 images and JPG format for negative COVID-19 images. The sample is shown in Figure 1 where the first two rows specify for COVID-19 positive CTs and the last two rows specify for COVID-19 negative CTs.
2) Kvasir-SEG Dataset [16]: it is an open-source of gastrointestinal polyp images and its corresponds to segmentation masks, manually annotated and verified by an experts gastroenterologist from Vestre Viken Health Trust in Norway. It comprises 8000 gastrointestinal tract images. All images are provided in JPG format. We take a sample of 20 different images from 20 different patients which are shown in Figure 2.
3) The Cancer Genome Atlas Low Grade Glioma (TCGA-LGG) Dataset [17]: this dataset contains brain MRI images together with manual FLAIR abnormality segmentation masks. They correspond to 110 patients included in The Cancer Genome Atlas (TCGA). All images are provided in TIF format. Masks are binary, 1-channel images. The images of TCGA-LGG Dataset were obtained from The Cancer Imaging Archive (TCIA) [18] which is a service that has a large archive of medical images of cancer that can be accessed by the public and allowed for downloading. From this dataset, we take a sample of 20 different images from 20 different patients which are shown in Figure 3.

So, in this work, we used 60 medical images; every 20 images of them belong to a different dataset and cover different type of medical images (CT, MRI, and gastrointestinal endoscopic images), different image formats (PNG, JPG, and TIF) and different image data type (color and grey-scale).

### C. Image Compression Application

Our application interface has been designing using GUIDE (Graphical User Interface Development Environment) [19]. Where the programming of the image compression code was written by MATLAB. Figure 4 show an image sample from COVID-19 (PNG format).

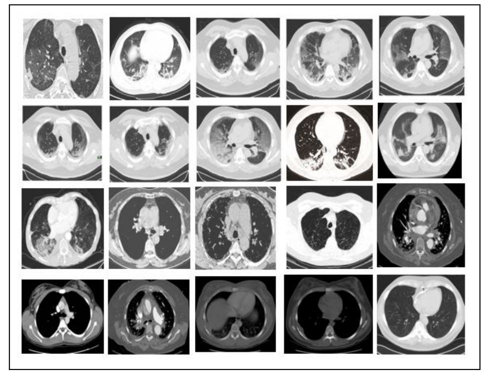

Fig. 1. Sample of 20 lung CT images from COVID-19 dataset [15].

Fig. 2. Sample of 20 gastrointestinal endoscopic images from Kvasir-SEG dataset [16].

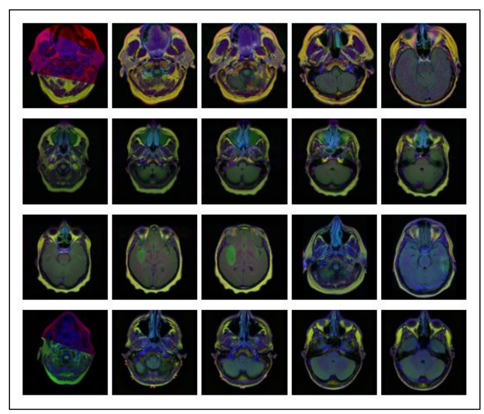

Fig. 3. Sample of 20 brain MRI images from TCGA-LGG dataset [17].

### D. Experiments Design

As mentioned early, in this study, we will compare the lossless DCT and HWT image compression algorithms based on three comparison/evaluation criteria. These criteria used for selecting the best-performing algorithms were defined as the following:

1) Compressed Image Size (CIS): which is the size of the image after compressed with the algorithm.
2) Compression Ratio (CR): which is the ratio of the original image size to the compressed image size.
3) Compression Time (CT): is the time taken by the algorithm to compress the image [20].

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DESCRIPTION OF THE USED MEDICAL IMAGE DATASETS

| Dataset # | Dataset Name | Image Type | Image Format | Image Size | Ref. |
|---|---|---|---|---|---|
| 1 | COVID-19 | Lung CT (grey-scale) | PNG and JPG | Vary from 43 to 375 Kb | [15] |
| 2 | Kvasir-SEG | Gastrointestinal endoscopic (color) | JPG | Vary from 23 to 53 Kb | [16] |
| 3 | TCGA-LGG | Brain MRI (color) | TIF | Fixed to 198 Kb | [17] |

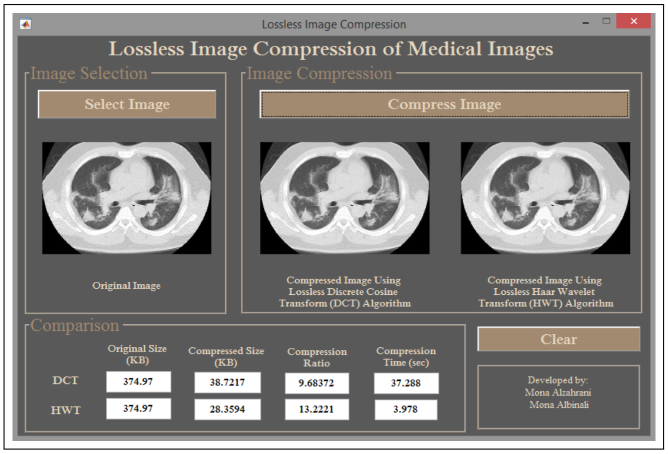


Fig. 4. PNG test image sample of COVID-19 dataset (image number 10) a) Original image, b) Compressed image with DCT, and c) Compressed image with HWT.

An algorithm with a higher compression ratio, compression size, and lower compression time implies better performance.

So, the experimental comparison conducted in this study is designed as three experiments. Where in each experiment, we compare the lossless DCT and HWT algorithms based on different medical image datasets. These experiments are:

- Experiment 1: Compress the 20 images from the COVID-19 dataset using our application to compare the lossless DCT and HWT algorithms in terms of compressing the grey CT medical image in PNG and JPG formats.
- Experiment 2: Compress the 20 images from the Kvasir-SEG dataset using our application to compare the lossless DCT and HWT algorithms in terms of compressing the color gastrointestinal endoscopic medical image in JPG format.
- Experiment 3: Compress the 20 images from the TCGA-LGG dataset using our application to compare the lossless DCT and HWT algorithms in terms of compressing the color MRI medical image in TIF format.

## IV. RESULTS AND DISCUSSION

Table II summarized the final averaged (A-) results of the three experiments. While Figures 5, 6, 7, 8, 9, and 10, represent the results of these experiments regarding the compressed image sizes and compression ratio which show the impact of compress different image types and format on the original sizes of these images. While the compression times for all the three datasets are shown in Figures 11, 13, and 12 which illustrated how much time taken to compress these images.

For *experiment 1*, using COVID-19 dataset, which contains PNG and JPG images format that have 137.465Kb average size, it can be observed that DCT and HWT are efficient techniques to reduce the images sizes to 13.505Kb and 27.51Kb average compressed size with average compression ratios of 10.5 and 5.0315 using DCT and HWT respectively which is helpful in the storage. However, as shown in Figures 5 and 8, the sizes of both the ten first images which represent PNG format, and the sizes of the rest ten images which represent JPG format, are dramatically reduced using both DCT and HWT techniques but DCT outperforms the HWT.

For *experiment 2*, using KVASIR-SEG dataset, which contains JPG images format that have 39.335Kb average size, it is obviously from Figures 6 and 9, that there are no much differences between the two algorithms in terms of compressed image sizes and compression ratio and that duo to the already small size of the original images. The image sizes were reduced to 30.16Kb and 26.265Kb average size with average compression ratios of only 1.29 and 1.5 using DCT and HWT respectively. But in general, HWT outperforms DCT in reducing the image sizes.

For *experiment 3* using TCGA-LGG dataset, which contain TIF images format the same 198.3Kb size, as illustrated in Figure 7, the results showed no noticeable differences of performance in terms of reducing the image sizes among the two compression algorithms. However, the DCT reduces the average image size to 6.339Kb while HWT reduces it to 15.427Kb which indicates that DCT has the best performance as also noticeable in the compression ratio in Figure 10.

According to results from the three previous experiments, we found the followings: i) DCT outperforms HWT in terms of compressed image size and compression ratio regarding PNG images and TIF images, which also represent CT-grey image and MRI-color image; ii) regarding compressed image sizes and a compression ratio of JPG format which represented by gastrointestinal endoscopic-color, DCT performs well when grey-scale images used where HWT outperform it when color images used which lead us to an open research question; and iii) HWT outperforms DCT in terms of compression time regarding all the image types and formats as obviously shown in Figures 11, 12, 13 and which illustrated how much time taken to compress these images.

## V. CONCLUSION AND FUTURE WORK

In this study, we conducted an experimental comparison between two famous lossless algorithms: DCT and HWT. Covering three different medical image datasets that containing different types of medical images: MRI, CT, and gastrointestinal endoscopic images; with different image formats: PNG,

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.

TABLE II
SUMMARY OF AVERAGED RESULTS USING LOSSLESS DCT AND HWT COMPRESSION ALGORITHMS (A- STAND FOR AVERAGE)

| Dataset # | Dataset Name | Image Format | Original Size (Kb) | DCT | | | HWT | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | A-CIS (Kb) | A-CR | A-CT (sec) | A-CIS (Kb) | A-CR | A-CT (sec) |
| **1** | **COVID-19** | PNG | 178.28 | 16.51 | 11.89 | 16.2 | 29.35 | 5.933 | 3.11 |
| | | JPG | 96.65 | 10.5 | 9.11 | 11.51 | 25.67 | 4.13 | 3.2 |
| | | PNG and JPG | 137.465 | 13.505 | 10.5 | 13.855 | 27.51 | 5.0315 | 3.155 |
| **2** | **Kvasir-SEG** | JPG | 39.335 | 30.16 | 1.29 | 30.56 | 26.265 | 1.5 | 3.265 |
| **3** | **TCGA-LGG** | TIF | 198.3 | 6.339 | 32.005 | 6.517 | 15.427 | 13.4445 | 3.4175 |

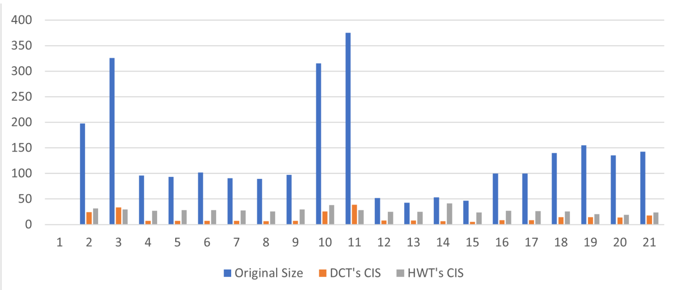


Fig. 5. Results of images size in Covid-19 Dataset

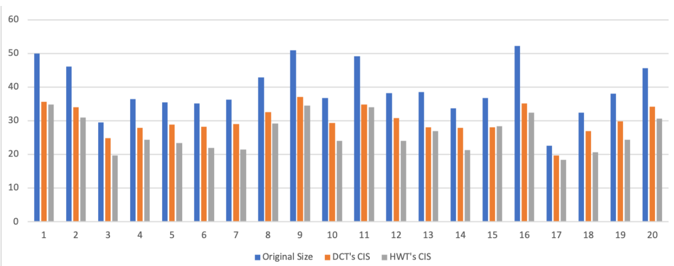


Fig. 6. Results of images size in Kvasir Dataset

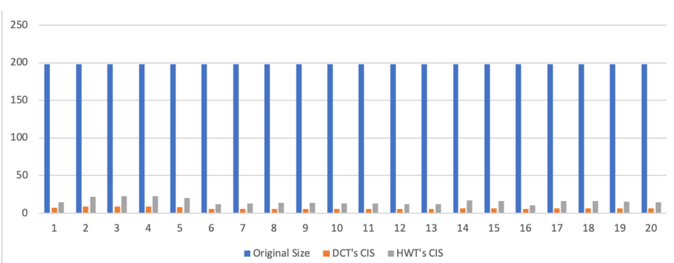


Fig. 7. Results of images size in TCGA Dataset

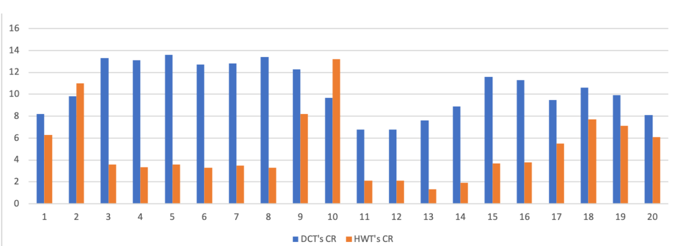


Fig. 8. Results of images compression ratio in Covid-19 Dataset

JPG, and TIF. Even more, these datasets covered the two types

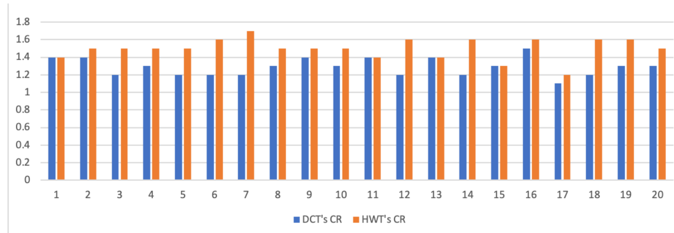


Fig. 9. Results of images compression ratio in Kvasir Dataset

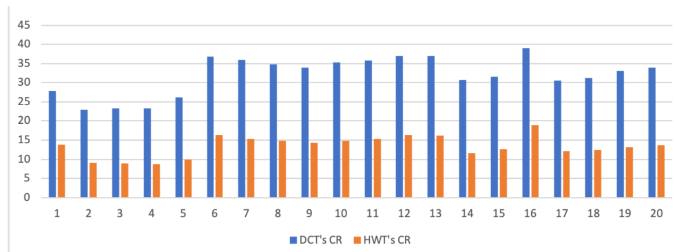


Fig. 10. Results of images compression ratio in TCGA Dataset

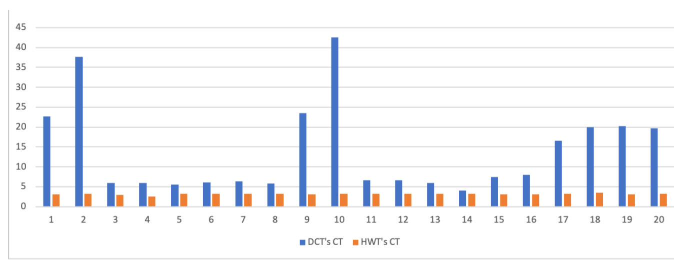


Fig. 11. Results of images compression time in Covid-19 Dataset

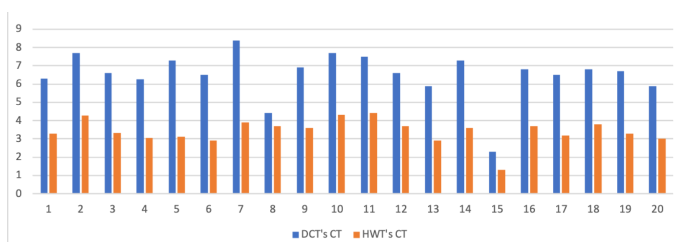


Fig. 12. Results of images compression time in TCGA Dataset

of color and grey-scale images.

For future work, this work could be extended by including the comparison evaluation criteria: RMSE, SNR, and compression percentage for the proposed method. Moreover,

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.

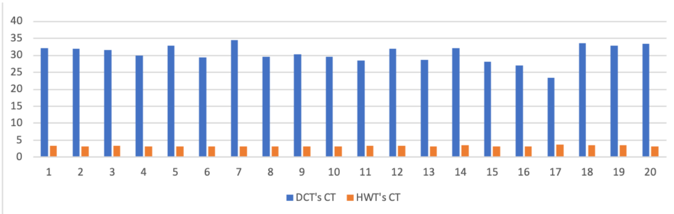


Fig. 13. Results of images compression time in Kvasir Dataset

we encourage the researcher to include the decompression techniques to restore the original images with focusing on image resolution. Even more, the datasets could be extended in terms of the number, types, and formats of medical images.

## VI. Acknowledgment

We would like to thank Dr. Imane Boudellioua for sharing her wisdom, valuable feedback, guidance, and encouragement in completing this work.

## References

[1] S. T. Kodukulla, "Lossless image compression using matlab: Comparative study," 2020.
[2] N. Sikka, S. Singla, and G. P. Singh, "Lossless image compression technique using haar wavelet and vector transform," in *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*. IEEE, 2016, pp. 1–5.
[3] X. Song, Q. Huang, S. Chang, J. He, and H. Wang, "Lossless medical image compression using geometry-adaptive partitioning and least square-based prediction," *Medical & Biological Engineering & Computing*, vol. 56, no. 6, pp. 957–966, 2018.
[4] G. Schaefer, R. Starosolski, and Shao Ying Zhu, "An evaluation of lossless compression algorithms for medical infrared images," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 1673–1676.
[5] S. S. Me, V. Vijayakuymar, and R. Anuja, "A survey on various compression methods for medical images," vol. 4, no. 3, 2012, p. 13.
[6] M. F. Ukrit, A. Umamageswari, and G. Suresh, "A survey on lossless compression for medical images," vol. 31, no. 8. Citeseer, 2011, pp. 47–50.
[7] H. Amri, A. Khalfallah, M. Gargouri, N. Nebhani, J.-C. Lapayre, and M.-S. Bouhlel, "Medical image compression approach based on image resizing, digital watermarking and lossless compression," *Journal of Signal Processing Systems*, vol. 87, no. 2, pp. 203–214, 2017.
[8] M. Ansari and R. Anand, "Comparative analysis of medical image compression techniques and their performance evaluation for telemedicine," in *Proceedings of the International Conference on Cognition and Recognition*. Citeseer, 2014, pp. 670–677.
[9] D. A. Clunie, "Lossless compression of grayscale medical images: effectiveness of traditional and state-of-the-art approaches," in *Medical Imaging 2000: PACS Design and Evaluation: Engineering and Clinical Issues*, vol. 3980. International Society for Optics and Photonics, 2000, pp. 74–84.
[10] S. S. Me, V. Vijayakuymar, and R. Anuja, "A survey on various compression methods for medical images," *International Journal of Intelligent Systems and Applications (IJISA)*, vol. 4, no. 3, p. 13, 2012.
[11] S. Dhawan, "A review of image compression and comparison of its algorithms," *International Journal of electronics & Communication technology*, vol. 2, no. 1, pp. 22–26, 2011.
[12] K. Kaur and S. Malhotra, "Image compression using haar wavelet transform and discrete cosine transform," *International Journal of Computer Applications*, vol. 125, no. 11, 2015.
[13] W. Philips, "The lossless dct for combined lossy/lossless image coding," in *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269)*. IEEE, 1998, pp. 871–875.
[14] R. S. Stanković and B. J. Falkowski, "The haar wavelet transform: its status and achievements," *Computers & Electrical Engineering*, vol. 29, no. 1, pp. 25–44, 2003.
[15] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient deep learning for covid-19 diagnosis based on ct scans," *medrxiv*, 2020.
[16] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," pp. 451–462, 2020.
[17] M. Buda, A. Saha, and M. A. Mazurowski, "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm," *Computers in Biology and Medicine*, vol. 109, pp. 218 – 225, 2019.
[18] The Cancer Imaging Archive (TCIA), year = 2020, url = https://www.cancerimagingarchive.net/, urldate = 2020-09-30.
[19] B. R. Hunt, R. L. Lipsman, and J. M. Rosenberg, *A guide to MATLAB: for beginners and experienced users*. Cambridge university press, 2014.
[20] S. Di and F. Cappello, "Fast error-bounded lossy hpc data compression with sz," in *2016 ieee international parallel and distributed processing symposium (ipdps)*. IEEE, 2016, pp. 730–739.

Authorized licensed use limited to: National Remote Sensing Agency. Downloaded on May 02,2024 at 05:48:05 UTC from IEEE Xplore. Restrictions apply.