DOI: 10.21917/ijivp.2022.0384

LOSSLESS IMAGE COMPRESSION USING DIFFERENT ENCODING ALGORITHM FOR VARIOUS MEDICAL IMAGES

T. Sujatha and K. Selvam
Department of Computer Applications, Dr. MGR Educational and Research Institute, India

Abstract
In the medical industry, the amount of data that can be collected and kept is currently increasing. As a result, in order to handle these large amounts of data efficiently, compression methods must be re-examined while taking the algorithm complexity into account. An image processing strategy should be explored to eliminate the duplication image contents, so boosting the capability to retain or transport data in the best possible manner. Image Compression (IC) is a method of compressing images as they are being stored and processed. The information is preserved in a lossless image compression technique which allows for exact image reconstruction from compressed data with retain the quality of image to higher possible extend but it does not significantly decrease the size of the image. In this research work, the encoding algorithm is applied to various medical images such as brain image, dental x-ray image, hand x ray images, breast mammogram images and skin image can be used to minimize the bit size of the image pixels based on the different encoding algorithm such as Huffman, Lempel-Ziv-Welch (LZW) and Run Length Encoding (RLE) for effective compression and decompression without any quality loss to reconstruct the image. The image processing toolbox is used to apply the compression algorithms in MATLAB. To assess the compression efficiency of various medical images using different encoding techniques and performance indicators such as Compression Ratio (CR) and Compression Factor (CF). The LZW technique compresses binary images; however, it fails to generate a lossless image in this implementation. Huffman and RLE algorithms have a lower CR value, which means they compress data more efficiently than LZW, although RLE has a larger CF value than LZW and Huffman. When fewer CR and more CF are recorded, RLE coding becomes more viable. Finally, using state-of-the-art methodologies for the sample medical images, performance measures such as PSNR and MSE is retrieved and assessed.

Keywords:
Lossless Image Compression, Huffman Coding, Lempel-Ziv-Weich, Run Length Encoding

1. INTRODUCTION

Medical research and clinical practice both rely heavily on image processing [1,2]. Medical image analysis has advanced significantly in recent years as a result of improved digital imaging systems. There have been a great number of medical images created, with growing variation and quality. Despite their moderate success, traditional medical image analysis approaches are incapable of coping with huge volumes of image data [3]-[5]. The use of digital computers to process digital images is known as digital image processing. Digital images are, in fact, a unique synthesis of a small number of components. Pixels, images, and image elements are all elements that have a position and a value. The term pixel refers to the components of a computer image. Medical images, such as CT scans, MRIs, and X-rays, are now the most widely used images in medical research. As a result, complex automated techniques are required to analyze these various image forms. IC is a method for storing and transmitting images effectively while maintaining the greatest possible quality. The image data storage is reduced by deleting redundant or undesirable data. IC major purpose is to lower the amount of data saved while preserving acceptable image quality. It consists of two parts namely encoding and decoding. The original pixels are handled to encode and turned to binary bits in encoding, whereas the output bits of encoded are applied to decode and converted back to the original pixels in decoding. Many applications and methodologies [6] [7] attempt to solve this compression challenge by striking a good compromise between reconstructed image quality and compression ratio.

For both domains, the strategies that involved lossless compression without impacting image quality. Lossy compression is a criterion for image compression that reduces psychovisual redundant data. When compared to lossless compression, it yields a high compression ratio. It accepts and compromises image quality decrease. It may sometimes cause serious problems in areas like as satellite imaging, biomedical imaging, and the medical industry. Smart gadgets, tablets, and laptops are becoming more popular every day, and they are gaining access to imaging fields in order to develop network topology. By using a simple technique with minimal computing cost, it does not tolerate information loss in decoded images.

There are four types of lossless compression algorithms: 1. Huffman Coding, 2. Run-Length Encoding, 3. Entropy Encoding, and 4. Arithmetic Encoding. The bit planes are used to encode the images using the RLE technique. The RLE technique is used to encode higher-level bit planes, whereas arithmetic coding is used to encode lower-level bit planes. A great compression efficiency is achieved by combining the arithmetic encoding with RLE. This compression approach is best for gray-scale images [8]. For lossy image compression of gray-scale images, the LZW method is used, with a compression effectiveness of roughly 40%. To accomplish lossless compression, a GIF encoder utilizing the LZW technique is used [9].

The gray-scale images are compressed using Huffman encoding as well. This is a lossless compression method that minimizes the size of the source symbols before compressing them with Huffman Coding. The approach achieves a 10% higher compression ratio than standard Huffman encoding [10]. This study work focuses on compressing various medical images using several encoding algorithms such as Huffman coding, LZW, and RLE to reduce the bit size of the image pixels for effective compression and decompression without any quality loss to reconstruct the image.

The following is a breakdown of how the paper is structured: Section 1 covers the introduction to lossless compression, Section 2 discusses related works based on lossless image compression, Section 3 discusses research methodology based on encoding approach, Sections 4 and 5 cover the results and conclusion, respectively.

2. LITERATURE REVIEW

Gajendra Sharma [11] looked into various techniques to see whether there are any variations in compression time, compression ratio and decompression time. Huffman Coding has a better location than LZW Coding. The LZW Coding system has a better compression ratio than the Huffman method. Huffman Coding takes longer than LZW Coding to complete. In other circumstances, time is not a factor since Huffman Coding may achieve a high compression ratio. Time is critical for some applications, such as real-time applications and LZW coding. Pourasad and Cavallaro [12] used two compression algorithms for IC in their study: lossless compression and lossy compression, both of which retain image quality. In addition, different image enhancing techniques is used to increase the quality of the compressed image. The efficiency of the measuring metrics outperforms that of other image processing approaches. The authors reviewed several medical image compression algorithms in [13]-[16]. The examined approaches have a distinct property; however, the medical images are compressed, which has certain limitations. As a consequence of the research, these weaknesses will be addressed, and the quality of a reconstructed medical image with a higher compression rate will be improved. In [17], the author proposed a novel method of image alteration for aesthetically acceptable images.

Karthikeyan et al. [18] examined several lossy and lossless compression algorithms, as well as experimental results, employing the Discrete Wavelet Transform (DWT) and Discrete Cosine Transform (DCT). In comparison to DCT based Arithmetic and DCT based Huffman Coder, DWT based arithmetic delivers good results in terms of CR and PSNR for input images. Arithmetic takes substantially longer than Huffman coder, based on the results of calculation time. The quality factor is raised, lowering the reconstructed image quality. Still, the most difficult aspect of image compression is achieving a high compression ratio with little deterioration in reconstructed image quality. Volume of Interest (VOI) coding, as described by Subramanian et al. [19], has the effect of boosting compression ratio while reducing bit rate. The VOI is effectively encoded and decoded using this manner. The VOI decreases the complexity of recreating the real 3D image volume with fewer reconstruction features by using the Selective Bounding Volume (SBV) approach. Furthermore, on evaluated 3D Volumes with low Bit Rate, our technique achieves a compression ratio that is twice as high as other available methods. Sowmyalakshmi et al. [20] propose an OLBG-LZMA image compression method that combines Optimized Linde–Buzo–Gray (OLBG) and Lempel Ziv Markov Algorithm (LZMA) coding to reduce microarray images without losing quality. In comparison to other compression algorithms, OLBG-LZMA coding obtained a substantial compression performance, according to the simulation findings.

Huffman Encoding, RLE, and LZW algorithms gain good interaction to control compression challenges, according to Divya et al. [21]. Each technique compresses data in a unique way, allowing this paper to come to a more satisfying conclusion. By encoding and decoding data, the above methods provide data security, allowing users to communicate more effectively. When compared to JPEG2000, PNG, and WebP, lossless IC has been deemed superior in several ways. It outperforms the JPEG2000, PNG, and WebP codecs in the lossless IC approach, which keeps the image quality while decreasing the bit (file size). For entropy coding, this probabilistic model of image compression is created [22] [23]. The bit planes have to encode the images using the RLE algorithm. The RLE algorithm is used to encode higher-level bit planes, while arithmetic coding is used to encode lower-level bit planes. A high compression efficiency is achieved by combining the arithmetic encoding and RLE. This compression approach is best for gray-scale images. In lossy IC of gray scale images, the LZW algorithm is used with a compression efficiency of around 40% [24]. In order to achieve lossless compression, a GIF encoder with the LZW algorithm is used [25]. The gray-scale images are compressed using Huffman encoding as well. This method that reduces the source symbols before compressing them using Huffman coding. The approach obtained 10% higher CR than standard Huffman encoding [26].

3. RESEARCH METHODOLOGY

DICOM (Digital Imaging and Communications in Medicine) is utilized in practically every imaging and radiation application in radiology, cardiology, and radiotherapy, as well as in tools in other medical professions comprising dentistry and ophthalmology. With hundreds of thousands of medical imaging equipment in use worldwide, one of the most often utilized healthcare communication protocols is DICOM. DICOM has enabled breakthrough medical imaging technologies that have changed the face of clinical care, much as the Internet has allowed current consumer information applications. From the emergency department to breast cancer detection and heart stress monitoring, DICOM is the standard that allows clinicians and patients to engage with medical imaging. This study used medical images from the National Library of Medicine Med-Pix to analyze and assess procedures. With approximately 500 patient cases, 800 themes, and over 1000 images, Med-Pix is a free, open-access online collection of medical images, case studies, and clinical topics that combines textual documentation with images. The images gathered are free of copyright issues and can be used by anybody [27].

3.1 LOSSLESS IMAGE COMPRESSION

One of the most essential image compression techniques is lossless compression. This lossless compression technique compresses an image without causing significant data loss. This implies that the image will be compressed but there will be no severe data loss, ensuring that all important information is preserved. This lossless compression is the ideal way for compressing medical images since it can compress the image without losing any data. Because of its recent quick expansion, lossless compression finds a lot of usage in the medical profession. As the number of hospitals and the quantity of case records grows, so does the demand for IC for storage and transmission. In these situations, it is critical that a case record be saved in a compressed format with no data loss due to compression. Lossless compression meets both instances in this fashion. However, as compared to lossy approaches, this comes at the cost of attaining low compression rate values. Entropy encoding methods are used in the majority of lossless compressions.

3.1.1 Huffman Coding (HC):

Huffman Coding is a typical encoding method used for lossless data compression. It is the basic Shannon source coding method improvisation algorithm. Shannon source coding is modified by Huffman coding, which developed a unique manner of allocating codewords so that longer codewords are given to less common symbols and short codewords are assigned to the most recurrent symbols. Initially, it computes the probability of every symbol appearing in an input. The probabilities are then sorted, and the lowest symbol probabilities are combined, leaving just two probabilities in the procedure. The codewords are allocated using the appropriately constructed Huffman tree. The data may be easily decrypted using built-in codewords. The algorithmic phases involved in HC are listed below.

Algorithm for HC

Step 1: [0, 1] is the range of real numbers. According to the symbols in the input, this range is separated into subranges. This subrange represents a real value equal to the source symbol probability.

Step 2: Determine which subrange each input symbol belongs to.

Step 3: To produce the next level subranges, these ranges are subdivided once more.

Step 4: Continue with the following symbol in the same manner.

Step 5: Steps 3 and 4 are repeated until the entire input has been parsed.

3.1.2 Run Length Coding (RLC):

The most basic way for compressing any form of visual data is run length coding. It operates by taking use of the redundancy in pixel values in an image. This approach generates an ordered pair consisting of the pixel intensity value and the pixel matching consecutive length. This method is simple to apply, but it only works effectively for images with a lot of repeating data, such as medical scans. The RLC algorithmic steps are mentioned below.

Algorithm for RLC

Step 1: From the initial pixel of an image, read the pixel constantly. If it is the last pixel, it is time to quit.

Step 2: If the following pixel value is the same as the previous one, the count value is increased by one. Otherwise, create a new array to hold the pixel value.

Step 3: Lastly, using the intensity and its matching length, the sequence of ordered pairs is obtained.

Step 4: For reconstruction, an empty array of input size is created.

Step 5: In order to construct the i^{th} row of compressed image, enter the run length value in the rebuild array from compressed array.

Step 6: Then, construct 1^{st} row, then next row, and so on to get the reconstructed image.

3.1.3 LZW Coding:

LZW Coding is a dictionary-based approach for identifying the number of repeating image pixels using a single index. It stores the character sequence that is dynamically chosen from a dictionary image. Each character sequence is assigned an index on a maximum of 4096 characters. It is also known as a greedy algorithm since each iteration modifies the dictionary for each new string or character sequence. The algorithmic phases of LZW Coding are listed below.

Algorithm for LZW

Step 1: The dictionary is constructed in such a way that all input strings have the same length.

Step 2: The longest string for the current input symbol is determined using a dictionary.

Step 3: This string includes the dictionary index, which is output and erased from input.

Step 4: Continue to the next symbol and repeat the procedure until all of the input symbols have been processed.

4. RESULT AND DISCUSSION

The results are displayed using a medical image from the MedPix® database. Specific outputs from each image is gathered and analyzed after compression and enhancement is performed on the sample medical images. MedPix® is an open-access online resource of restorative images, educational cases, clinical topics, coordinating images, and printed documentation that contains over 1000 images and 500 understanding case scenarios. It primarily targets a target audience of doctors and medical attendants, as well as other healthcare workers, medical understudies, nursing understudies, and those interested in therapeutic understanding.

The encoding algorithm verify the collection of medical images based on consistent results. The quality of the reconstructed image has been investigated with the CR and CF which perform as an analyzing metrics of compression efficiency.

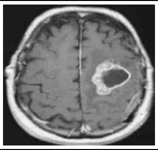
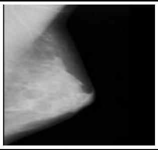
4.1 ANALYZING METRIC OF COMPRESSION EFFICIENCY

One of the main metrics for analyzing the efficiency of compression by CR and it can be expressed as given below formulae

$$CR = \text{(Total bits in compressed data)/(Total bits in uncompressed data)}$$

In order to accomplish high compression efficiency, the CR measured value need to be as minimum as possible and vice versa in the case of CF metrics because CR is inversely proportional to CF. Subsequently, the bite rate is represented in term of required amount of bit for executing each pixel present in the image. The Table.1 additionally categorizes the values extracted from these diverse medical images using the HC algorithm. The values take from this various medical image for LZW algorithm are also categorized in Table.2. The values take from this various medical image for RLE algorithm are also categorized in Table.3.

Table.1. Analysis of compression efficiency and computation time for HC algorithm

Image Name	Sample Input Medical Images	HC Algorithm			LZW Algorithm			RLE Algorithm		
		No of pixel in image	CR	CF	No of pixel in image	CR	CF	No of pixel in image	CR	CF
Brain MRI image		114840	114840	114840	114840	0.0952398	1.1958623	114840	0.07562398	1.1958623
Dental mandibular x-ray image		3705000	3705000	3705000	3705000	0.125379314	7.56893245	3705000	0.025379314	6.56893245
Dental mandibular cropped image		4012000	4012000	4012000	4012000	0.147895623	5.56893256	4012000	0.037895623	4.46893256
Mammogram image		1048576	1048576	1048576	1048576	0.56893256	6.23589641	1048576	0.46893256	5.33589641
Skin cancer image		270000	270000	270000	270000	0.058965896	1.89025896	270000	0.058965896	2.79025896
Hand bone x-ray image		588471	588471	588471	588471	0.5236984	1.98856745	588471	0.2236984	2.78856745

The Table.1 illustrates the comparative analysis of proposed coding scheme with encoding algorithm such as LZW, RLE and Huffman based on compression performance efficiency in terms of CR and CF. From the figure, it is apparently shown the average CR value of RLE attains the lower CR value which performs the better compression efficiency than the Huffman and LZW. Likewise, the average CR value of Huffman attains lower CR value which performs better compression efficiency than the LZW. The CF is the inverse of CR, where the RLE attains higher CF value than the LZW and Huffman.

4.2 EVALUATION METRICS

The values are calculated after reconstruction of the compressed image on host computer and then solving for the PSNR and MSE parameters. The PSNR is a well-known metric for evaluating the quality of a reconstructed image. The PSNR value is calculated using the formula as in Eq.(2).

$$PSNR=10\log_{10}(MAX^2/MSE) \quad (2)$$

where *MAX* denotes the image maximum intensity pixel value

The MSE is the difference of cumulative squared errors between the compressed and original image. The MSE value is calculated using the technique outlined below equation.2.

$$MSE = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}\{lm\ (x,y) - lm'(x,y)\}^2 \quad (3)$$

where, *M* and *N* are the dimensions of a reconstruction matrix, respectively. $lm(x,y)$ and $lm'(x,y)$ represent original and compressed image pixels, respectively. The goal of compression methods is to reduce the MSE.

The mean value of mean square error of six sample medical images is shown in Table.2. The real-time medical image dataset is put to the test, and the PSNR values that is shown performed better. RLE has a higher PSNR average value than HC and LZW.

Table.2. MSE and PSNR Value of medical images

Medical image	MSE			PSNR		
	HC	LZW	RLE	HC	LZW	RLE
Brain MRI	0.87	0.91	0.74	44.23	42.89	46.12
Dental mandibular x-ray	1.51	1.54	1.21	43.51	42.56	43.98
Dental mandibular cropped	1.74	1.54	1.35	43.87	43.96	45.78
Mammogram	3.14	2.78	2.15	42.13	43.14	45.67
Skin cancer	4.25	3.98	2.87	43.87	40.69	48.65
Hand bone x-ray	4.89	4.05	3.21	40.21	41.52	45.69

The Table.2 depicts MSE performance evaluation for various medical image compressions. From 0 to 4, it reflects the MSE value of HC, LZW, and RLE. PSNR values vary from 41 (dB) to 48 (dB).

5. CONCLUSION

The encoding process both minimizes the size and maintains the quality of the medical image. Experiments is carried out on a variety of medical images, with good reconstruction quality. Image compression methods reduce the number of bits required for image representation while maintaining output quality. As a consequence, lossless compression methods is investigated in this study, and it can be concluded that the RLE approach is better suited for lossless compression based on the test image quality. This RLE algorithm will result in reduced MSE, increased PSNR for various test images, so it can be used for transmission and reception of images. The CR value of RLE attains the lower value which performs the better compression efficiency than the Huffman and LZW whereas RLE attains higher CF value than the LZW and Huffman. As a result, the RLE encoding technique may be used to provide lossless compression for high-resolution images without any delays or complications.

REFERENCES

[1] N. Karimi, R. Ranjbarzadeh Kondrood and T. Alizadeh, "An Intelligent System for Quality Measurement of Golden Bleached Raisins using Two Comparative Machine Learning Algorithms", *Journal of the International Measurement Confederation*, Vol. 107, pp. 68-76, 2017.

[2] A.C.B. Monteiro, R.P. França and P.D.M. Negrete, "Metaheuristics Applied to Blood Image Analysis", *Lecture Notes in Electrical Engineering*, Vol. 696, pp. 117-135, 2021.

[3] Z. Li, X. Zhang and S. Zhang, "Large-Scale Retrieval for Medical Image Analytics: A Comprehensive Review", *Medical Image Analysis*, Vol. 43, pp. 66-84, 2018.

[4] R. Ranjbarzadeh and S. Baseri Saadi, "Automated Liver and Tumor Segmentation based on Concave and Convex Points using Fuzzy C-Means and Mean Shift Clustering", *Journal of the International Measurement Confederation*, Vol. 151, pp. 1-18, 2020.

[5] W.A. Ali, K.N. Manasa and P. Sandhya, "A Review of Current Machine Learning Approaches for Anomaly Detection in Network Traffic", *Journal of Telecommunications and the Digital Economy*, Vol. 8, No. 1, pp. 64-95, 2020.

[6] M.L. Hachemi, M. Omari and M. Baroudi, "Enhancement of DCT-Based Image Compression Using Trigonometric Functions", *Proceedings of International Conference on Computing Sciences and Engineering*, pp. 1-5, 2018.

[7] N.V. Kousik and M. Saravanan, "A Review of Various Reversible Embedding Mechanisms", *International Journal of Intelligence and Sustainable Computing*, Vol. 1, No. 3, pp. 233-266, 2021.

[8] Med Karim Abdmouleh, Atef Masmoudi and Med Salim Bouhlel. "A New Method which Combines Arithmetic Coding with RLE for Lossless Image Compression", *Scientific Research Publishing*, Vol. 2021, pp. 1-7, 2021.

[9] J. Surendiran, S. Theetchenya and P.M. Benson Mansingh, "Segmentation of Optic Disc and Cup Using Modified Recurrent Neural Network", *BioMed Research International*, Vol. 2022, pp.1-8, 2022.

[10] S.S. Sivasankari, J. Surendiran and M. Ramkumar, "Classification of Diabetes using Multilayer Perceptron", *Proceedings of IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics*, pp. 1-5, 2022.

[11] Gajendra Sharma, "Analysis of Huffman Coding and Lempel–Ziv–Welch (LZW) Coding as Data Compression Techniques", *International Journal of Scientific Research in Computer Science and Engineering*, Vol. 8, No. 1, pp. 37-44, 2020.

[12] Yaghoub Pourasad and Fausto Cavallaro. "A Novel Image Processing Approach to Enhancement and Compression of X-ray Images", *International Journal of Environmental Research and Public Health*, Vol. 18, pp. 1-12, 2021.

[13] R. Ahmad and N.S. Choubey, "Review on Image Enhancement Techniques using Biologically Inspired Artificial Bee Colony Algorithms and its Variants", *Lecture Notes in Computational Vision and Biomechanics*, pp. 249-271, 2018.

[14] G. Kaur, N. Bhardwaj and P.K. Singh, "An Analytic Review on Image Enhancement Techniques based on Soft Computing Approach", *Proceedings of International Conference on Advances in Intelligent Systems and Computing*, pp. 255-265, 2018.

[15] C.R. Nithyananda and A.C. Ramachandra, "Review on Histogram Equalization based Image Enhancement Techniques", *Proceedings of International Conference on Electrical, Electronics, and Optimization Techniques*, pp. 1-6, 2016.

[16] R. Maini and H.A. Aggarwal, "Comprehensive Review of Image Enhancement Techniques", *International Journal of Innovative Research and Growth*, Vol. 8. pp. 1-9, 2010.

[17] C.S.G. Dhas and T.D. Geleto, "D-PPSOK Clustering Algorithm with Data Sampling for Clustering Big Data Analysis", Academic Press, 2022.

[18] N. Karthikeyan and S.R. Mugunthan, "Comparative Study of Lossy and Lossless Image Compression Techniques", *International Journal of Engineering and Technology*, Vol. 7, pp. 950-953, 2018.

[19] Boopathiraja Subramanian, Kalavathi Palanisamy and V.B. Surya Prasath, "On a Hybrid Lossless Compression Technique for Three-Dimensional Medical Images", *Medical Imaging*, Vol. 2020, pp. 1-12, 2020.

[20] R. Sowmyalakshmi, Mohamed IbrahimWaly, T. Mohamed Yacin Sikkandar, Jayasankar Sayed Sayeed Ahmad, Rashmi Rani and Suresh Chavhan, "An Optimal Lempel Ziv Markov Based Microarray Image Compression Algorithm", *Computers, Materials and Continua*, Vol. 69, No. 2, pp. 1-14, 2021.

[21] M.S. Divya, J. Chandrashekhara, S. Vinay and A. Ramadevi, "Lossless Compression for Text Document Using Huffman Encoding, Run Length Encoding, and Lempel-Ziv Welch Coding Algorithms", *International Journal of Computer Science and Mobile Computing*, Vol. 9, No. 3, pp. 100-105, 2020.

[22] J. Karam Lina, "*Lossless Image Compression*", Academic Press, 2009.
[23] Bin Xiao, Gang Lu, Yanhong Zhang, Weisheng Li and Guoyin Wang, "Lossless Image Compression based on Integer Discrete Tchebichef Transform", *Neurocomputing*, Vol. 214, pp. 587–593, 2016.
[24] Med Karim Abdmouleh, Atef Masmoudi and Med Salim Bouhlel, "A New Method which Combines Arithmetic Coding with RLE for Lossless Image Compression", *Scientific Research Publishing*, Vol. 2021, pp. 1-12, 2021.
[25] S.W. Chiang and L.M. Po, "Adaptive Lossy LZW Algorithm for Palletized Image Compression", *Electronics Letters*, Vol. 33, No. 10, pp. 852-854, 1997.
[26] Paul G. Howard and Jeffrey Scott Vitter, "Parallel Lossless Image Compression using Huffman and Arithmetic Coding", *Proceedings of International Conference on Data Compression*, pp. 299-308, 1992.
[27] Yaghoub Pourasad and Fausto Cavallaro, "A Novel Image Processing Approach to Enhancement and Compression of X-ray Images", *International Journal of Environmental Research and Public Health*, Vol. 18, pp. 6724-6744, 2021.