The Combined Fuzzy C-means, Data Encryption and Reversible Data Hiding Method to Enhance the Integrity and Authenticity of the Electronics Health Records (EHR) for Tele Radiology

Fepslin Athish Mon (fepslin@gmail.com)
Kalasalingam University: Kalasalingam Academy of Research and Education https://orcid.org/0000-0002-1023-0593

Suthendran K
Kalasalingam University: Kalasalingam Academy of Research and Education

Jarin T
Jyothi Engineering College

Research Article

Keywords: Medical Images, Data Hiding, EMR, Digital Signature, Authentication Tag.

Posted Date: June 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-647542/v1

Abstract

The transferring of medical image to the medical practitioner through cloud network with field expects around the globe to get better diagnostics and expect suggestions ever-increasing significantly. Enormous possibility of security treats arises while transferring the medical image through cloud services. In this paper, the trustworthiness, information integrity and authenticity of the patient EMR during storage and transmission over network can be improved without compromising medical quality and diagnosis accuracy. This can be achieved with the help of combined segmentation, cryptography and reversible data hiding on medical images. In this work, we store the patient basic details in the medical image using reversible data hiding with data capacity and security enhancements. The proposed method uses medical images of EMR as cover image, the patient's details can be encoded in the cover image as data and finally encrypt the image using asymmetric key encryption. The proposed method provides more improvement in integrity, security, authenticity, data capacity and error rate. This method also maintains PSNR value above 51dB.

Introduction

The vital need of a human being is healthy life. Teleradiology [1] is the evolving technology for the enhancement of medical system. With this technology, the medical practitioners can transfer the EMR though networks with various physicians all over the world. Nowadays, medical diagnosis of diseases has been improved due to the growth in the information technology. All patients' information is stored and transmitted as electronics medical records or electronic health records (EMR/EHR)

The enormous patient's data are stored as electronic medical records with the help of evolving information technologies. For the better treatment and physicians' suggestions, EMR will be shared with the medical practitioners in specific diagnosis area. The shared EMR can be accessed by physicians from any location of world through network at any time. However the EMR shared over network creates some sensitive issues like patient data security and privacy [9]. Standardisation, maintaining medical ethics, patient information privacy and data security are the major concerns in the digital information management [10]. Security issues may leads to unauthorised access of patient information which causes major thread in diagnosis, treatment and research.

The security has a major role in maintaining confidentiality, providing authenticity, integrity and accountability. These issues can be addressed by methods like data hiding and encryption. The patient's information is encrypted and encoded into medical image with reversible data hiding methodology. With this reversible data hiding methodology, the original medical image and patient information can be retrieved without any loss [2].

A. Reversible Data Hiding

In the case of a medical image, a small loss of data may affect the diagnosis process and treatment procedures. So, the recovery of the original medical image is very important.

Figure 1 shows the data hiding process in an image. The image is pre-processed, and the data can be inserted into the medical image using data hiding algorithm. The embedded image will contain data with cover image.

Figure 2 shows the mining of data from the embedded image and recovery of previous cover image with no data loss. The data extraction algorithm and image recovery algorithm play a vital role in image recovery

B. Electronic Medical Records (EMR) Security

The integrity of EMR is the most important security concept in teleradiology. The use of information technology in medical field leads to issues in related to integrity of medical images [18]. The term that more related with integrity is authenticity. The identification of ownership of EMR and maintaining its originality is an important task [19]. The integrity and authenticity of EMR can be offered by using digital signature.

Digital signature is the process of generating digitally signed information which cannot be forged without the secret key used. The asymmetric key encryption is used with two keys namely private key and public key. The private key is used only by the signer and the public key can be used by anyone to check the validity of the signature. The private key cannot be derived from the public key, but both are interrelated.

Figure 3 explains the digital signature generation process. The data can be hashed and with a private key asymmetrically encoded to produce a digital signature. The private key provides the authentication information of the patient.

The verification of the digital signature (figure 4) uses the public key to create the hash value from the digital signature and it will be matched with the hash value generated from original data. If both the hashes are the same then it will be considered as authentic data and secured.

Reversible Data Hiding, Image Integrity And Authenticity Approaches

In recent days, the research on the security of electronic medical records has been addressed. Metadata and reversible data hiding are the two major approaches used for providing security for EMR. By considering digital signature as metadata, it can be kept alongside with EMR. In the case of DICOM standard medical images, the header portion is used to store the digital signature [20]. This methodology has been executed and tested [21]. The DICOM image header can be encrypted using the metadata approach for providing the confidentiality of DICOM medical images [22].

The information can be inserted in a cover image by the technique of data hiding. By using reversible data hiding methodology, the cover image and the information inserted can be retrieved back without any loss. Yun[3] proposed an LSB method for reversible data hiding in which the least significant bit is replaced with data. The image will not be affected in visual perspective, regeneration of the original image may have a minute variations. Wen [4] proposed a method in which a separate record is

maintained for recording the changes in pixel position and values. The different image histogram method and transform coefficient histogram methods are proposed by Ho[5]. Yongjian[6] suggested a method named even and odd number based embedding. All these approaches show better data capacity and image recovery.

Boundaries in current approaches

In the current methodologies, the security data and medical images are not linked together. This may make the medical images non-trustable and weaken the security information. The initial data will be retrieved back by applying reversible data hiding approaches but the current approaches may degrade the image quality, not enough data capacity to hide large data and also shows variation in file size.

Proposed Methodology

The proposed approach provides a new methodology to maintain the integrity and authenticity of EMR by combining asymmetric key encryption and reversible data hiding. The figure 5 shows the proposed technique of data embedding process. The original EMR is encrypted and encrypted version is shared in teleradiology. The proposed methodology will maintain the original file size along with improvement in data embedding capacity, integrity, and authenticity.

The patient information is encrypted with an authentication key (private key) and the encrypted data is generated along with the patient key. The encrypted data are inserted using reversible data hiding technology in a medical image. The output of this process is the data embedded medical image and the patient key for provenance and authenticity of the medical image.

The encryption key is derived from the authenticity and integrity of patient data. The medical practitioner or any other third party will not retrieve a patient's data without the patient security key. This provides a strong link between the EMR and the patient's data. The third party can get only the encrypted data of the patient's details if the data in the image is removed for the diagnosis purpose of medical images. The following figure 6 explain the proposed process of data mining and image retrieval.

The EMR verification deals with the patient's data decryption and authentication. The secret key can be retrieved from the patient key. The data embedded on the medical image can be extracted with reversible data hiding technique. The data extracted from the image will be in encrypted form and the data can be decrypted by the secret key generated from patient key. The authentication tag and patient's data can retrieve from the decrypted data. The integrity and authenticity of EMR can be verified with the authentication tag.

Algorithm

In this proposed method, three methods are joint together to enhance integrity, security, authenticity, data capacity and reduced error rate of EMR. The region of interest (ROI) is identified by using Fuzzy C-Mean algorithm. The improved XOR based reversible data hiding mechanism is employed to enhance the data

hiding capacity while embedding the data. The asymmetric key encryption method is used to provide integrity and authenticity. In this algorithm, the key for encryption and initialization vector are derived from patients basic information. The encrypted patient data is inserted on the medical image by applying reversible data hiding, so that medical image and encrypted data will be obtained when required.

Pre-processing and location mapping:

The image has to be pre-treated to keep the image boundary value within the range. In the grey scale image, the pixel value can be ranges from 0 to 255. The embedding single bit of data on this image may leads to undefined value (-1 or 256)

For this, every pixel value lies in the upper boundary (255) must be altered to 254 and the lower boundary pixel value (0) must be altered to 1. Now the embedding of one bit data on the pixel will not create indeterminate values in boundary values

To increase the image data embedding capacity, the pear pair pixel frequency have to be increased. This can be performed by using XOR operation on the pixel values which can be reversible. The XOR operation with value 255 is performed over the pixel value which is less than 128 (middle grey value) to generate new pixel value (Px) and remaining pixel values (greater than 128) will be remain same.

Phase 1: Region of non-interest (RONI)

The medical image can be segmented two sections namely: Diagnostic sections as region of interest (ROI), and non-diagnosis section as region of non-interest (RONI).

RONI Extraction: The RONI section of the medical image can be extract using several methods by following various morphological operations. In this we proposed a Fuzzy C – Means algorithm and Canny edge detection algorithm to get the ROI from the medical image. The figure 7 explains the process of extraction of region of non-interest region.

The fuzzy C-Mean algorithm is applied on the grey scale medical image to transform the input image into binary image. The canny edge detection algorithm is used for automating the calculation of limit value by using an anticipated quantity of edge and light image pixels. The morphological operations: dilation and erosion operations is applied to obtain the output image. The binary mask is generated by obtaining the white section of image called ROI. By subtracting the ROI from the original grey scale medical image, we can obtain the region of non-interest.

Phase 2: Signing and Encoding Process

The encoding process produces the stream of encoded patient information and authentication tag, retains the information about patient information integrity. The authentication tag will be applied along with the private key of owner to create digital signature.

The patient data encryption and encrypted data embedding (Figure 8) of the suggested process is as follows. First the patient data (or some part of data) is hashed, which creates fixed size bits for indicating its integrity. Single fixed part of the output is used as key and another section as the initialization vector of the encoding process. The key and the initialization vector together known as security data.

The authentication tag can be generated after encryption process, which is an evidence for integrity of patient information. Then the tag can be endorsed with the private key of the patient, the digital signature is generated and stored along with public key as patient identity which can be carry on by patient as safety code.

Phase 3: Reversible Data-Hiding

For data embedding, two peak pixel values (P_1, P_2) are produced from histogram of the medical image, peak value P_2 should be greater than P_1. The encoded data bits (b) can be inserted into medical image.

Data Embedding

The encoded patient information can be inserted in the pre-treated medical image as follow:

a. Subtract pixel value by 1 (ie Px- 1) for all pixel (Px) value less than P_1; (ie $Px < P_1$)
b. Subtract pixel value with encoded data bit(b) (ie $P_1 - b(k)$) for all pixel value equal to P_1; (ie $Px = P_1$)
c. No change in pixel value between P_1 and P_2.
d. Add pixel value by 1 for all pixel (Px) value greater than P_2; (ie $Px > P_2$)
e. Add pixel value with encoded data bit(b) (ie $P_2 + b(k)$) for all pixel value equal to P_2; (ie $Px = P_2$)

The pixel values P_1 and P_2 can be inserted in the LSB of last 16 pixel's position. If all the data is not embedded then repeat the above process till data embedding complete.

Data Extraction

The data extraction process is the reverse process of data embedding. First the peak values can be extracted from the LSB of last 16 pixel. The data extraction can be performed as follows:

a. Scan the pixel value of embedded medical image sequentially.
b. if the pixel value is equivalent to $P_1 - 1$ or $P_2 + 1$ then encoded data b(k) is 1.
c. If pixel value is equivalent to P_1 or P_2 then encoded data b(k) is 0.

Thus the encrypted data (b) can be extracted from the embedded medical cover image

Medical Image Recovery

The image can be retrieved completely as same as previous medical image with no loss. The recovery of original medical images as follows:

a. Add pixel value by 1 (ie Px + 1) for all pixel value less than P_1 -1; (ie Px < P_1 - 1)
b. Keep value as P_1 for all pixel with P_1 - 1 or P_1
c. Keep value as P_2 for all pixel with P_2 + 1 or P_2
d. Subtract pixel value by 1 (ie Px - 1) for all pixel value greater than P_2 +1; (ie Px > P_2 +1)

Decryption and Verification Process

The figure 9 show the patient information decryption and verification of integrity and authenticity. Same patient information used for the encryption has been recovered and hashed to produce the key and the initialization vector in the similar process used for encoding.

The original patient data can be decrypted from the encoded data which is separated from the medical image. And its authentication tag is also separated and validated compared to the signature stored in patient identity. Thus, we can verify the integrity and authenticity of patients information.

Implementation and Evaluation

The above algorithm is implemented in MATLAB simulation tool and evaluated its performance in various terms. First, the Whirlpool algorithm is used to hash the information. This is an 512 bit output, strong hash function which is recommended as a project namely New European Schemes for Signature, Integrity and Encryption (NESSIE)[23]. It is standardised by International Standards Organization ISO [24]. For this operation, patient ID and patient name is carried for hashing and other data are taken for encoding. The patients name is hashed as encryption key for encrypting patients information

The AES algorithm with 256 bits key size and 96 bits initialization vector is used for data encryption. This algorithm provides universal hashing for authenticated encryption, generation of cipher text and authorization tags for integrity verification. The National Institute of Standards and Technology (NIST) standardized this approach[25].

The 256 bit key Elliptic Curve Digital Signature Algorithm (ECDSA) is used as authentication tag. ECDSA is generally recognized and utilized for creating short signature than the initial DSA with the similar security.

The encoding will be result in change of patient information into a cipher text which cannot be read by anyone if the key is tempered. The data after decryption will recover the original patient information. The third party who obtained the data from the embedded medical image only able to get an cipher text. And only with the patient identity, physician can decrypt patients information from the cipher text.

Table 1: Correlation between original record and encrypted record

Data Set	Correlation
Patient record 1	[corr] < 0.001
Patient record 2	[corr] < 0.001
Patient record 3	[corr] < 0.003
Patient record 4	[corr] < 0.001
Patient record 5	[corr] < 0.002
Patient record 6	[corr] < 0.001

The cipher data after encryption will not have any relation in appearance with the original data. Correlation are calculated for various patients records (Table 1). Low correlation value indicates that there is low similarity of patient record after encoding.

In the reversible data hiding part in this proposed method, the data inserting capacity has been enhanced by using extended XOR method. It shows a sensible increase in the size of data embedded capability without any visible change in medical images and size of medical images. The extended XOR method of duplicating the peak pixel value focussed more in data capacity and PSNR value.

Sets of 8 EMR is used to test the proposed algorithm. The experimental result show that the data capacity has been enhanced to bits per pixel and it fixed for the entire data set and PSNR value after inserting data bits is above 51dB

In this recommended method, the results shows that the duplication of peak pair values works well in expanding the data inserting capability, PSNR and the maintain the original file size.

Table 2: Assessment on Payload and PSNR value

EMR	PSNR (dB)	Pure Payload (bits)
DICOM 1	51.1365	262128
DICOM 2	50.9592	262128
DICOM 3	51.9463	262128
DICOM 4	51.1256	262128
DICOM 5	51.0022	262128
DICOM 6	50.8557	262128
DICOM 7	51.6813	262128
DICOM 8	51.0051	262128

From the above table 2, it is identified that the data capacity reaches it maximum capacity. That is, data capacity is total pixel size minus reserved 16 pixels for peak value embedding. The Figure 10 shows the comparison of PSNR for various DICOM images.

For evaluation of file size, the image with 512 X 512 is analysed with the same algorithm. The file size should be maintaining similar before and after embedding of data. If the size of the file get change then it will show that there is some data hidden inside the image. Normally file size can be calculated by size required to save a pixel information multiplied by number of pixels. While testing with various image with various resolution, the file size remains constant value as of original image, after embedding and after data recovery.

In the case of reversibility, many of the data hiding method shows some distortion in the due to data embedding and cannot retrieved the original image without any loss. In our approach, we are dealing with reversible data hiding approaches that will have no data loss or insignificant deficiency of data after recovery.

Table 3: Reversibility Analysis table

Test Dataset	PSNR(dB)		
	Initial dataset PSNR (dB)	After Data inserting PSNR (dB)	After data Removal PSNR (dB)
DICOM 1	99.32	51.1365	99.37
DICOM 2	99.43	50.9592	99.39
DICOM 3	99.63	51.9463	99.70
DICOM 4	99.68	51.1256	99.63
DICOM 5	99.72	51.0022	99.68
DICOM 6	99.59	50.8557	99.63
DICOM 7	99.92	51.6813	99.87
DICOM 8	99.48	51.0051	99.42

In the above figure 11, the PSNR value is retained to highest value after the extraction of data. This figure 11 shows that the PSNR value of medical image after removal of data has a insignificant difference from the previous image.

Conclusion

This paper proposed a new technique to provide integrity and authenticity of medical image along with reversible data hiding. Compare with the previous methods, this approach provides a trustworthiness of EMR without compromising their quality and file size. This method also provide hiding of patient's information in the medical image of same EMR.

The new algorithm proposed for improve authenticity and integrity of medical image and patient information was implemented, evaluated in various aspects and result is analysed. From the result, it confirmed that the proposed algorithm has a better result in various aspects comparing with many existing approaches.

The goal is to improve the security of electronic medical record in teleradiology, which is the emerging area in modern health care system. This paper is the part of providing integrity and authenticity of patient information along with reversible data hiding methodology.

Declaration Statements

Funding:

No funding sources available for this research work.

Conflicts of interest/Competing interests:

No conflict of interest

Availability of data and material:

All data generated or analysed during this study are included in this article.

Code availability:

Own code written and implemented in MATLAB tool. The Whirlpool algorithm is used to hash the information, AES algorithm with 256 bits key size and 96 bits initialization vector is used for data encryption. 256 bit key Elliptic Curve Digital Signature Algorithm (ECDSA) is used as authentication tag.

References

[1] P. Viswanathan and P. V. Krishna, "A Joint FED Watermarking System Using Spatial Fusion for Verifying the Security Issues of Teleradiology," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 3, pp. 753-764, May 2014. doi: 10.1109/JBHI.2013.2281322

[2] Arjun K P et. al., "Implementation of reversible Data Hiding in Encrypted Image using A-S Algorithm", IEEE International Conference on Green Computing and Internet of Things (ICGCIoT), Oct. 2015.

[3] Yun Q. Shi, “Reversible Data Hiding”, Lecture Notes in Computer Science, Springer, Vol. 3304, Jan. 2005.

[4] Wen Chung Kuo et. al., “Reversible Data Hiding Based on Histogram”, Lecture Notes in Computer Science, Springer, Vol. 4682, Aug. 2007

[5] Ho Thi Huong Thom et. al., “Steganalysis For Reversible Data Hiding”, Communication in Computer and information Science, Springer, Vol. 64, Mar. 2009.

[6]. Yongjian Hu et. al., “Reversible Data Hiding Using Prediction Error Values Embedding”, Lecture Notes in Computer Science, Springer, Vol. 5041, Aug. 2008.

[7] Qiming Li et. al., “A Reversible Data Hiding scheme for JPEG Images”, Lecture Notes in Computer Science, Springer, Vol. 6297, Dec. 2010.

[8]. United Nations Department of Public Information, Universal Declaration of Human Rights. (1948). [Online]. Available: http://www.unhchr.ch/udhr/ lang/eng.htm

[9]. “Oath of Hippocrates,” in Harvard Classics, vol. 38. Boston, MA: Collier, 1910.

[10]. W. Raghupathi and J. Tan, “Strategic IT applications in health care,” Commun. ACM, vol. 45, no. 12, pp. 56–61, 2002.

[11]. Hsiang Cheh Huang et. al., “Reversible Data Hiding with Hierarchical Relationships”, Lecture Notes in Computer Science, Springer, Vol. 7198, Sep. 2010.

[12]. Z. Ni et. al., “Reversible Data Hiding”, IEEE Transaction on Circuits and Systems For Video Technology, Vol. 1, Mar. 2006.

[13]. Wei Liang Tai et. al., “Reversible Data Hiding Based on Histogram Modification of Pixel Differences”, IEEE Transaction on Circuits and Systems for Video Technology, Vol. 19, Jun. 2009.

[14]. X. Zhang, “Reversible Data Hiding in Encrypted Images”, Signal Processing Letters, Springer, Vol. 18, Apr. 2011.

[15]. X. Zhang, “Separable Reversible Data Hiding in Encrypted Image”, IEEE Transactions on Information Forensics and Security, Vol. 7, Apr. 2012.

[16]. W. Hong et. al., “An Improved Reversible Data Hiding in Encrypted Images Using Side Match”, IEEE Signal Processing Letters, Vol. 19, Apr. 2012.

[17]. Kede Ma et. al.,”Reversible Data Hiding in Encrypted Images by Reserving Room Before Encryption”, IEEE Transactions on Information Forensics and Security, Vol. 8, Mar. 2013.

[18] D. L. Hamilton, “Identification and evaluation of the security requirements in medical applications,” in Proc. 5th Annu. IEEE Symp. Comput.-Based Med. Syst., Session 2B: Pictural Archival Commun. Syst., 1992, pp. 129– 137

[19] A. L. Rector, W. A. Nolan, and S. Kay, “Foundations for an electronic medical record,” Methods Inf. Med., vol. 30, pp. 179–186, 1991

[20] Digital Imaging and Communications in Medicine (DICOM) Standard, DICOM. (2006). [Online]. Available: http://medical.nema.org/dicom/ 2006/

[21] M. Kroll, B. Schutze, T. Geisbe, H.-G. Lipinski, D. H. W. Gr ¨ onemeyer, ¨ and T. J. Filler, “Embedded systems for signing medical images using the DICOM standard,” in Proc. Int. Congr. Series 2003, vol. 1256, pp. 849– 854.

[22] J. Bernarding, A. Thiel, and A. Grzesik, “A JAVA-based DICOM server with integration of clinical findings and DICOM-conform data encryption,” Int. J. Med. Inf., vol. 64, pp. 429–438, 2001.

[23] P. S. L. M. Barreto and V. Rijmen. (2003). The WHIRLPOOL Hashing Function [Online]. Available: http://planeta.terra.com.br/informatica/ paulobarreto/whirlpool.zip.

]24] ISO. (2004). ISO/IEC 10118-3:2004 Standard [Online]. Available: http:// www.iso.org/iso/en/ CatalogueDetailPage. CatalogueDetail?CSNUMBER =39876&scopelist=

[25] D. A. McGrew and J. Viega, “The Galois/counter mode of operation (GCM),” NIST Comput. Security Div., Comput. Security Resour. Center, Gaithersburg, MD, Tech. Rep., 2005.

Figures

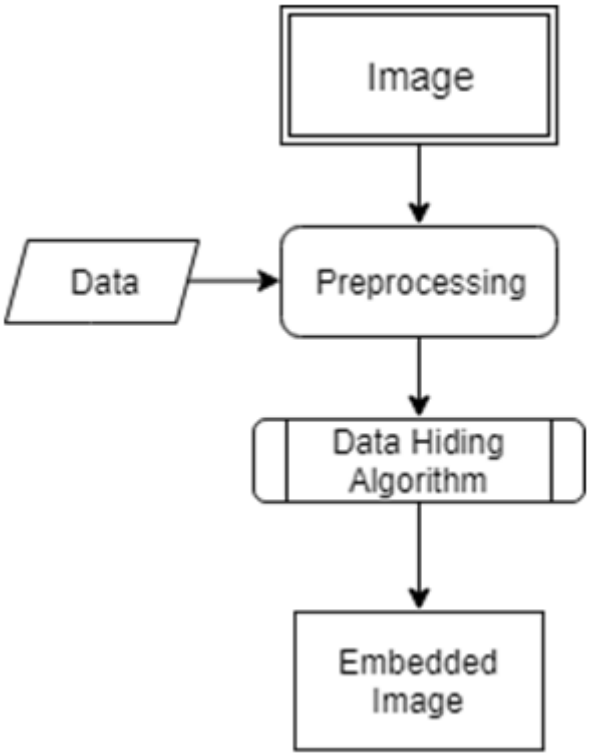

Figure 1

Process of data hiding in medical Image

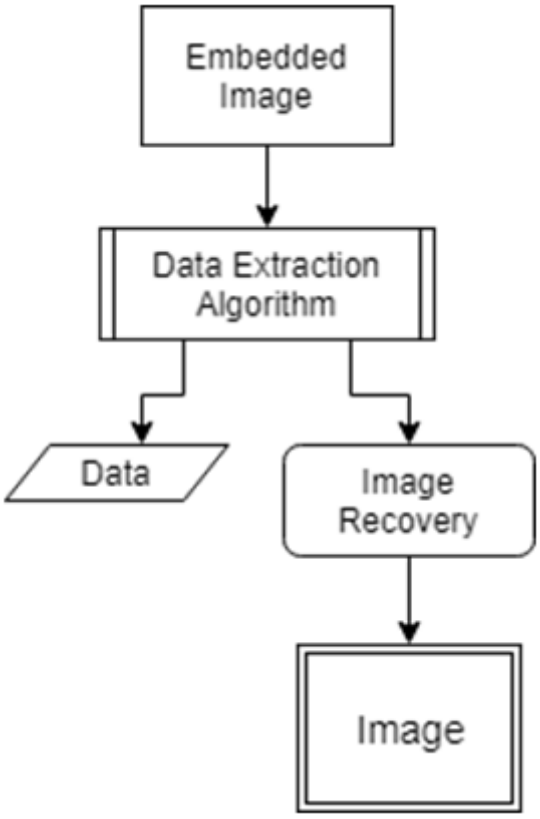

Figure 2

Data Extraction processof Embedded Image

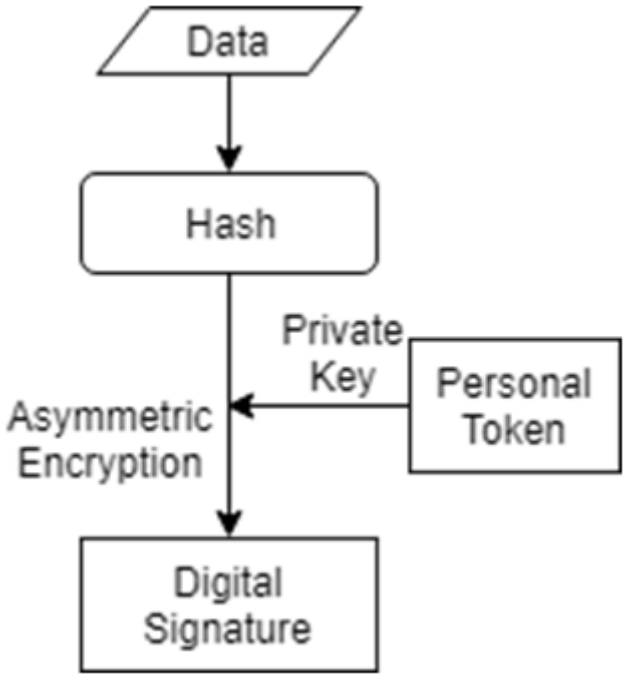

Figure 3

Digital Signature generation process

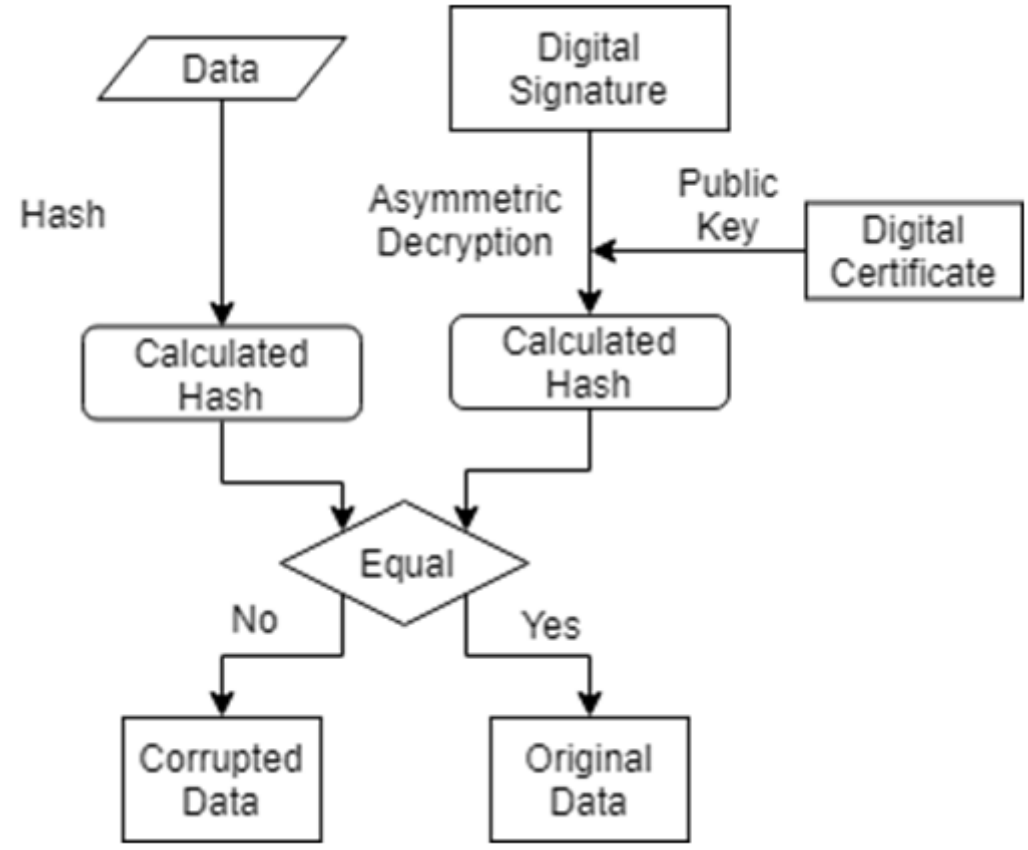

Figure 4

Verification of Digital Signature

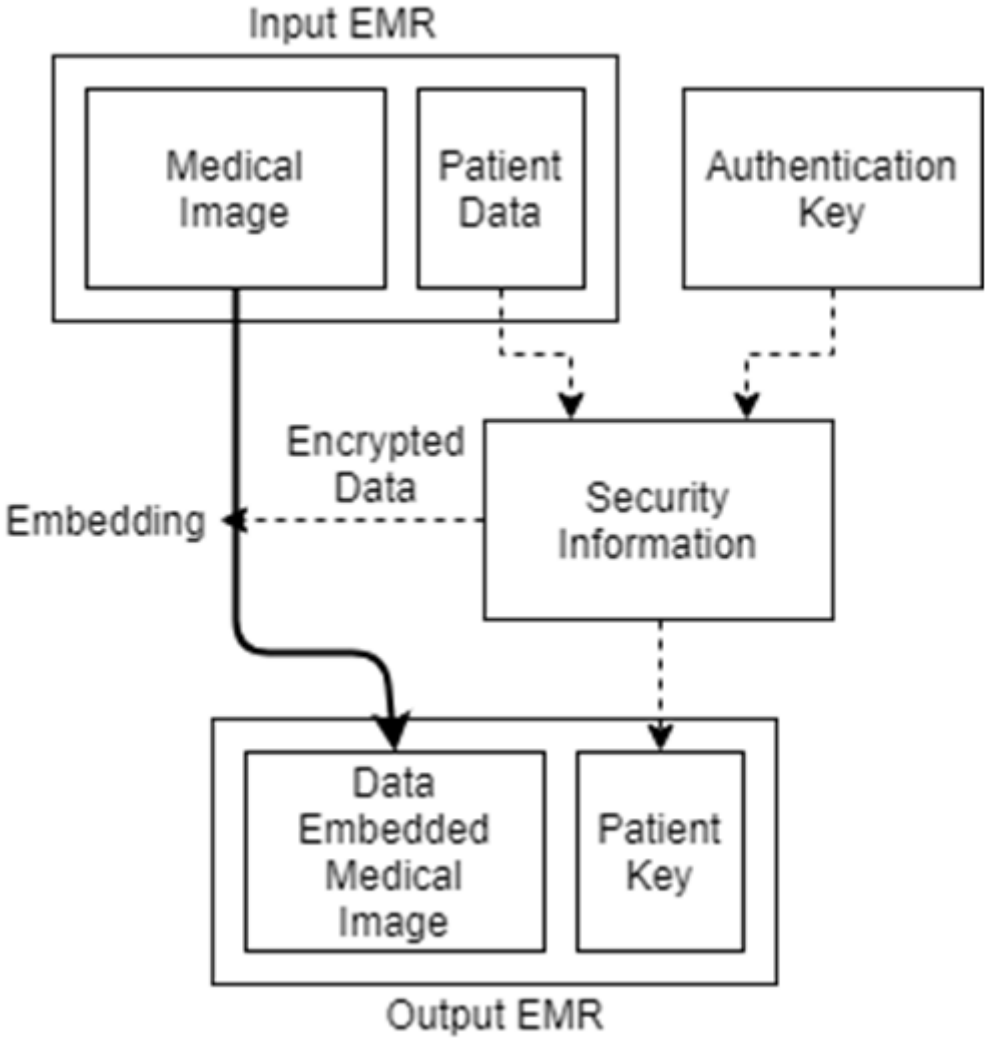

Figure 5

Proposed Technique (Data Embedding)

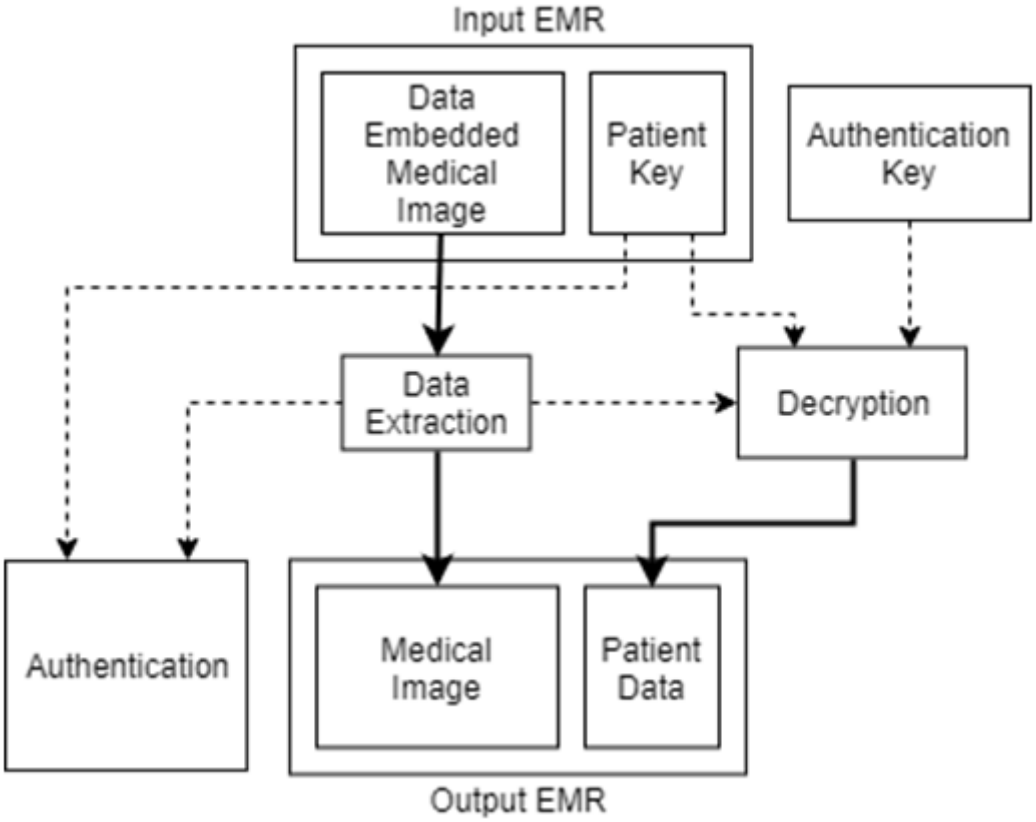

Figure 6

Proposed Data Mining and Image retrieval

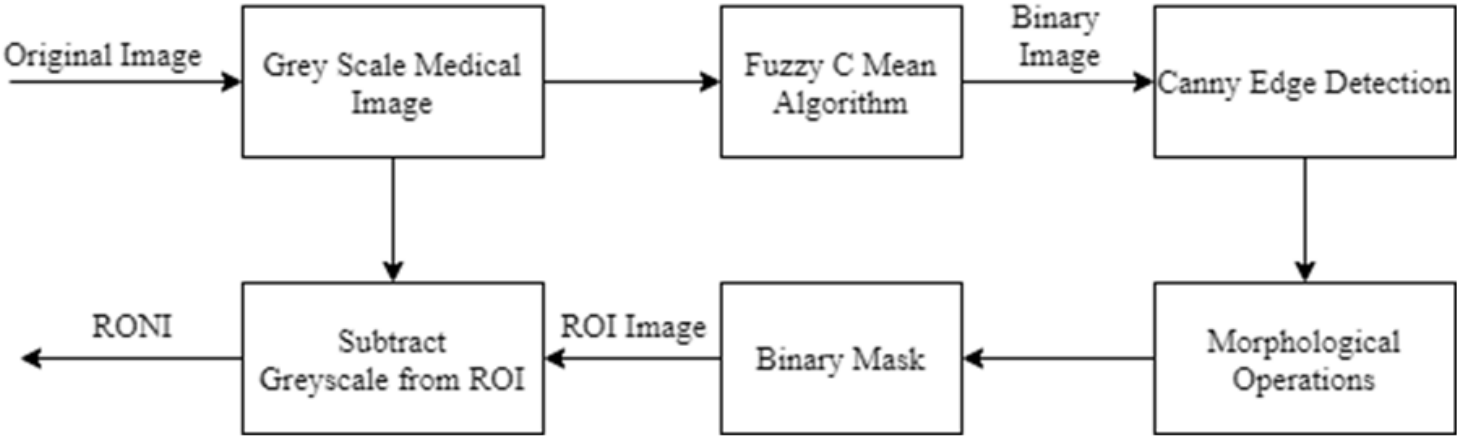

Figure 7

Region of Non Interest

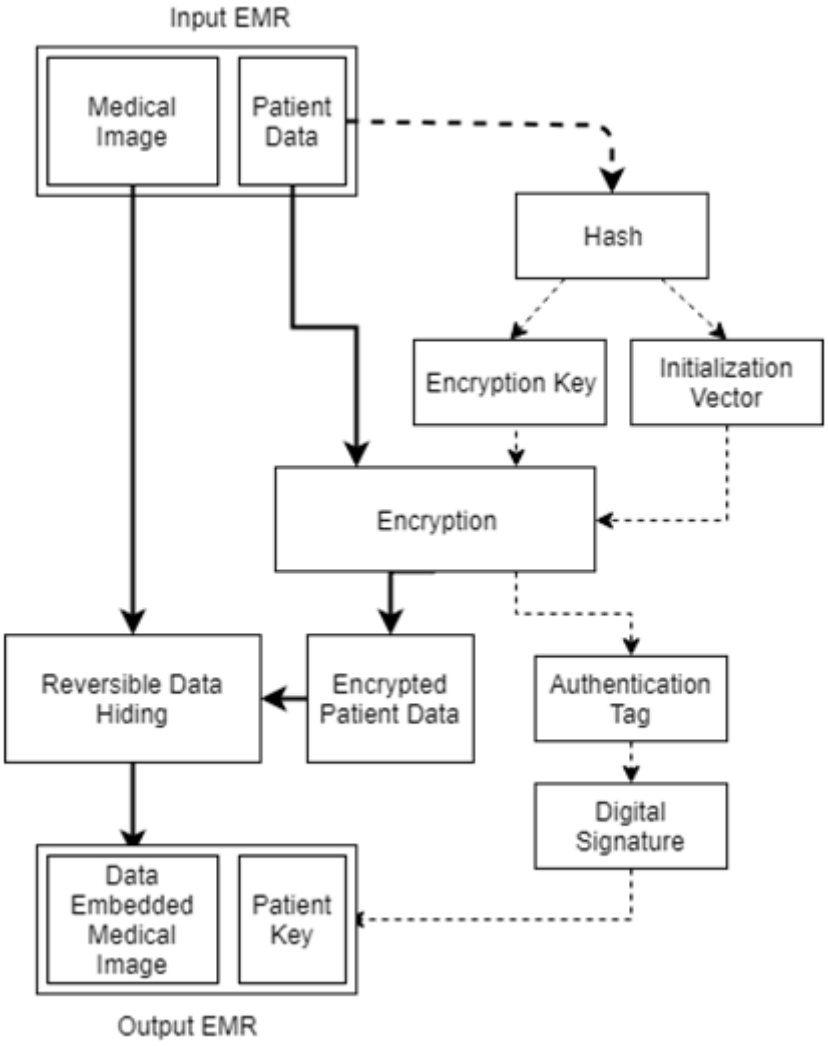

Figure 8

Encryption and Data Embedding

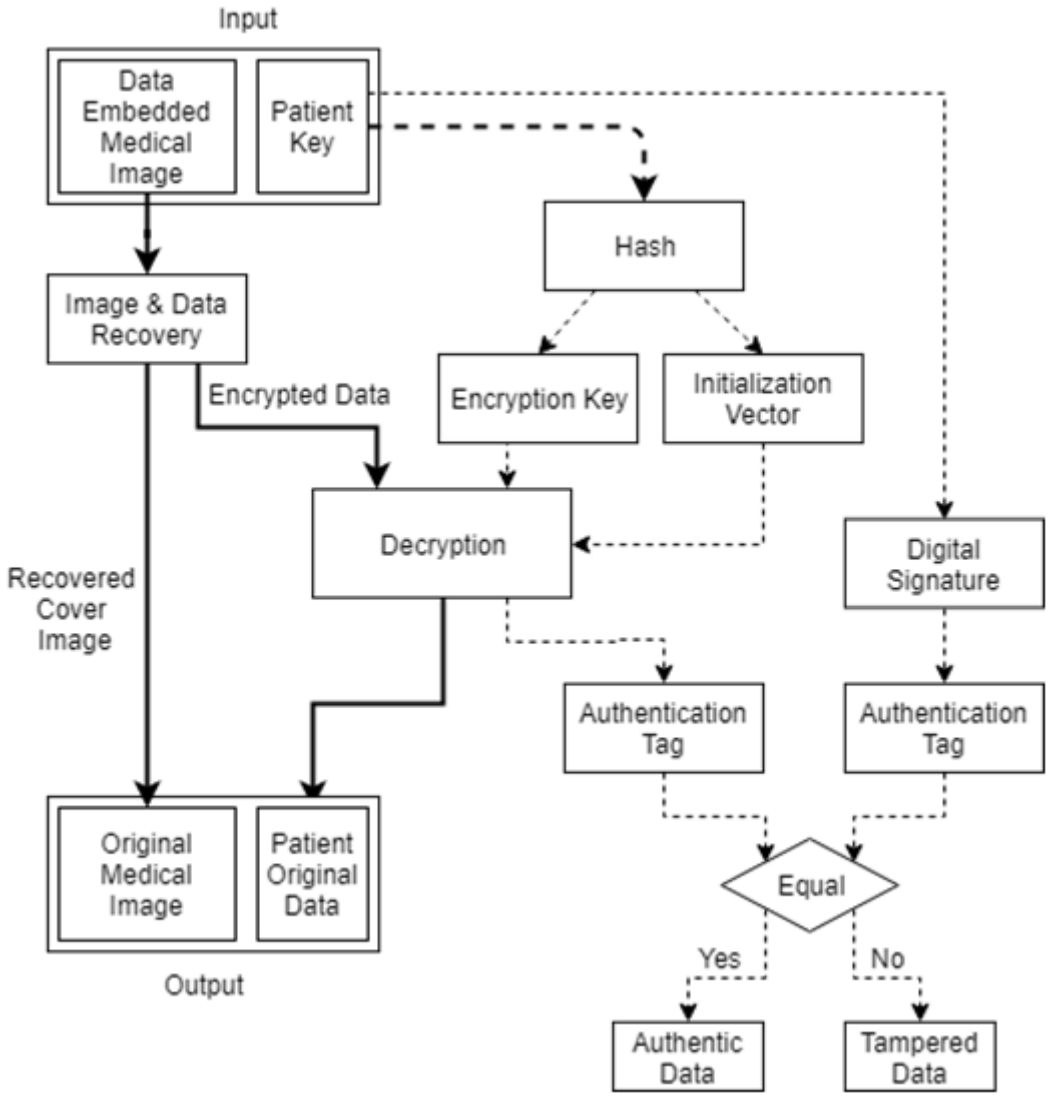

Figure 9

Data recovery, decryption and verification

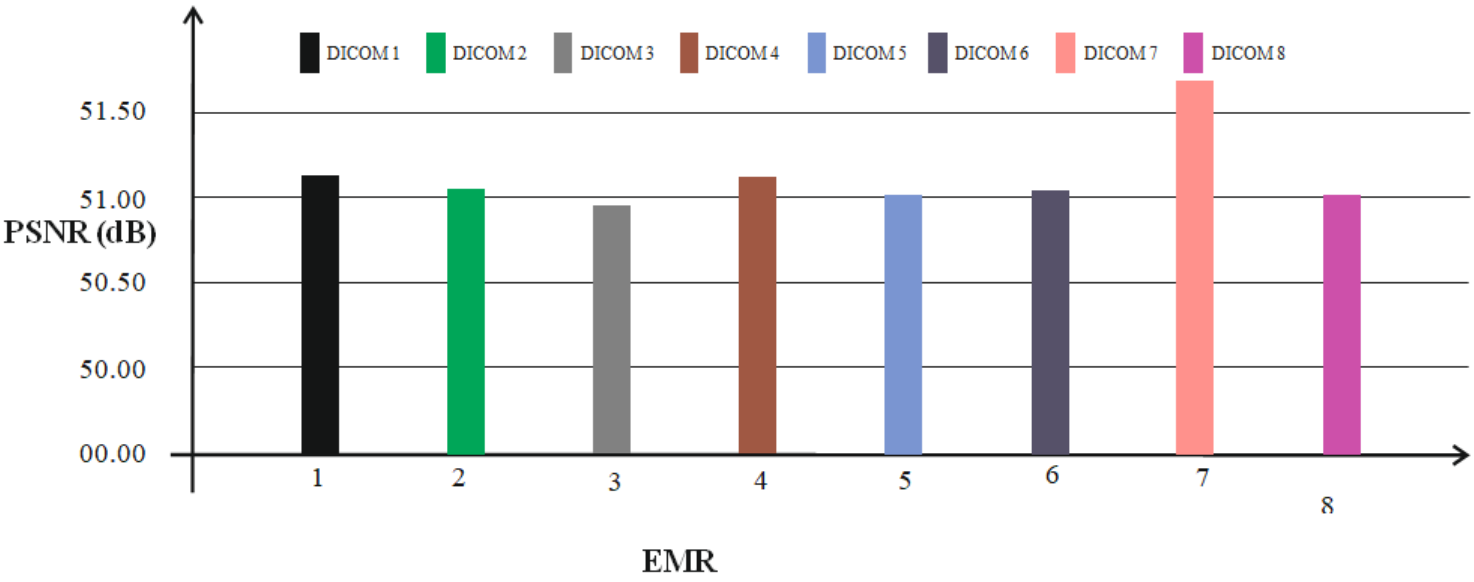

Figure 10

Similarity of PSNR value

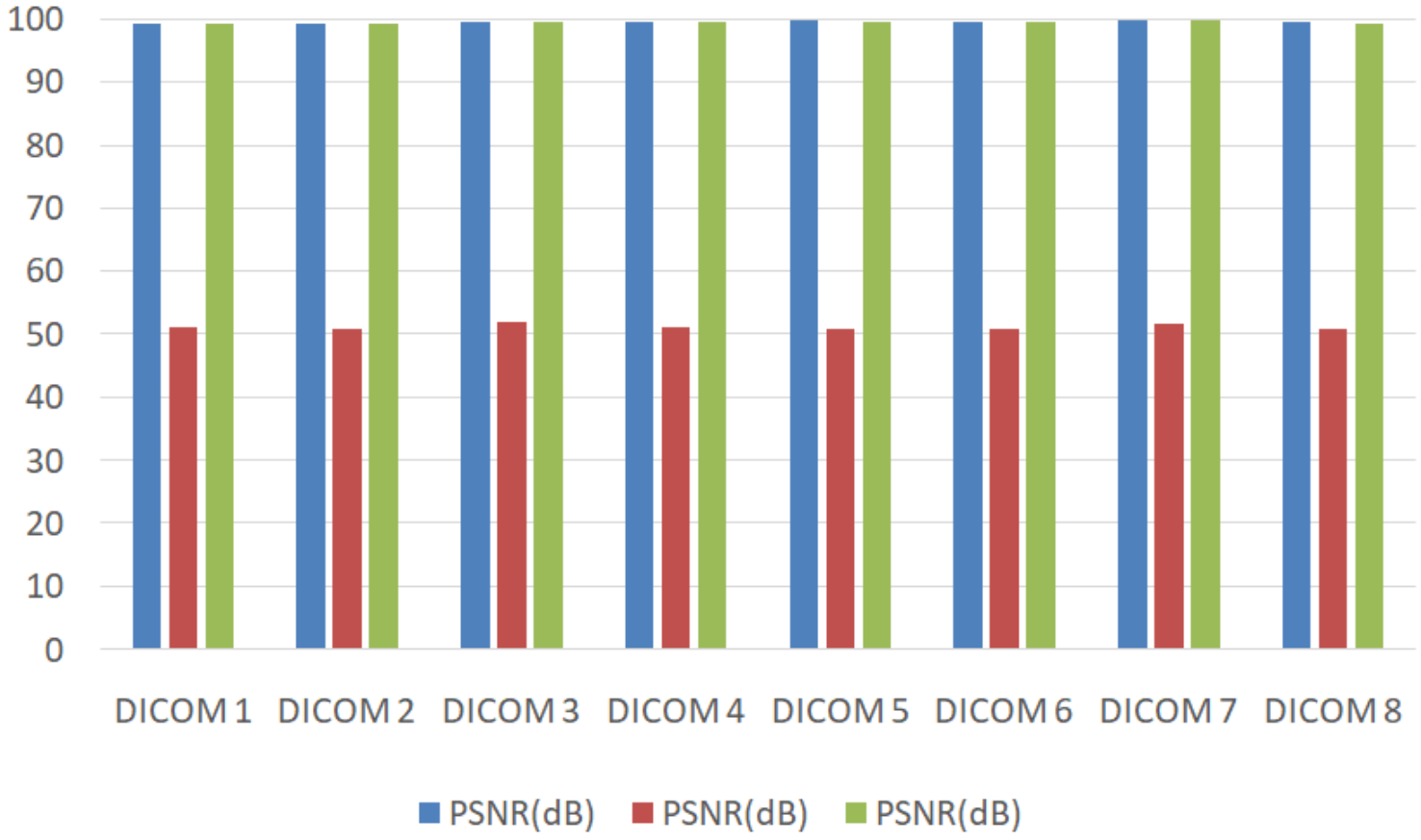

Figure 11

Reversibility Analysis