Indonesian Journal of Electrical Engineering and Computer Science
Vol. 27, No. 2, August 2022, pp. 885~891
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v27.i2.pp885-891

Low feature dimension in image steganographic recognition

Ismail Taha Ahmed[1], Norziana Jamil[2], Baraa Tareq Hammad[1]
[1]Department of Computer Science, College of Computer Sciences and Information Technology, University of Anbar, Anbar, Iraq
[2]Department of Computing (CCI), College of Computing and Informatics, Universiti Tenaga Nasional, Kajang, Malaysia

Article Info

Article history:

Received Jul 26, 2021
Revised May 9, 2022
Accepted Jun 9, 2022

Keywords:

Ada-Boost classifier
Gaussian discriminant analysis classifier
Steganalysis recognition gray level co-occurrence matrix
Steganography

ABSTRACT

Steganalysis aids in the detection of steganographic data without the need to know the embedding algorithm or the "cover" image. The researcher's major goal was to develop a Steganalysis technique that might improve recognition accuracy while utilizing a minimal feature vector dimension. A number of Steganalysis techniques have been developed to detect steganography in images. However, the steganalysis technique's performance is still limited due to their large feature vector dimension, which takes a long time to compute. The variations of texture and properties of an embedded image are clearly seen. Therefore, in this paper, we proposed Steganalysis recognition based on one of the texture features, such as gray level co-occurrence matrix (GLCM). As a classifier, Ada-Boost and Gaussian discriminant analysis (GDA) are used. In order to evaluate the performance of the proposed method, we use a public database in our proposed and applied it using IStego100K datasets. The results of the experiment show that the proposed can improve accuracy greatly. It also indicates that in terms of accuracy, the Ada-Boost classifier surpassed the GDA. The comparative findings show that the proposed method outperforms other current techniques especially in terms of feature size and recognition accuracy.

Corresponding Author:

Norziana Jamil
Department of Computing (CCI), College of Computing and Informatics, Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
Email: norziana@uniten.edu.my

1. INTRODUCTION

Due to the rapid growth of social networking sites, we may see or receive a large number of photographs, but we have no way of knowing whether these images are original or encrypted. Steganography [1] is a method of hiding private information in media such as text, audio, image, and video without leaving any trace that they are encrypted as shown in Figure 1. Therefore, we urgently require methods to distinguish photos containing an encrypted object. The goal of blind Steganalysis is to detect steganographic data without knowing the embedding algorithm or the ‘cover' image.

Figure 2 depicts a general taxonomy of Steganalysis techniques, which is separated under multimedia data types and domains. Steganalysis approaches are classified into two types, signature steganalysis and statistical steganalysis, according to Steganalysis detection methods in literature review [3]. Statistical steganalysis is the process of seeking to find such statistical traces. When compared to signature steganalysis, statistical steganalysis is a more powerful tool since mathematical procedures are more sensitive than visual perception [2]-[4].

The majority of steganalysis approaches rely on image statistical calculations such as first and second order statistics. Statistical and signature steganalysis can be divided into two categories: specific and universal. Specific steganalysis is created for a particular steganographic embedding algorithm, such as least significant bit (LSB) embedding, LSB matching, spread spectrum, bit-plane complexity segmentation (BPCS), joint photographic experts group (JPEG) compression, and other transform domains [5], [6], whereas universal steganalysis is a general class steganalytic technique that can be used with any steganographic embedding algorithm, including unknown algorithms [7], [8].

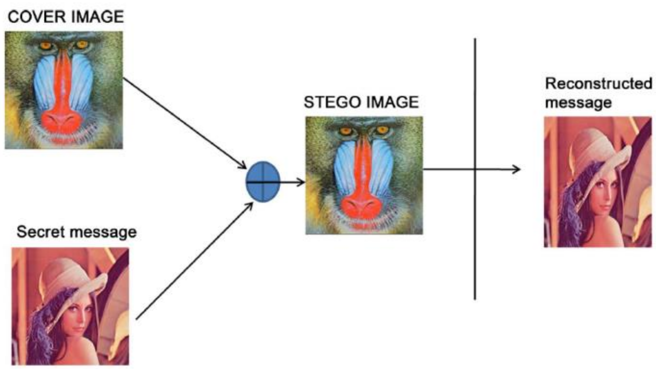

Figure 1. General architecture of steganography [2]

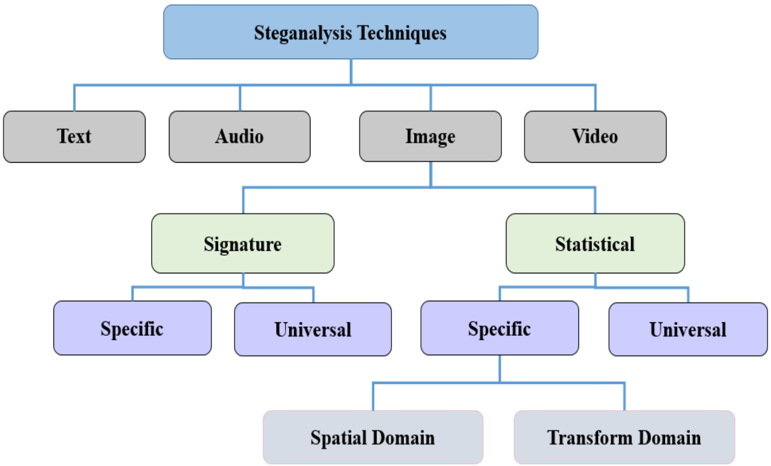

Figure 2. General taxonomy of steganalysis

Various methods for detecting the hidden image have been developed in the literature. However, most of Steganalysis recognition methods [9]–[13] rely on handcrafted features. Jyoth *et al*. [9], proposed a steg analysis recognition based on gray level co-occurrence matrix (GLCM), discrete wavelet transform (DWT) and contourlet transform (CT) and as well as an Adaboost classifier. Song *et al.* [12] proposed a steganalysis recognition based on the Shannon entropy of 2D Gabor wavelets and as well as an ensemble classifier. Karimi *et al*. [13] proposed a steganalysis recognition based on discrete cosine transform (DCT) coefficients and as well as an ensemble classifier. Gui *et al*. [14] proposed a steganalysis recognition method based on local binary pattern (LBP). In summary, smooth pixels are used to extract multi-scaled rotation invariant LBPs as distinguishing features. After that, linear support vector machine (SVM) is used to train and classify features. Zhang and Ping [15] presented a steganalysis recognition method relied on statistical analyses of differential image histograms. Lin *et al*. [16] presented a steganalysis recognition method relied on local ternary pattern (LTP) and path integral (pi-LBP) features combined. Liu *et al*. [17] proposed a steganalysis of LSB matching steganography based on generalized Gaussian distribution (GGD) in the wavelet domain. Chhikara and Bansal [18] proposed a steganalysis recognition based on GLCM as well as J48, sequential minimal optimization (SMO) and Naïve Baye's classifier.

Although the preceding studies have many benefits, it also has certain limitations, including a high feature dimension and a long computation time. Because of these constraints, we concentrate our efforts on an efficient few feature that, at the same time, help the classifier perform better. Therefore, in this research, we proposed Steganalysis recognition based on one of texture features such as GLCM. As a classifier, Ada-Boost and Gaussian discriminant analysis (GDA) are utilized.

The remainder of this paper is organized as shown in section 2 discusses the GLCM features. In section 3, suggested methods are described. Section 4 discusses the findings and analysis. Finally, conclusions can be formed in Section 5.

2. GLCM FEATURE

Texture analysis is crucial in a variety of applications. Steganalysis is one of the most significant [19], [20]. The reason for this is that the information hidden in the images is very difficult to discover or distinguish with human eye, therefore texture analysis is used to uncover information that the human eye cannot see it [21]. For example, when any image embedding the secret data in an image", the texture and characteristics in an image deviated. Therefore, texture analysis may easily uncover these buried details.

GLCM descriptors are typical texture features that are used to extract texture features. The GLCM descriptors [22] are based on statistical moments and are obtained from a co-occurrence matrix. In order to know the process of GLCM calculation, Figure 3 [23] illustrates an example of GLCM calculation. For more details, see [22]–[24].

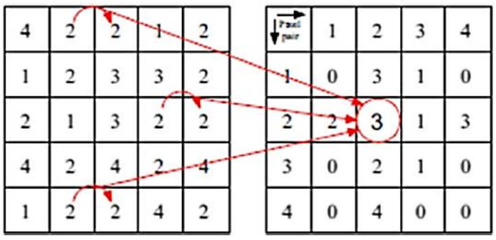

Figure 3. A GLCM computation procedure [60B]

3. THE PROPOSED METHODS

Any steganalysis recognition technique aims to recognize steganography in images from data sets that comprise both the cover image and the hidden image. Figure 4 depicts the three phases of the suggested technique. The following are the steps and algorithms:

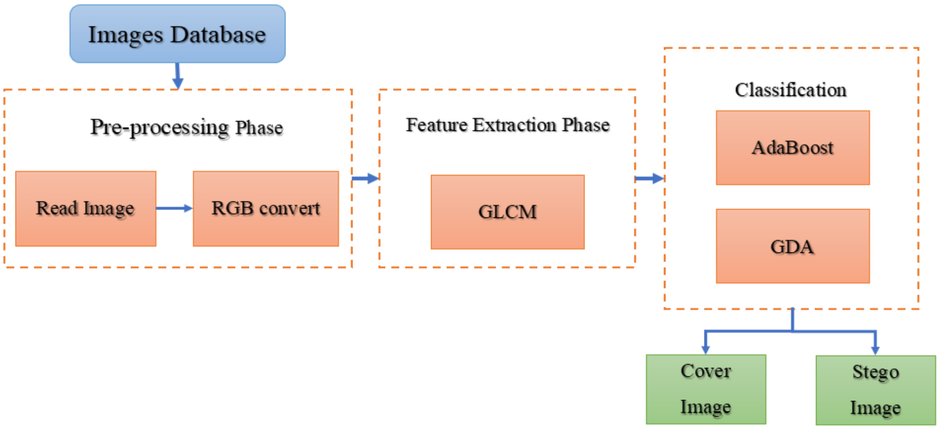

Figure 4. Proposed method block diagram

3.1. Preprocessing

The majority of previously proposed approaches relied on a crucial step known as preprocessing. This step is important to minimize the overall computational complexity. The operation is done by producing a grayscale image from RGB image [25].

3.2. Feature extraction

There are two primary considerations to think about when choosing appropriate features: reducing dimensionality and avoiding redundancy. For each gray scale image, GLCM features are extracted. The dimension of obtained GLCM feature vector is 1×14.

3.3. Classification

Image steganalysis recognition is a two-class problem in this case, with cover and stego being the two classes. Therefore, we must devise ways for classifying those images. As a classifier, the adaptive boosting method (Ada-Boost) is applied. Adaptive boosting is a well-known supervised-learning method that employs many sequential learners, each with a different weight [26]. Furthermore, the Gaussian discriminant analysis (GDA) [27], [28] is a well-known generative model used to execute the classification task [29].

4. RESULTS AND ANALYSIS

The results and analysis section discusses the experimental findings and evaluates the proposed method's performance. The proposed method has also been compared to others. We used the public database IStego100K (Large-scale Image Steganalysis Dataset) [30] as the data set for the proposed approach, which contains 8,104 images with cover/stego sizes of 1,024*1,024. There are 4,052 images that are covered and another 4,052 that are stego.

As shown in (1) was utilized to calculate the accuracy of our proposed recognition approach [31].

$$Accurcy = \frac{(Cover-Stego\ images\ detected\ )}{(Total\ Cover-Stego\ images)} \times 100\ \% \qquad (1)$$

The accuracy can be determined using (1) by computing the proportion of covered and stego images properly detected in the IStego100K dataset.

We use the Ada-Boost and GDA classifiers to classify the GLCM feature vector since different classifiers have varying classification performance. We performed 10-fold cross-validation in order to get accurate results. The recognition accuracy of GLCM feature extraction, Ada-Boost and GDA classifier over the IStego100K database is shown in Table 1.

Table 1. Recognition accuracy of two classifiers across IStego100K database

Feature	Classifier	
	Ada-Boost	GDA
GLCM	97.36%	69.48%

Despite the fact that all classifiers use the same feature vector, they yield different outputs. This is due to the fact that each classifier has its own range of attributes. Figure 5 shows that the Ada-Boost classifier has a 97 percent accuracy. As a result, it's reasonable to believe that the Ada-Boost classifier outperforms the GDA. The Ada-Boost classifier, according to the results, is the best of our proposed method.

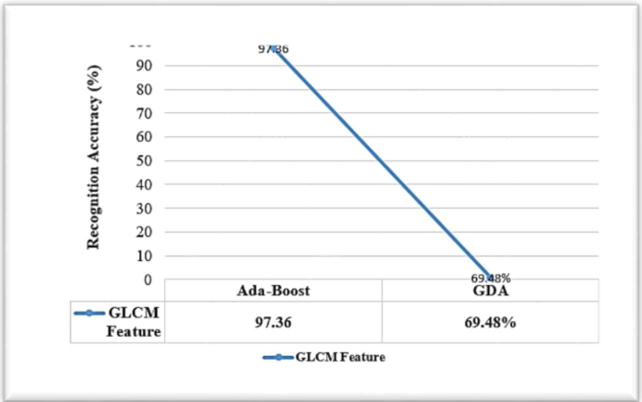

Figure 5. Recognition accuracy of Ada-Boost and GDA classifiers across IStego100K database

Table 2 compares the performance of our proposed method to that of prior methods [9], [10], [32] in the term of recognition accuracy and feature vector dimension. According to recognition accuracy and small feature vector dimension, the presented method surpasses other existing techniques, as shown by the results. The presented method has fewer feature vector dimensions than other previous techniques, as seen in Figure 6. It simplifies the techniques in terms of computing. Both techniques [9] and [32] produced positive outcomes. However, they both use 416 and 22,130 feature vectors. The high feature dimension necessitates a significant amount of computation. Finally, as shown in Table 2, our proposed method can reduce the number of features needed in image Steganalysis while maintaining classification accuracy.

Table 2. Performance comparison with previous methods

Methods	Jyoth *et al.* [9]	Farshid and Ghaemmaghami [10]	Qin *et al*. [32]	Proposed
Feature Vector Dim	416	128	22,130	14
Features kind	GLCM+DWT+CT	Clouds-Min-Sum and Local-Entropies-Sum	GLCM	GLCM
Recognition Accuracy (%)	93.87	78	83	97.36

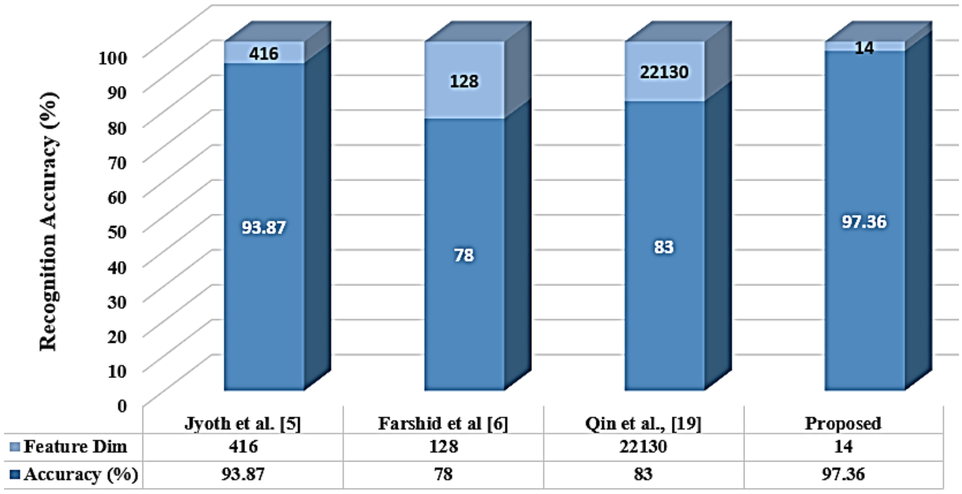

Figure 6. Current methods' performance across different types of feature vector dimensions

5. CONCLUSION

The researcher's major goal was to develop a Steganalysis technique that might improve recognition accuracy while utilizing a minimal feature vector dimension. In this paper, we proposed Steganalysis recognition based on one of the texture features, such as GLCM. As a classifier, Ada-Boost is used. The recognition accuracy of the proposed method using GLCM and Ada-Boost over the IStego100K database is found to be 97 percent. The proposed method outperforms other current methods in terms of recognition accuracy and a small feature vector dimension, as shown by the results. The presented method has fewer feature vector dimensions than other previous techniques. It simplifies the techniques in terms of computing. Finally, our proposed method can minimize the feature dimension needed in image Steganalysis while maintaining classification accuracy.

ACKNOWLEDGEMENTS

This research is supported by Uniten BOLD publication fund 2022.

REFERENCES

[1] S. Dhawan and R. Gupta, "Analysis of various data security techniques of steganography: A survey," *Information Security Journal*, vol. 30, no. 2, pp. 63–87, 2021, doi: 10.1080/19393555.2020.1801911.

[2] J. Babu, S. Rangu, and P. Manogna, "A survey on different feature extraction and classification techniques used in image steganalysis," *Journal of Information Security*, vol. 08, no. 03, pp. 186–202, 2017, doi: 10.4236/jis.2017.83013.

[3] N. Subramanian, O. Elharrouss, S. Al-Maadeed, and A. Bouridane, "Image steganography: A review of the recent advances," *IEEE Access*, vol. 9, pp. 23409–23423, 2021, doi: 10.1109/ACCESS.2021.3053998.

[4] Y. JinaChanu, K. Manglem Singh, and T. Tuithung, "Image steganography and steganalysis: a survey," *International Journal of Computer Applications*, vol. 52, no. 2, pp. 1–11, 2012, doi: 10.5120/8171-1484.

[5] P. C. Mandal, "An extensive review of current trends in steganalysis," *Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, no. 7, pp. 215–220, 2012.

[6] Z. Wang, Z. Li, X. Song, and Y. Zhang, "Robust JPEG image steganography based on SVD and QIM in stationary wavelet domain," in *Communications in Computer and Information Science*, 2021, vol. 1424, pp. 551–560, doi: 10.1007/978-3-030-78621-2_46.

[7] A. Nissar and A. H. Mir, "Classification of steganalysis techniques: A study," *Digital Signal Processing: A Review Journal*, vol. 20, no. 6, pp. 1758–1770, 2010, doi: 10.1016/j.dsp.2010.02.003.

[8] I. T. Ahmed, B. T. Hammad, and N. Jamil, "Common gabor features for image watermarking identification," *Applied Sciences (Switzerland)*, vol. 11, no. 18, p. 8308, 2021, doi: 10.3390/app11188308.

[9] T. S. Jyothy, G. Sreelatha, R. Pradeep, and V. Sajith, "Texture-based multiresolution steganalytic features for spatial image steganography," in *Proceedings of the 2nd International Conference on Smart Systems and Inventive Technology, ICSSIT 2019*, 2019, pp. 966–971, doi: 10.1109/ICSSIT46314.2019.8987907.

[10] F. Farhat and S. Ghaemmaghami, "Towards blind detection of low-rate spatial embedding in image steganalysis," *IET Image Processing*, vol. 9, no. 1, pp. 31–42, 2015, doi: 10.1049/iet-ipr.2013.0877.

[11] G. Gul and F. Kurugollu, "A new methodology in steganalysis: Breaking highly undetectable steganograpy (HUGO)," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011, vol. 6958 LNCS, pp. 71–84, doi: 10.1007/978-3-642-24178-9_6.

[12] X. Song, Z. Li, L. Chen, and J. Liu, "Entropy feature based on 2D gabor wavelets for JPEG steganalysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 10067 LNCS, pp. 59–72, doi: 10.1007/978-3-319-49145-5_7.

[13] H. Karimi, M. G. Shayesteh, and M. A. Akhaee, "Steganalysis of JPEG images using enhanced neighbouring joint density features," *IET Image Processing*, vol. 9, no. 7, pp. 545–552, 2015, doi: 10.1049/iet-ipr.2013.0823.

[14] X. Gui, X. Li, and B. Yang, "Steganalysis of lsb matching based on local binary patterns," in *Proceedings - 2014 10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2014*, Aug. 2014, pp. 475–480, doi: 10.1109/IIH-MSP.2014.125.

[15] T. Zhang and X. Ping, "A new approach to reliable detection of LSB steganography in natural images," *Signal Processing*, vol. 83, no. 10, pp. 2085–2093, 2003, doi: 10.1016/S0165-1684(03)00169-5.

[16] Q. Lin, J. Liu, and Z. Guo, "Local ternary pattern based on path integral for steganalysis," in *Proceedings - International Conference on Image Processing, ICIP, 2016*, vol. 2016-August, pp. 2737–2741, doi: 10.1109/ICIP.2016.7532857.

[17] Q. Liu, A. H. Sung, J. Xu, and B. M. Ribeiro, "Image complexity and feature extraction for steganalysis of LSB matching steganography," in *Proceedings - International Conference on Pattern Recognition*, 2006, vol. 2, pp. 267–270, doi: 10.1109/ICPR.2006.684.

[18] Ashu, R. R. Chhikara, and D. Bansal, "GLCM based features for steganalysis," in *Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit*, 2014, pp. 385–390, doi: 10.1109/CONFLUENCE.2014.6949284.

[19] Y. Q. Shi, C. Chen, G. Xuan, and W. Su, "Steganalysis versus splicing detection," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2008, vol. 5041 LNCS, pp. 158–172, doi: 10.1007/978-3-540-92238-4_13.

[20] I. T. Ahmed, B. T. Hammad, and N. Jamil, "Forgery detection algorithm based on texture features," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 226–235, 2021, doi: 10.11591/ijeecs.v24.i1.pp226-235.

[21] Z. Liu, J. Deng, Z. Lin, J. Xie, L. Li, and Y. Jiang, "A color image steganography scheme based on texture features," in *Proceedings - 2021 International Conference on Computer Technology and Media Convergence Design, CTMCD 2021*, 2021, pp. 10–13, doi: 10.1109/CTMCD53128.2021.00010.

[22] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979, doi: 10.1109/PROC.1979.11328.

[23] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1311–1328, 2021, doi: 10.1007/s00521-020-05017-z.

[24] P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image texture feature extraction using GLCM approach," *International Journal of Scientific & Research Publication*, vol. 3, no. 5, pp. 1–5, 2013.

[25] I. T. Ahmed, B. T. Hammad, and N. Jamil, "Image steganalysis based on pretrained convolutional neural networks," in *2022 IEEE 18th International Colloquium on Signal Processing \& Applications (CSPA*), 2022, pp. 283–286, doi: 10.1109/CSPA55076.2022.9782061.

[26] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 904, no. 1, pp. 23–37, 1995, doi: 10.1007/3-540-59119-2_166.

[27] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, 1995, doi: 10.1109/76.350779.

[28] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[29] B. T. Hammad, I. T. Ahmed, and N. Jamil, "A steganalysis classification algorithm based on distinctive texture features," *Symmetry*, vol. 14, no. 2, p. 236, 2022, doi: 10.3390/sym14020236.

[30] Z. Yang, K. Wang, S. Ma, Y. Huang, X. Kang, and X. Zhao, "IStego100K: large-scale image steganalysis dataset," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12022 LNCS, pp. 352–364, 2020, doi: 10.1007/978-3-030-43575-2_29.

[31] I. T. Ahmed, B. T. Hammad, and N. Jamil, "Effective deep features for image splicing detection," in *2021 IEEE 11th International Conference on System Engineering and Technology, ICSET 2021 - Proceedings*, 2021, pp. 189–193, doi: 10.1109/ICSET53708.2021.9612569.

[32] J. Qin, X. Xiang, Y. Deng, Y. Li, and L. L. Pan, "Steganalysis of highly undetectable steganography using convolution filtering," *Information Technology Journal*, vol. 13, no. 16, pp. 2588–2592, 2014, doi: 10.3923/itj.2014.2588.2592.

BIOGRAPHIES OF AUTHORS

Ismail Taha Ahmed received his B.E. and M.Sc. degrees in Computer Science from College of Computer Science and Information Technology, University of Anbar, Anbar, in 2005 and 2009, respectively. He received his Ph.D. degrees in Computer Science from College of Computer Science and Information Technology, Universiti Tenaga Nasional, Putrajaya, Malaysia, in 2018. His research interests Include Image Processing, Image Quality Assessment, Deep Learning, and Computer Vision. He can be contacted at email: ismail.taha@uoanbar.edu.iq.

Norziana Jamil received her BSc (Information Technology), 2000, form Universiti Kebangsaan Malaysia, and she received her M.Sc. (Information Security), 2005, from Royal Holloway Universiti of London, UK, while she finish her PhD (Security in Computing), 2013, from UPM university, She is interested in cryptography, Authentication, SCADA system, wireless sensor network. She can be contacted at email: norziana@uniten.edu.iq.

Baraa Tareq Hammad received her B.E. and M.Sc. degrees in Computer Science from College of Computer Science and Information Technology, University of Anbar, Anbar, in 2005 and 2012, respectively. She received her Ph.D. degrees in Computer Science from College of Computer Science and Information Technology, Universiti Tenaga Nasional, Putrajaya, Malaysia, in 2018. Her research interests Include Information Security, IoT and Network Security. She can be contacted at email: baraa.tareq@uoanbar.edu.iq.