

## Fast discrimination of fake video manipulation

Wildan Jameel Hadi<sup>1</sup>, Suhad Malallah Kadhem<sup>2</sup>, Ayad Rodhan Abbas<sup>2</sup>

<sup>1</sup>Department of Computer Science, Baghdad University, Baghdad, Iraq

<sup>2</sup>Department of Computer Science, Al-Technology University, Baghdad, Iraq

### Article Info

#### Article history:

Received Jul 28, 2021

Revised Nov 10, 2021

Accepted Nov 30, 2021

#### Keywords:

Deepfake

Focus measures

Multimedia forensics

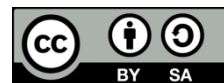
Video manipulation

Visible artifacts

### ABSTRACT

Deepfakes have become possible using artificial intelligence techniques, replacing one person's face with another person's face (primarily a public figure), making the latter do or say things he would not have done. Therefore, contributing to a solution for video credibility has become a critical goal that we will address in this paper. Our work exploits the visible artifacts (blur inconsistencies) which are generated by the manipulation process. We analyze focus quality and its ability to detect these artifacts. Focus measure operators in this paper include image Laplacian and image gradient groups, which are very fast to compute and do not need a large dataset for training. The results showed that i) the Laplacian group operators, as a value, may be lower or higher in the fake video than its value in the real video, depending on the quality of the fake video, so we cannot use them for deepfake detection and ii) the gradient-based measure (GRA7) decreases its value in the fake video in all cases, whether the fake video is of high or low quality and can help detect deepfake.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Wildan Jameel Hadi

Department of Computer Science, Baghdad University

Bagdad, Iraq

Email: wildanjh\_comp@cs.w.uobaghdad.edu.iq

## 1. INTRODUCTION

At present, when we talk about artificial intelligence, we do not just mean a set of theories; on the contrary, it has entered many practical applications [1]–[6]. It has become an essential part of many industries. On the other hand, by looking at artificial intelligence algorithms that depend on deep learning, we see the potential of these algorithms to solve complex problems because they try to simulate the human mind. Also, with advanced deep learning techniques, tampering with media is possible, provided large amounts of data are obtained for training. The main concern here is using these artificial intelligence (AI) tools for malicious purposes, such as creating vulgar videos or creating false advertising campaigns. In recent years, fake multimedia has become a problem that must be focused on and solved, especially after the emergence of so-called deepfakes. Deepfakes are images and video clips that were tampered with using artificial intelligence [7], such as generative adversarial networks (GANs) [8]–[11] or auto encoders (AEs) [12], [13].

The term deepfake came from a Reddit user who converted celebrity faces into porn videos by developing a machine learning algorithm [14]. Technology companies and those responsible for social media platforms have made a wide effort to uncover so-called deepfakes attacks. For example, Microsoft designed software to detect deepfakes. This program checks the images and videos and then generates a percentage indicating the extent to which the input material is artificially created [15]. In 2019, Facebook announced the deepfake detection challenge (DFDC) in cooperation with major technology companies. This initiative stimulated the creation of new methods that detect deepfakes. They created a giant database of more than 100,000 videos [16]. Facebook also banned all the videos created using artificial intelligence in a way that

could not be detected by the users of this platform [17]. The defense advanced research projects agency (DARPA) and the air force research laboratory (AFRL) have funded the media forensics research project (MediFor), whose purpose is to find solutions and technologies to ensure the integrity of digital images and videos [18]. Three main components are considered in the MediFor program, namely: i) digital integrity, ii) physical integrity, and iii) semantic integrity [19].

In many cases, video evidence is assumed in criminal investigations to be reliable. But with the development of technology, where it became possible to tamper with videos, and audio recordings, it is likely soon to subject such evidence to a test to ensure its authenticity [20]. Therefore, finding ways to help detect deepfakes is of great importance to preserve the legitimacy of digital evidence.

In this paper, we investigate the detection of these tampered video contents, especially deepfake videos. Our method uses four focus measures (Laplacian and gradient) to evaluate the amount of blur in the frames belonging to the fake video and compare it to the real video. We focus on assessing the amount of blurring because most deepfake applications use blurring to hide the border of the cropped face. Our contribution to this work can be summarized as follows: i) we use a simple way to detect the blurring residual in fake videos using the focus measures; ii) a small dataset is used to implement our work without using a large dataset for training; and iii) our method uses simple measures to detect the forgery instead of using deep learning methods, which do not easily detect blurring residuals.

## 2. RELATED WORK

This section will briefly mention some of the works related to deepfake detection and focus on methods that explore artifacts in fake video frames. Concerning the detection of deepfakes, image manipulation is not a new problem that has been highlighted since the emergence of deepfakes; for example, image modification with Photoshop tools is being used up to the present day. Also, multimedia forensics science is still dealing with this problem [17]. Deepfake algorithms leave traces on media that are difficult to detect with the naked eye, so more research is devoted to helping detect deepfakes. Figure 1 shows the number of research papers that have provided solutions in the field of deepfake detection.

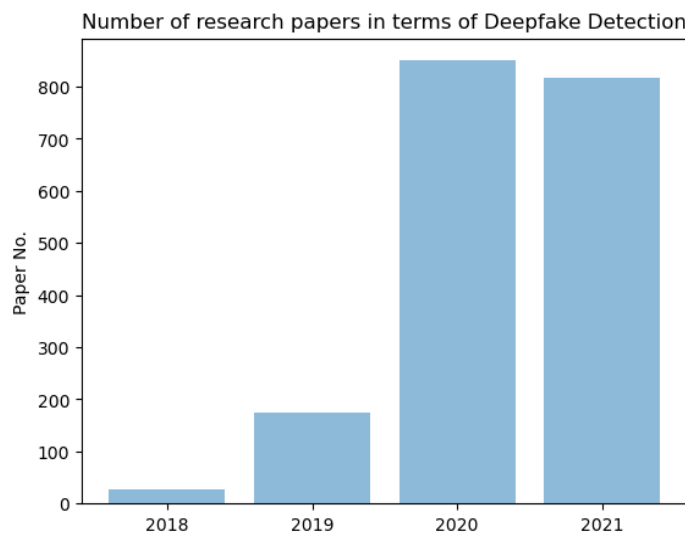


Figure 1. The number of research papers that are implemented in terms of deepfake detection [21]

Methods for detecting fakes in a video can be classified into two categories [7]: techniques that depend on temporal features and those that rely on visible artifacts within frames. We concentrate on the second type, which breaks down the video into frames and then works on each frame by searching for the visual artifacts to obtain distinctive properties or features and then feeding them to a classifier. For example, Koopman *et al.* [20] used photo response non-uniformity (PRNU) to detect deepfakes manipulation. The analysis of PRNU is an attractive method because it finds that tampering with the facial area region affects the local value of the PRNU pattern in video frames. This analysis is widely used in the field of video forensics [22]. Compression artifacts are also exploited in this area. It was found that when the manipulated JPEG-image is compressed a second time, the whole image has the effects of double compression, except for

the fake face area [23]. In addition to the defects generated by recompression, the process of deepfakes also generates a set of precious traces, which need post-processing operations to hide those traces or defects. Bahrami *et al.* [24] divided the blurred image into blocks to extract the characteristics of the blur type. Finally, these local blur characteristics are combined to classify the image blocks into motion or out of focus. These artifacts, especially the artifacts resulting from blurring the border of the fake face, were used in this experiment for detecting fakes in videos.

### 3. FOCUS MEASURES

Focus measures (FM) are useful for the measurement of the amount of blur contained in an image. Pertuz *et al.* [25] divided these measures into six categories: gradient-based (GB) and Laplacian-based (LB) operators, which are based on the first derivative (gradient) and second derivative to measure the degree of focus and the number of edges found in the input image, respectively. Wavelet-based (WB) and DCT-based (DCT-B) operators use the fact that the coefficients of the discrete wavelet transform and discrete cosine transform can be used to measure the focus level. Statistics-based (SB) operators use image statistics to find focus levels. Finally, miscellaneous operators, which do not belong to any of the prior groups, can measure the amount of blur founded in the image.

Our work was based on the focus measures operators mentioned in [25]. The selection fell on the Laplacian family because of its good performance in common imaging conditions. Also, we use the gradient-based operator because it showed promising results from among the 11 high-performance scales tested in [25]. We did not refer to other standards, such as deep cryogenic treatment (DCT), due to their dependence on their applications. All focus measures considered in this paper are presented in Table 1, along with their exact abbreviation from [25].

Table 1. Focus measures used and their abbreviation

Focus Measure	Abbr.
Energy of Laplacian	LAP1
Modified Laplacian	LAP2
Diagonal Laplacian	LAP3
Tenengrad Variance	GRA7

## 4. METHOD

### 4.1. Dataset

The dataset used in our experiment is the UADFV dataset [26]. It is simple and has only 49 videos with two classes, fake and real. Figure 2 shows a sample of this dataset. We use ten videos from this dataset to implement our experiment.

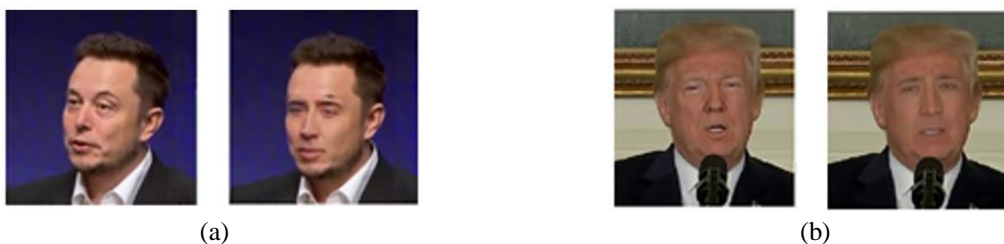


Figure 2. Sample of the dataset used (a) is an example of deepfakes with low quality (notable edges), while (b) shows an example of deepfakes with high quality (polished edges)

### 4.2. Focus measure operators' analysis

Each video in the dataset is converted into a series of frames. This step is followed by cropping each frame into eight parts to apply the focus measure mentioned in Table 1 on each part. Figure 3 shows how to divide each frame into eight parts. Before applying the focus measure operators to the split image, the grayscale channel of the image is taken into consideration. After applying each measure on each part, the average value of the eight measure values is calculated for each frame image. This process is repeated for all video frames, whether real or fake, to calculate the final value of the focus measure operators.

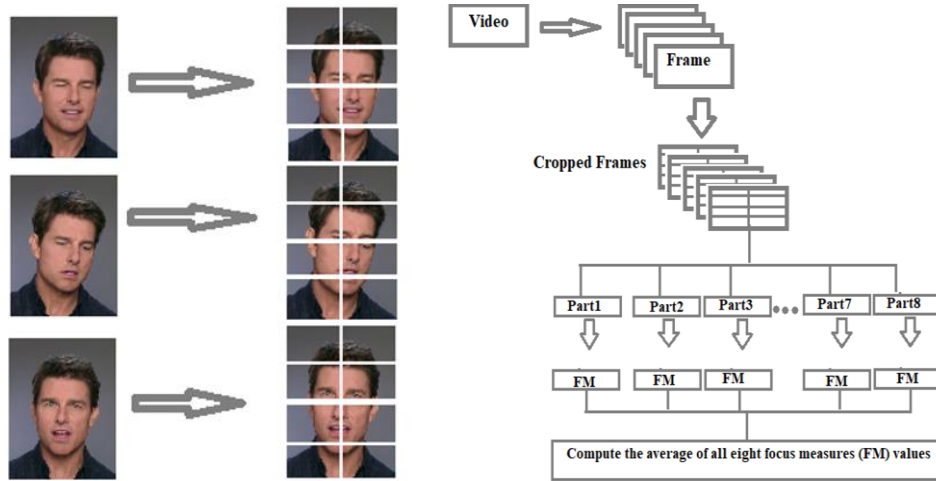


Figure 3. Overall steps for analyzing the quality of focus measures. The video frames are cropped into eight parts where an average value of the focus measure (FM) is calculated for each frame

### 5. RESULTS

The focus measure for each video was calculated. Figures 4(a)-(d) show the results energy of Laplacian, modified Laplacian, Laplacian variance, and Tenengrad variance, respectively. From these results, we note that the values of the Laplacian-based operators were lower or higher than their values in the real video, depending on the number of edges present in the fake video. We know that the edges decrease as more blurring is applied to the image.

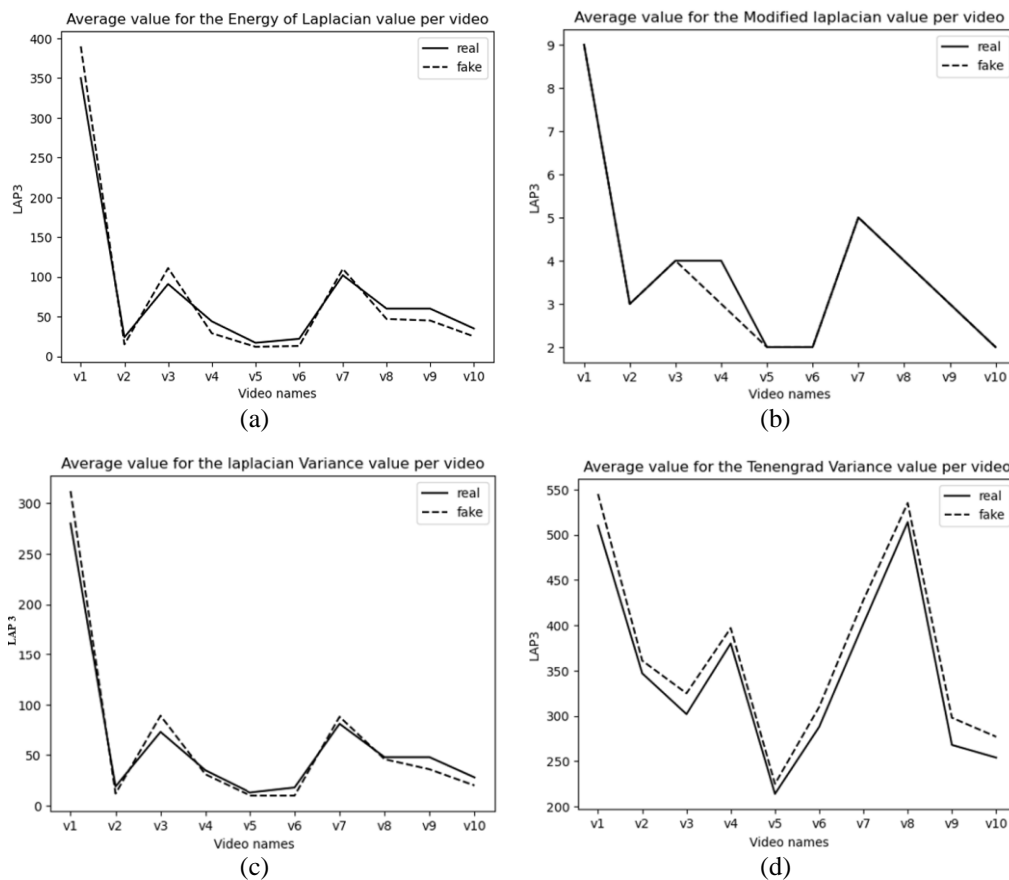


Figure 4. The results of applying focus measure operators: (a) energy of Laplacian, (b) modified Laplacian, (c) Laplacian variance, and (d) Tenengrad variance, per fake videos against real one

By looking at the results achieved by the gradient-based operator, we find that the value of this measure is less than its value in a real video in all cases, whether the fake video is of high quality or low. The results indicate that we can use the gradient-based measure to detect deepfakes. This explains the algorithm's performance in terms of deepfake detection. As for time, the response time of a single video is measured in seconds, so if we compare the time taken for all videos, it is negligible in relation to the time taken by deep learning-based methods.

## 6. CONCLUSION

In this analysis, a methodology for deepfake detection is suggested. Our method exploits the visible artifacts (blur inconsistencies) generated via post-processing steps applied on fake videos. Compared to other methods that use deep learning, our method does not require high computational power and extensive data for training. Still, it is based only on the comparison of the performance of various focus measure operators. The selection of operators has been chosen after a comprehensive review of the results of much recent literature. Four focus measure operators are selected to evaluate the degree of focus between fake videos and real ones. Various mathematical principles were used in the analysis and testing of the focus measure operators. Out of 36 operators, the best four operators were selected to analyze and compare their performance. The selection fell on the Laplacian family and the gradient-based operator. Experiments have been implemented on a test set of both fake and real videos. Experiments showed that the values of the Laplacian-based operators were lower or higher than their values in the real video, depending on the number of edges present in the fake video. This means the Laplacian family cannot be used in deepfake detection. On the other hand, the gradient-based measure (GRA7) can distinguish fake videos from real ones. The dataset used in our experiment is overly small to subedit guidelines for likelihood ratios, as is desired in forensic sciences. Nevertheless, we have a simple collection of the focus measures. However, better results could have been achieved with more complex collection strategies. Due to the increasing sophistication of deepfake technology, it becomes necessary to keep pace with these developments and establish robust detection methods. Because the multimedia forensic tools worked on studying and detecting malicious manipulations in multimedia content before the occurrence of deep learning technology, so it was necessary to take advantage of these methods and combine them with deep learning techniques to build a strong detection model.




## REFERENCES

- [1] D. H. Abd, A. T. Sadiq, and A. R. Abbas, "Classifying political arabic articles using support vector machine with different feature extraction," in *Communications in Computer and Information Science*, Springer International Publishing, 2020, pp. 79–94.
- [2] A. R. Abbas and A. O. Farooq, "Human skin colour detection using bayesian rough decision tree," in *Communications in Computer and Information Science*, Springer International Publishing, 2018, pp. 240–254.
- [3] D. H. Abd, A. T. Sadiq, and A. R. Abbas, "Political articles categorization based on different naïve bayes models," in *Communications in Computer and Information Science*, Springer International Publishing, 2020, pp. 286–301.
- [4] E. A. Kadhim, H. B. A. Wahab, and M. Suhad, "Proposed approach for key generation based on elliptic curve (EC) algebra and metaheuristic algorithms," *Engineering Technology Journal*, vol. 32, no. 2, pp. 333–346, 2014.
- [5] A. T. Sadiq and A. G. Hamad, "BSA: a hybrid bees' simulated annealing algorithm to solve optimization and NP-complete problems," *Engineering Technology Journal*, vol. 28, no. 2, pp. 271–281, 2010.
- [6] H. H. Salih, A. T. Sadiq, and I. K. Ali, "Attack on the Simple Substitution Ciphers Using Particle Swarm Optimization," *Eng. & Tech. Journal*, vol. 28, no. 11, 2010.
- [7] L. Verdoliva, "Media forensics and DeepFakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, Aug. 2020, doi: 10.1109/JSTSP.2020.3002101.
- [8] X. Mao and Q. Li, *Generative adversarial networks for image generation*. Singapore: Springer Singapore, 2021.
- [9] X. Wang, W. Li, G. Mu, D. Huang, and Y. Wang, "Facial expression synthesis by U-Net conditional generative adversarial networks," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, Jun. 2018, pp. 283–290, doi: 10.1145/3206025.3206068.
- [10] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 4396–4405, doi: 10.1109/CVPR.2019.00453.
- [11] Y. Yu, Z. Gong, P. Zhong, and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *International Conference on Image and Graphics*, 2017, pp. 97–108.
- [12] Y. Zhou and B. E. Shi, "Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, Oct. 2017, pp. 370–376, doi: 10.1109/ACII.2017.8273626.
- [13] H. Zhu, Q. Zhou, J. Zhang, and J. Z. Wang, "Facial aging and rejuvenation by conditional multi-adversarial autoencoder with ordinal regression," *arXiv:1804.02740*, Apr. 2018.
- [14] M. Westerlund, "The emergence of Deepfake technology: a review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 39–52, Jan. 2019, doi: 10.22215/timreview/1282.
- [15] V. Mehta, P. Gupta, R. Subramanian, and A. Dhall, "FakeBuster: a deepfakes detection tool for video conferencing scenarios," in *26th International Conference on Intelligent User Interfaces*, Apr. 2021, pp. 61–63, doi: 10.1145/3397482.3450726.
- [16] A. A. Pokroy and A. D. Egorov, "EfficientNets for DeepFake detection: comparison of pretrained models," in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, Jan. 2021, pp. 598–600, doi:




- 10.1109/EIConRus51938.2021.9396092.
- [17] L. Guarnera, O. Giudice, C. Nastasi, and S. Battiato, "Preliminary forensics analysis of DeepFake images," in *2020 AEIT International Annual Conference (AEIT)*, Sep. 2020, pp. 1–6, doi: 10.23919/AEIT50178.2020.9241108.
- [18] L. Verdoliva and P. Bestagini, "Multimedia forensics," in *Proceedings of the 27th ACM International Conference on Multimedia*, Oct. 2019, pp. 2701–2702, doi: 10.1145/3343031.3350542.
- [19] M. Hogan, "Replicating reality advantages and limitations of weaponized deepfake technology PIPS." The Project on International Peace and Security, Global Research Institute, College of William & Mary, 2020.
- [20] M. Koopman, A. M. Rodriguez, and Z. Geradts, "Detection of deepfake video manipulation," in *Proceedings of the 20th Irish Machine Vision and Image Processing conference*, 2018, pp. 133–136.
- [21] Digital Science & Research Solutions Inc., "Number of research papers in terms of deepfake detection," *Dimensions*. [https://app.dimensions.ai/discover/publication?search\\_mode=content&search\\_text=Deepfake Detection&search\\_type=kws&search\\_field=full\\_search](https://app.dimensions.ai/discover/publication?search_mode=content&search_text=Deepfake%20Detection&search_type=kws&search_field=full_search) (accessed Nov. 05, 2021).
- [22] W.-C. Yang, J. Jiang, and C.-H. Chen, "A fast source camera identification and verification method based on PRNU analysis for use in video forensic investigations," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6617–6638, Feb. 2021, doi: 10.1007/s11042-020-09763-z.
- [23] T. Pevný and J. Fridrich, "Estimation of primary quantization matrix for steganalysis of double-compressed JPEG images," in *Proceedings of SPIE-The International Society for Optical Engineering*, Feb. 2008, doi: 10.1117/12.759155.
- [24] K. Bahrami, A. C. Kot, Leida Li, and Haoliang Li, "Blurred image splicing localization by exposing blur type inconsistency," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 999–1009, 2015, doi: 10.1109/TIFS.2015.2394231.
- [25] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognition*, vol. 46, no. 5, pp. 1415–1432, May 2013, doi: 10.1016/j.patcog.2012.11.011.
- [26] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8261–8265, doi: 10.1109/ICASSP.2019.8683164.

## BIOGRAPHIES OF AUTHORS






**Wildan Jameel Hadi**    has received her bachelor's degree in computer science from Baghdad University in the year 2006. She has completed her master's degree in Computer Science from the University of Technology in the Year 2008. She is currently pursuing her Ph.D. degree in computer science at Technology University, Iraq, and working in the Department of Computer Science, College of Science for women, University of Baghdad, Iraq. Her research interests include image processing, video analysis, deep learning, face detection. She can be contacted at email: wildanjh\_comp@csu.uobaghdad.edu.iq.



**Suhad Malallah Kadhem**    has finished her Ph.D. in Computer Science from the Department of Computer Science at the Technology University. She has completed her bachelor's and master's degree in Computer Science from the University of Technology (UOT), Baghdad, Iraq in 1997. Suhad is a faculty member in the Computer Science Department at UOT since 1997, where she became the Head of the artificial intelligent branch at UOT in 2003. Her research interests focus on artificial intelligence, natural language processing (especially Arabic language processing), and computer security (especially steganography). She can be contacted at email: 110102@uotechnology.edu.iq.



**Ayad Rodhan Abbas**    has finished his Ph.D., Artificial Intelligent, Wuhan University, School of Computer Science, China, 2009, M.Sc., Computer Science, University of Technology, Computer Science Department, Iraq, 2005, B.S., Computer Science, University of Technology, Computer Science Department, Iraq, 2003, B.S., Chemical Engineering, University of Baghdad, Chemical Engineering Department, Iraq, 1999. His research interests focus on artificial intelligent, machine learning, natural language processing, deep learning, data mining, web mining, information retrieval, soft computing, E-learning, E-Commerce, and recommended system. He can be contacted at email: 110010@uotechnology.edu.iq.