

A digital image recommendation system using Semantic Segmentation

Omar Muayad Abdullah

Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq.
omaraldewachy@uomosul.edu.iq

Article history: Received Jan 27, 2026 | Revised Feb 23, 2026 | Accepted Maret 7, 2026

The research aims to determine the nature and features of a digital image in order to suggest or recommend relative images (candidate) to the image (query) that is determined by the user which may contain high scientific or valuable contents. The proposed method processes the semantic segmentation firstly, then computing the feature similarity, by segmenting the image into meaningful object regions then extracting isolating features for the determined objects, then the system will enhance the matching range between the recommendation query and the output of the semantic level. The sample was consisting of 1380 images based on 70/30 split with selected 3 labels (*cars, persons and trees*). The preprocessing step has been applied where the features were extracted from the determined image in order to determine its contents then we can recommend (candidate) the images with the relative features, this process is applied using the concept of semantic segmentation, where the procedure is partitioned into several steps: The first step includes grouping (*cars, persons and trees*) images from multiple online and real-world datasets, the grouped raw data that can be processed by the system normalization have been processed using Min-Max normalization that is used for standardizing the input data. The second step is the **feature extraction process** which is achieved using a modified VGG-16 net and the fully connected layers have been removed and the output will be a **2-D features map**, then we convert this map to a 1-D vector called **feature vector** using a Global Average Pooling. The next step is obtaining **feature labels** by applying a Support Vector Machine SVM classifier, then we recommended the images with relative features to the determined (query) image, this step is achieved using a Pearson correlation coefficient. The final step is constructing a Confusion Matrix and applying the (F1-score, Recall, Precision and Accuracy) metrics in order to estimate the performance.

Keywords: Semantic segmentation, Pearson Correlation Coefficient, re commendation system, Global average pooling.



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

1. INTRODUCTION

In the last few years, and with the development of new image processing fields and advances in the deep learning have increased the importance of digital image recommendation systems where these systems help users in retrieving digital images in the visual and semantic form [1]. In many fields that deal with digital images, they can enterprise from the recommendation systems in a way that consistent with the requirements through gathering images relative to the needs of the users [2]. Traditional recommendation systems like content-based filtering or collaborative filtering encounter some difficulties when dealing with analyzing the contents of video segments or images, particularly in understanding visual semantics and determining consistence similarity measures [3]. Semantic segmentation tries to classify the image's pixels into determined features, most deep learning algorithms depend on modern image recommendation systems on CNN for extracting global features from a determined image, although these algorithms give good results but often, they deal with the image in overall as a one feature, without distinguishing between background and foreground components [4]. For achieving a succeeded image recommendation system, there are many of procedures that can be used to construct the required system, one of these procedures is using the semantic segmentation concept which is the process of connecting each pixel in the determined image with a relative label, while the computation of similarity can produce in accurate background information and this may lead to in accurate recommendations in the images that contain multiple objects [5]. There are many of studies use the neural networks that are designed to work with many of 2D forms for computing the similarity between images without understanding and analysis the objects in the determined image, because of the absence of the concept of the understanding level of the object [6]. Semantic segmentation is considered as

an important factor in many application fields especially in the field of image processing which classifies the image pixels into meaningful objects [7]. For a deep understanding of the image details, and for more accuracy, the semantic segmentation gives high performance in the analysis of the image contents [8]. Semantic segmentation provides the concept of understanding the whole scene in the image using the idea of forged pixel processing [9]. For accurate analysis, the semantic segmentation has become an interest research focus, it enables constitution identification for each object in the image [10].

2. GAP STATEMENT AND OBJECTIVES

The traditional image recommendation systems are depending on the concept of global feature extraction that is obtained from the CNN, while these approaches fail to distinguish between multi semantic entities or objects in the same digital image, so this leads to decreasing the recommendation accuracy, also the lack of semantic analysis prior to measurement of similarity, and many methods have been compute thus similarity depending on overall representation of the image like extracting one global feature vector for the image without isolating between the contained objects in the image and also this may leads to inaccurate recommendations, so for processing this gap the contribution of the proposed method is processing the semantic segmentation firstly, then computing the feature similarity, by segmenting the image into meaningful object regions then extracting isolating features for the determined objects, then the system will enhance the matching range between the recommendation query and the output of the semantic level.

3. METHOD

The most important step in this paper is applying a semantic segmentation for determining the features that contained in the determined image in order to recommend the relative images. First step includes grouping (*cars, persons and trees*) images from multiple online and real-world datasets, the grouped raw data that can be processed by the normalization have been processed using Min-Max normalization through scaling pixel values to the range [0–1] based on both the minimum and maximum values of each image. Using both Min and Max ensure that the smallest pixel value becomes 0, the largest becomes 1, and all other values are scaled proportionally, providing consistent input for the machine learning models. The second step is a feature map extraction. VGG-16 net is consisting of 16 layers where this map is considered as a resulted middle output from the convolution layer inside the VGG-16 net [11]. We applied data augmentation (rotation, cropping, flipping, zooming and scaling) processes on the original 230 image data set and we got 1380 image and this is considered as our new dataset. The result from the VGG-16 net is the feature map which encrypt low and high level of the determined image [12]. We just used 13 convolution layers in order to extract the feature map and we ignore the remaining 3 layers as explained in the following formula:

$$\beta = H * W * \alpha \quad (1)$$

where: H: is the 2-D map height, W: is the width, α : is the number of channels (filters) where every channel represents a feature or property like (edge, angle or texture etc...).

In the Vgg16 net, the convolution layer has 3*3 filter with stride equals to 1 and can be represented using the following formula:

$$C^{(l)} = f(W^{(l)} * O^{(l-1)} + b^{(l)}) \quad (2)$$

Where: $O^{(l-1)}$: the output of the previous layer, $W^{(l)}$: the input to the next layer, $C^{(l)}$: output (feature map) and $b^{(l)}$: the bias.

The images' pixels have been down sampled and reduced based on the frequency of the samples and where the image processing is applied [13]. For features that compressing and resetting the Global Average Pooling are considered one of the best procedures for obtaining one scalar value [14]. We used the Global Average Pooling (GAP) for down sampling the feature map as explained in the following formulas:

$$X \in H * W * C$$

Where: C,R: represent the feature map dimensions and C: no. of channels(filters).

$$Y_c = \frac{1}{H \times W} \cdot \sum_{k=1}^H \sum_{l=1}^W X_{klc}$$

$$Y = [y_1, y_2, \dots, y_c] \quad (3)$$

As we demonstrated, the output of the previous step is a 1-D feature vector. The SVM classifier employed in the regression and classification processes according to its abilities like segments the determined image into a group of sightings through placing good hyperplane for disjointing between the determined features in the image also provides a balance between both the small samples and model reliability [15] [16]. Once the features have been extracted from the image, they are considered as an input to the SVM classifier and the classifier converts these inputs into feature space [17]. In the proposed work, we got Semantic segmentation by extracting convolutional feature maps from the updated VGG-16 net, now we proposed that each spatial position in the feature map corresponds to a corresponding field in the original image. then these feature maps will be reconstructed where each pixel or may also called spatial unit will be represented by a feature vector, then these feature vectors will be classified using the Support Vector Machine (SVM) classifier in order to produce pixel-wise semantic labels, the derived classification map will be resized in order to match the used original image resolution using the bilinear interpolation. The use of SVM to determine object in the image demonstrates the model's potential applications in many fields [18]. Now first we built the decision boundaries between the contained features using the formula:

$$f(x) = w^T \cdot x + b \quad (4)$$

where w : weights vector, x : feature vector(inputs) and b : bias.

The feature then is determined using the following formula:

$$z = \text{sign}(f(x)) \quad (5)$$

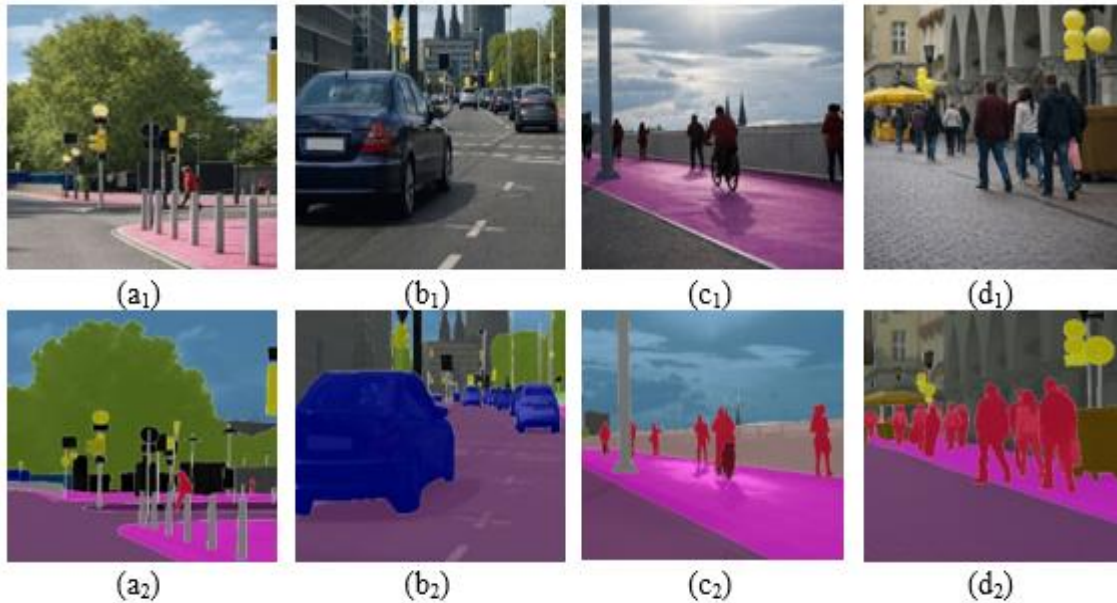


Figure 1. Original images with foreground semantic segmentation mask

Now for applying the recommended images, we need to determine the features similarity between the query image and the candidates, this achieved using the Pearson Correlation Coefficient using the following formula:

$$Pxy = \frac{\sum_{i=1}^k (xi - \bar{x}) - (yi - \bar{y})}{\sqrt{\sum_{i=1}^k (xi - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^k (yi - \bar{y})^2}} \quad (6)$$

Where k : sample size, x_i, y_i : individual features index with i, \bar{x}, \bar{y} : sample mean as displayed in figure2. We note that the recommended (candidate images) differ in the appearing percentage of the feature vector (determined label), so we used the Spatial Coverage formula to determine the appearance percentage in the candidate images

$$\sigma = \frac{\sum_{i=1}^C \sum_{j=1}^R 1(y_{ij}=c)}{C \times R} \quad (7)$$

where: $C.R$: represent the feature map dimensions, c : object feature (*Car, Person, Tree*), 1: Pointer function (1 if the pixel belongs to the determined object).

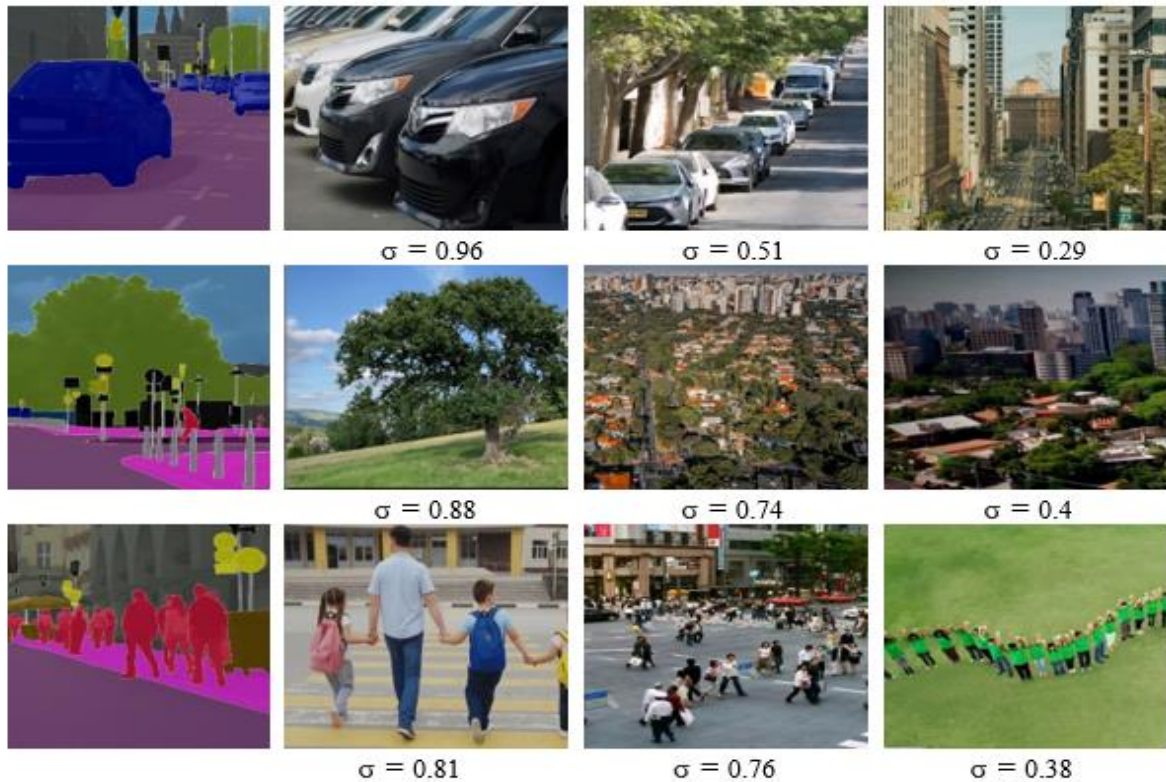


Figure 2. Original and recommended (candidate) images based on Pearson Correlation Coefficient and spatial coverage range between $[\sigma = 0.29-0.96]$

4. RESULTS AND DISCUSSION

4.1. RESULTS

Now the final step is estimating the proposed work through constructing a confusion matrix. The confusion matrix provides for each class the situation of accuracy predictions that has been made by the determined classifier and assess the efficiency of that classifier [19]. We used 1380 images as a sample and using the 70/30 split, the training set was consisting of 966 images and the testing set was 414 images.

		<u>Prediction</u>		
		Car (C)	Person (P)	Tree (T)
<u>Actual</u>	Car (C)	138		3
	Person (P)		135	5
	Tree (T)	2	7	124

Figure 3. Confusion matrix of the proposed method

Table 1. Performance metrics for evaluating proposed method using SVM classifier

Feature	C	P	T	Average
Precision	0.98	0.97	0.95	0.96
Accuracy	0.98	0.95	0.93	0.95
Recall	0.97	0.96	0.93	0.95
F1-Score	0.97	0.95	0.92	0.94

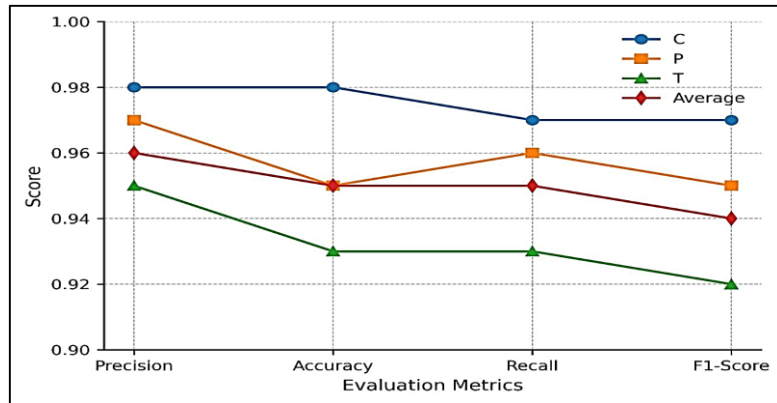


Figure 4. Graphical representation for the proposed model evaluation

4.2. DISCUSSION

The paper attempts to design a model for digital image recommendation system and after applying the method, we evaluate the obtained results, which have been derived from the metrics with another machine learning algorithms, we found that the proposed method got high evaluation, where the accuracy of the designed model was 0.95 as displayed in table1, the accuracy also was the highest among other deep learning algorithms with 0.95 as displayed in table2.

Table 2. Performance comparison and evaluation

Alg.	Proposed	LR	RF	KNN
Precision	0.96	0.94	0.93	0.96
Accuracy	0.95	0.93	0.95	0.94
Recall	0.95	0.93	0.91	0.94
F1-Score	0.94	0.91	0.92	0.92

Table 3. Evaluation of the proposed model with deep learning models

Model	Precision	Accuracy	Recall	F1-Score
Proposed	0.96	0.95	0.95	0.94
U-Net	0.93	0.92	0.92	0.90
Mask R-CNN	0.94	0.94	0.93	0.91

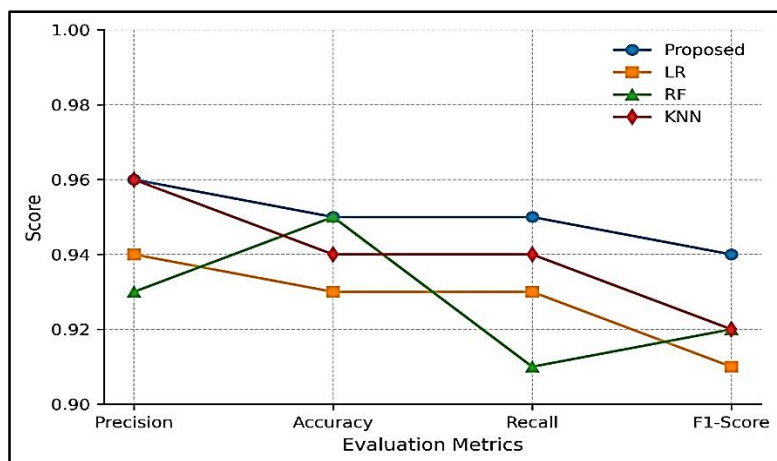


Figure.5 Evaluation of the proposed model with the machine learning models

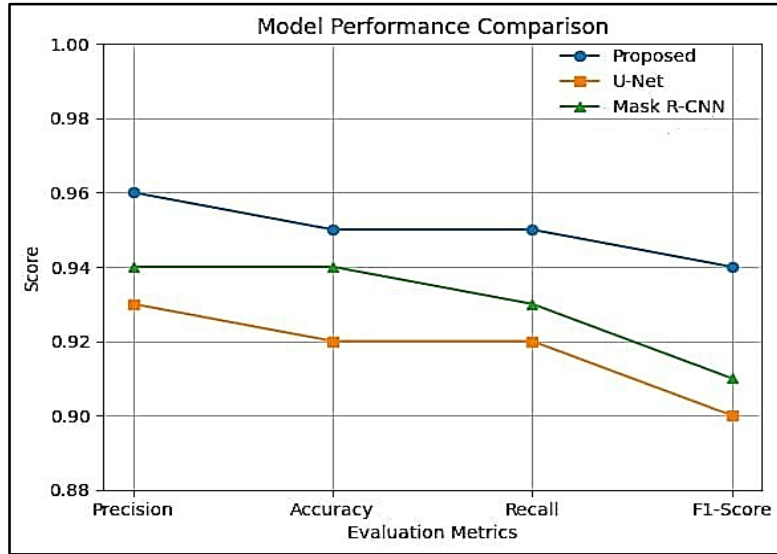


Figure.6 Evaluation of the proposed model with deep learning models

From table 3, we noticed that the accuracy satisfied the highest value among other deep learning algorithms as displayed in table 3.

Also, from figure2, we noticed that the Pearson measures the similarity of feature level and the spatial coverage (σ) reflects the structural similarity achieved by the user. Our analysis indicates that high efficiency recommendations are satisfied when both Pearson correlation and spatial coverage values are high.

5. CONCLUSION

In this research, we designed a digital image recommendation system based on the concept of semantic segmentation, we used a Support Vector Machine SVM classifier in order to extract the feature labels vectors for the original (query) and the recommended (candidate) images and depending on the comparison with other machine learning algorithms like LR,RF and KNN as displayed in table 2, we noticed that the proposed method has been achieved high precision and accuracy between recommending images and the query one, a;so we compared the proposed work with other deep learning algorithm like U-Net and R-CNN and we noticed that the obtained results from the proposed work were equivalent to the results of these algorithms and in some cases the proposed work got higher performance. The proposed work can be merged with other algorithms for extracting other features. The proposed system has some limitations. First, low recommendation quality when dealing with overlapping objects, also high brightness conditions may affect the recommendation quality, another limitation is the visual noise may lead to inaccurate recommendations. The contribution of the proposed method is processing the semantic segmentation firstly, then computing the feature similarity, by segmenting the image into meaningful object regions then extracting isolating features for the determined objects, then the system will enhance the matching range between the recommendation query and the output of the semantic level.

ACKNOWLEDGEMENT

The author would like to express his thanks for the assistance that is provided by the college of Computer Science and Mathematics/ Computer Science department/ University of Mosul during the duration of this work.

REFERENCES

- [1] Shefali Gupta, Meenu Dave, A recommendation system: Trends and Future, International journal of Engineering and Advanced Technology, (IJEAT), Blue Eyes Intelligence Engineering & Sciences Publication, Vol. 8, Issue 6S3, pp. 1361-1364, 2019, ISSN: 2249-8958.
- [2] He, X.; Liu, Q.; Jung, S. The Impact of Recommendation System on User Satisfaction: A Moderated Mediation Approach. J. Theor. Appl. Electron. Commer. Res. 2024, Vol.19, pp.448-466, DOI: 10.3390/jtaer19010024.
- [3] Vicky Rolanda, Teddy Surya Gunawan and Wanayumini, Content-Based Filtering Recommendation System using Categories Search Engine, IJRVOAS, 2023, Vol. 2, No. 4, Special issue- INCOSTIG, pp. 120-125, DOI: 10.53893.
- [4] Krzysztof Wolk, Marek S. Tatar, A Review of Semantic Segmentation and Instance Segmentation Techniques in Forestry using LiDAR and Imagery Data, MDPI Electronics, 2024, Vol. 13, No. 20, 4139, DOI: 10.3390/electronics13204139.

- [5] Aakanksha, Shanu Sharma, Semantic segmentation: Systematic Analysis from state-of-the-Art Techniques to Advance Deep Networks, Journal of Information Technology Research, 2022, Vol. 15, Issue 1, pp.1-28, DOI: 10.4018/JITR.299388.
- [6] Venkata S., Bonthu K, Lingamallu R., Bisht D, Chhabra S, " Content Based Image Retrieval using Deep Learning", AIP Conference proc., 2025, 3157, 030003, DOI: 10.1063/5.0261658.
- [7] M.K. Kar, M.K. Nath, and D.R.Neog, A Review on Progress in Semantic Image Segmentation and Its Application to Medical Images, SN Computer Science, 2021, Vol.2, No.5, DOI: 10.1007/s42979-021-00784-5.
- [8] Yutong Li, The Application of Semantic Segmentation on 2D images, Highlights in Science, Engineering and Technology, IFMPT, 2023, Vol.31, pp.88-96.
- [9] Abhishek Thakur, Shamneeh Sharma and Tushar Sharma, "Design of Semantic Segmentation Algorithm to classify Forged Pixels", 12th IEEE International Conference on Communication System and Networks Technologies, 2023, pp.409-413, DOI: 10.1109/csnt.2023.73.
- [10] Muqing Li, Ziyi Zhu, and others, "Research on Image Classification and Semantic Segmentation Model based on Convolutional Neural Network", Journal of Computing and Electronic Information Management, 2024, Vol. 12, No. 3, pp. 95-100, DOI: 10.54097/qg7hakzu.
- [11] Khaleghian S, Ullah H, and others, "Sea Ice Classification of SAR Imagery based on Convolution Neural Network", Remote Sens., 2021, Vol. 13, 1734, pp.1-20, DOI: 10.3390/rs13091734.
- [12] Mogan J N, Lee C P, and others, "VGG16-MLP:Gait Recognition with Fine-Tuned VGG-16 and Multilayer Perception", Appl. Sci., 2022, Vol. 12, 7639, pp.1-12, DOI: 10.3390/app12157639.
- [13] Pavel L., Nataliya S, Nikolay N, Maxim B, "High-Speed Wavelet Image Processing using Winograd method with Downsampling", MDPI Mathematics, 2023, Vol. 11, No. 22, 4644, DOI: 10.3390/math11224644.
- [14] Tae-Wan Kim, Keun-Chang Kwak, "Pixel-Level Fuzzy Rule Attention Maps for Interpretable MRI Classification", 2025, Vol. 17, 2187, DOI: 10.3390/sym17122187.
- [15] Pandey S., Mishra P., " An efficient image classification using SVM, CNN", International journal of Computer Applications ICAIDSC, 2025, Vol.1, pp. 5-10, DOI: 10.5120/icaidsc202403.
- [16] Jonatha Sousa Pimentel, Raydonal Ospina, Anderson Ara, "A Novel Fusion Support Vector Machine integrating weak and sphere models for classification challenges with massive data", Decision Analytics Journal, 2024, Vol.11, DOI: 10.1016/j.dajour.2024.100457.
- [17] Guido R, Ferrisi S, Lofaro D and Conforte D, "An Overview on the advancement of Support Vector Machine in Healthcare Applications: A Review", Information, 2024, Vol. 15, No. 4, 235, DOI: 10.3390/info15040235.
- [18] James G, Umoren I., Ekong A., "Analysis of Support Vector Machine and Random Forest models for classification of the impact of technodress in covid and post-covid era", Journal of the Nigerian Society of Physical Sciences, 2024, Vol. 6, No. 3, DOI: 10.46481/jnsps.2024.2102.
- [19] Fahmy Amin, Moahmoud, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial", Journal of Engineering Research, 2022, Vol. 6, No. 5, Article1, pp.1-12.

BIOGRAPHIES OF AUTHORS



Omar Muayad Abdullah, received a Bachelor of Computer Science Degree from the Department of Computer Science from the University of Mosul in 2001 and a Master of Science Degree from the Department of Computer Science from the University of Mosul in 2004 and a PhD degree in the specialty Multimedia from the Odessa National Polytechnic University (ONPU) from Ukraine in 2016. Omar Muayad Abdullah is a Lecturer at the Faculty of Computer Science and Mathematics, Mosul University, Iraq. Researching about being in Multimedia processing, Computer vision, DSP, Virtual and Augmented reality programming. he can be contacted at email: omaraldewachy@uomosul.edu.iq.