Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 3, June 2022, pp. 1343~1349
ISSN: 2302-9285, DOI: 10.11591/eei.v11i3.3730

Convolutional neural network for color images classification

Nora Ahmed Mohammed[1], Mohammed Hamzah Abed[2], Alaa Taima Albu-Salih[2]
[1]Department of Mechanical Engineering, Faculty of Engineering, Al-Qadisiyah University, Al Diwaniyah, Iraq
[2]Department of Computer Science, Faculty of Computer Science and Information Technology, Al-Qadisiyah University, Al Diwaniyah, Iraq

Article Info

Article history:

Received Feb 27, 2022
Revised Apr 29, 2022
Accepted May 17, 2022

Keywords:

Classification
CNN
Color images
Deep learning
Features extraction

ABSTRACT

Artificial intelligent and application of computer vision are an exciting topic in last few years, and its key for many real time applications like video summarization, image retrieval and image classifications. One of the most trend method in deep learning is a convolutional neural network, used for many applications of image processing and computer vision. In this work convolutional neural networks CNN model proposed for color image classification, the proposed model build using MATLAB tools of deep learning. In addition, the suggested model tested on three different datasets, with different size. The proposed model achieved highest result of accuracy, precision and sensitivity with the largest dataset and it was as following: accuracy is 0.9924, precision is 0.9947 and sensitivity is 0.9931, compare with other models.

Corresponding Author:

Mohammed Hamzah Abed
Department of Computer Science, Faculty of Computer Science and Information Technology
Al-Qadisiyah University
XVVR+WFR, Al Diwaniyah, Iraq
Email: mohammed.abed@qu.edu.iq

1. INTRODUCTION

Image feature extraction and classification has always been a fundamental research area in computer vision [1]. As the advanced and trained neural network has the ability to extract features and characteristics of existing images as well as classification of the extracted objects. It is also considered a very important scientific and research branch in the field of neural networks. Convolutional neural networks (CNN) have the property that the features of each layer are activated by the local area of the preceding layer via a convolution kernel that shares weights. Because of this property, CNN are better suited for image object detection and expression than other neural network methods [2].

Motivation and contribution; deep learning (DL) and deep CNN (DCNN) have significantly improved performance over state-of-the-art areas. Whereas CNN models can collect higher-level information from the post-level convolution layer in addition to extracting detailed texture data from pre-level convolution networks. In this field, several researchers have introduced pre-trained DCNN models, such as ResNet [3], VGG [4], AlexNet [5], the YOLO model, and GoogleNet [6]. On the other hand, another set of enhancements focused on densely training and testing the networks across the entire image and at several scales [7]. The proposed model of color images classification based on CNN was suggested. And the model divided into two part: first is a preprocessing to enhance the appearance of the image and second part: features extraction and classification based on CNN model.

Paper layout; the rest of this paper is organized as a following, related work and literature survey are described in in section 2, section 3 shows the materials and methods are used in this paper. In section 4, the proposed model and mathematical formula of the suggested model are represented and discussed, experimental result and discussion discuss and shows in section 5, finally conclusion of the research work will be presented in section 6.

2. RELATED WORK

CNN have proven to be effective in object detection and classification [3], the technique of deep learning algorithms based on the data input in the form of video or images, add weights and bias to various features of the image and then distinguish them by localizing every object in the bounding box from another object. Researchers seek in this ambit to take advantage of the spatial information contained within the pixels of an image. As a result, they are built on the concept of discrete convolution. There are numerous network layers in the CNN paradigm, that work to optimize the model's fault tolerance through several functional strategies consist of convolution, pooling, and dropout layer [8]. As it's recognized, convolution and pooling are two of these procedures that are required in current CNN models. Zhiqiang and Jun [9] showed the drawbacks and disadvantages in the features that are made manually by suggesting models extract them and also developing a detection algorithm, but there were also some defects like low resolution and occlusion.

MobileNet employed separable convolution to lower computing expenses while attempting to strike a compromise between accuracy and speed [10]. While deep ResNet, which debuted in 2015 with its residual function, enabled deeper network architectures to contain hundreds of layers [8]. Lin *et al.* [11] used CIA, Morph, and CACD2000 database and apply evolutionary-fuzzy-integral-based CNN (EFI-CNNs) for age and gender categorization depend on the fuzzy integral theory of the faces. In comparison to previous technologies such as GoogLeNet, CNN's AlexNet, and VGG16, this method has improved accuracy.

As well as in regards of X-rays and their applications in [12]. They evaluated the application of CNN comprehensively in classification and detection tasks within the X-ray luggage images. A comparison of CNN and classic (bag of visual words) BoVW model based on handcrafted features are employed in research using deep CNN with transfer learning to overcome limited object data availability. Additionally, they finally train a CNN support classifier for vector machine (SVM). Based on AlexNet characteristics, they attain an accuracy of 0.994. as well as the researcher in [13] proposed model based on pretrained CNN and compare the accuracy that achieved between them, the model tested on architectural heritage images dataset.

Research by Zhao *et al*. [14] deep learning was used to develop a system for fine-grained object classification and semantic segmentation. This approach differentiates between subordinate-level groups, such as dog breeds and bird species. On the ImageNet dataset, they achieved a 3.57 percent error rate. According to Fang [15] the technique to handle image classification difficulties has been proposed. They gain a better understanding of deep learning by analyzing misclassified situations of emotions and facial recognition in their work.

Kadhim and Abed [16] proposed model for satellite image classification based on three different pre-trained CNN. The suggested work tested on SAT4, SAT6 and UC merced and achieve a good result and accuracy 95.8, 94.1 and 98. Out from the above mentioned, most of the above studies focusing on pretrained CNN, for image classification, unlike the proposed solution which design CNN suitable for image classification mission with less number of layers, to decrease training time and achieve good result.

3. MATERIALS AND METHODS

3.1. Materials

The experiment of the proposed work deploys three different datasets: UC merced land, architectural heritage elements dataset and animal image dataset (Dog, Cat and Panda):

a. UC Merced land, this dataset consists of 21 classes land images, each class contain 100 images with 256×256 dimension, these images were collected from large dataset images from the USGS national map urban area imagery collection [17];
b. architectural heritage elements dataset, the dataset was published with two versions with 10 classes, the complete version contains of 10235 labelled images;
c. animal image dataset, this dataset consists of three different classes (Dog, Cat and Panda) collected from Kaggle. The Figure 1 show samples of dataset that used to test the proposed model.

3.2. Methods

Machine learning is a study of giving the computer ability to learning without any human interaction based on set of data known as a training dataset to predicate a new data. Machine learning basically classified into three major types based on learning method. One of the most famous methods of neural network in deep learning is a CNN. CNN is specially designed for image classification and recognition, it contains many layers of NN, for features extraction and preprocessing the data to predicate which class data belong to Figure 2 show the basic architecture of CNN.

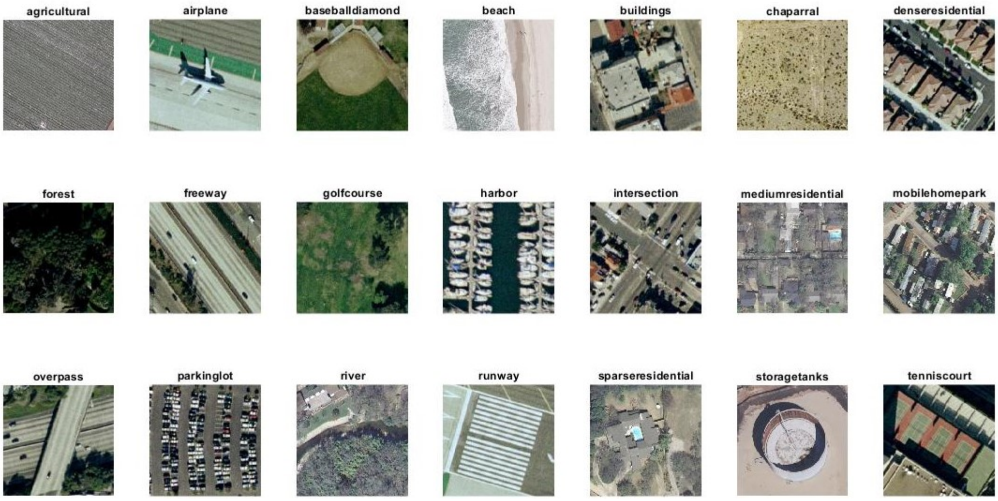

Figure 1. sample of UC merced dataset

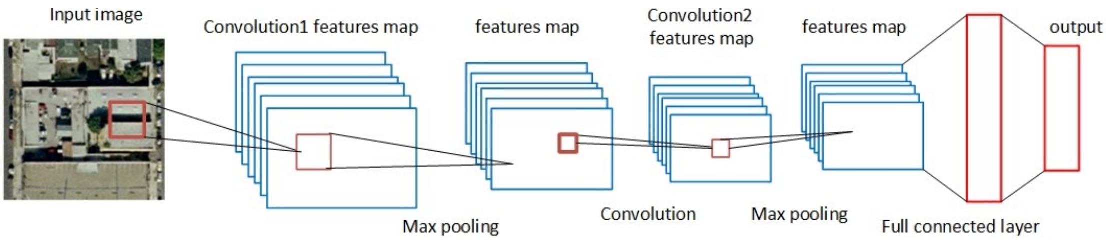

Figure 2. Basic architecture of CNN

A CNN contains of three types of layer, these layers constructed by CNN are:
- Convolutional layer is the layer which responsible of features extractions from input matrix, the earlier convolutional layer extracts the low features of images like edge, and line. As well as deeper level of it used to extract the deep features of input images or matrix.
- Pooling layer is a second type of NN layer, used to reduces the input image matrix space using mask size 2×2 or 3×3. The efficiency of pooling layer to reduce dimension of features help the CNN to speed up the computation time.
- Fully connected layer is used in the last of the CNN architecture to combine all the features together.

4. FEATURES EXTRACTION BASED ON CNN

Image classification based on CNN are proposed and evaluated based on three different datasets.

4.1. Problem statement

The growing need many efficiently methods for analysis and understanding the images as one application of computer vision in medical system as well as in robotics. One of the most important and primary problems in image processing is a classification, in which classification refers to the task of labeling an image based on their features.

4.2. CNN architecture of proposed model

Suggested CNN architecture contain a multiple convolutional and pooling layers that ended by fully connected layer [18], [19]. Each convolutional layer has its own weights across input, in which each input data entered for the current layer comes from subset of features from the previous layer [20]. Same concept with pooling layer the conduct output of the convolutional layer, it's to minimize the set of features to avoid the complexity cost of features data that moving to the depth [21], [22]. We can see, images feature representations and extracted by each layers is consider as a local features, therefore some fully connected layers introduced in sequence to find the global features which depend on the output of the previous layers fully-connected layers have a complete connection to all the activities in previous connected as a hierarchical structure, that give the CNN ability to extract more discriminative feature representations from the lower layer to the higher layer. The Figure 3 formulate the proposed CNN layers.

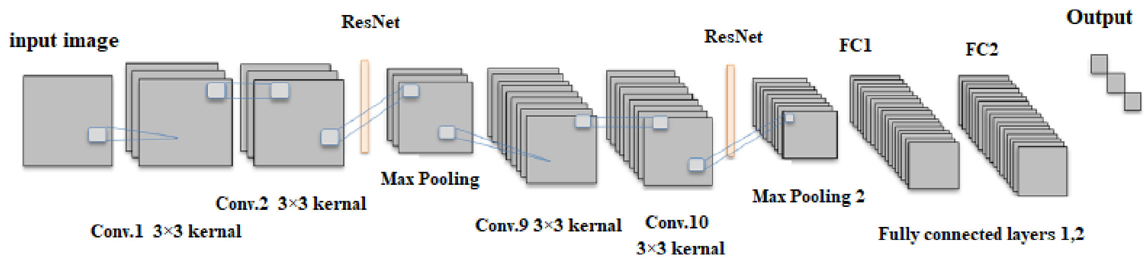

Figure 3. Proposed CNN architecture

Features extracted from convolutional layer, in this suggested model deep color images features can be extracted directly from convolutional and fully connected layers. To increase the performance of color images classification some of preprocessing phase are applied before CNN features extraction phase. It contains color normalization and enhance the appearance of each band spritely [20]. In the convolution layer the entire images contain three sub matrix (R, G and B) called color layers [23], [24]. The convolution matrix to do the filtering for each channel. The mechanism of convolution is, each area of 3*3, doing filtering for each channel 3*3 respectively, then add them together to get final numbers [25]. The proposed CNN architecture contains two fully connected layers FC1 and FC2 to extract the images features from enhance images dataset then from classifications layer and output make decision of classification result. The following algorithm explain the steps of color image classification using proposed architecture of CNN.

Algorithm of color image classification CNN

Input image matrix

Output classification result

Algorithm Steps

While iteration < = max_number of dataset images do:

1. Enhance the images appearance using Adaptive histogram equalization
2. Feature’s extraction and classification
3. Read image

   While iteration of layer<= 3 do

   a. Read image blocks

   b. Appy convolution layer for all image pixels

   i. Sliding mask features and image match path

   ii. Multiply each input image by mask of features pixel Features_I(x,y)= I(x,y)*mask(3,3)

   iii. Average_features= $\sum$ features_ I(x, y)/ no of

   c. Pooling layer

   i. Apply [2*2] max pooling on convolution features

4. Fully connected layers FC1 and FC2

5. **Output** classification result

5. EXPERIMENTAL RESULT AND ANALYSIS

We evaluate the performance of the color image classification is used three datasets have been mentioned above and tested in the current work. UC merced land, architectural heritage elements dataset and animal image dataset, each one has multi classes. These datasets divided into 70% for training and 30 % for testing. The proposed model consists of two main phases: first phase focusing on preprocessing the images of dataset and color normalization, help the features extraction phase to achieved a good result. The second phase is features extraction and classification based on CNN. Table 1 shows the configuration of CNN’s architecture. The proposed CNN was built and evaluated using the Matlab R2020 and using deep network designer tool for CNN’s architecture Figure 4 shows the training and loss function of the dataset.

Table 1. Configuration of proposed CNN

Layer number	Name	Activations	Properties
1	Input image	128*128*3 AHE data set 256*256*3 UC Merced land 256*256*3 Animal images	Color images Different dimensions
2	conv1_1	Filter size [3,3]	The convolution mask size is 3 *3 padding [0 0 0 0] and stride [1 1]
3	Relu1	Run length activation function	
4	Pool max1	Pooling size [3,3]	The pooling is max value in window2 *2 padding [0 0 0 0] and stride [2 2]
5	Conv2_1	Filter size [3,3]	The convolution mask size is 3 *3 padding [0 0 0 0] and stride [1 1]
6	Relu2	Run length activation function	-
7	Pool max2	Pooling size [3,3]	The pooling is max value in window2 *2 padding [0 0 0 0] and stride [2 2]
8	Conv3_1	Filter size [3,3]	The convolution mask size is 3 *3 padding [0 0 0 0] and stride [1 1]
9	Relu3	Run length activation function	-
10	Pool max3	Pooling size [3,3]	The pooling is max value in window2 *2 padding [0 0 0 0] and stride [2 2]
11	Fully connected layer	FC1	Features layer
12	Fully connected layer	FC2	Features layer
13	Classification layer		Output layer

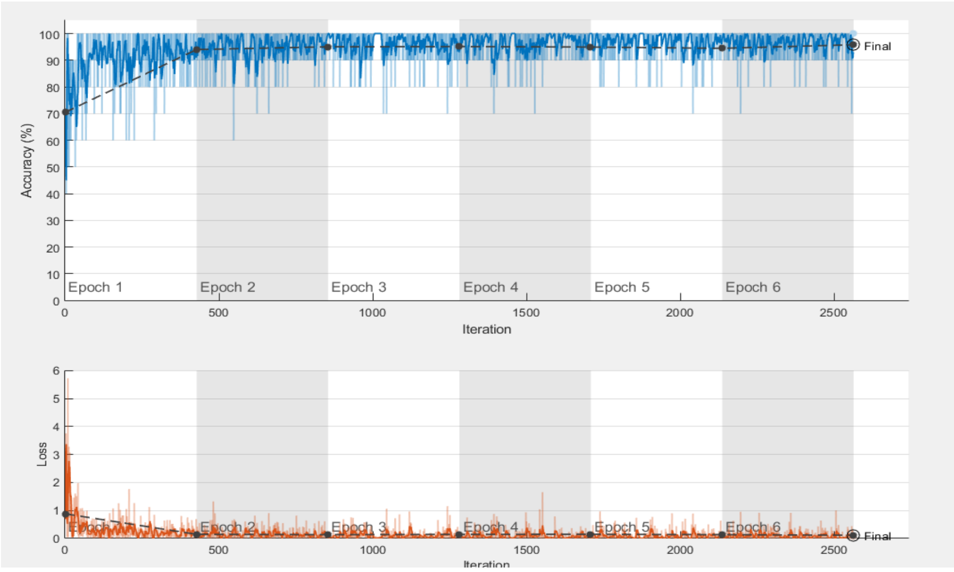

Figure 4. Training of proposed CNN

6. RESULT

Three metrics have been calculated to evaluate the performance proposed model, some of analysis were used to measure the result based on standard metrics, to evaluate the accuracy performance of the suggest model. The measurement that calculated is an accuracy, precision and sensitivity of a method determines how correct class are predicted, (1), (2) and (3) shows the accuracy calculation, precision and sensitivity.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

where TP is true postive, TN is true negative, FP is false positive and FN is false negative

Table 2 shows the summary of performance of color image classification based on CNN as well as it is contain value of TP, TN, FP, FN, accuracy, sensitivity and precision for all classes of dataset. The accuracy ratio in second dataset was the highest between other datasets, same as the precision and sensitivity ratio. The main reasons of this success ratio were because the number of images in training and testing is larger than other dataset, and CNN model in general need large of dataset to work well. Table 3 show the comparison of color image classification.

Table 2. Summary of color image classification performance

Dataset	TP	TN	FP	FN	Accuracy	precision	sensitivity
UC Merced land	287	331	4	8	0.9810	0.9863	0.9729
Architectural heritage elements dataset	6410	3747	44	34	0.9924	0.9947	0.9932
Animal image	166	154	8	4	0.9639	0.9540	0.9765

Table 3. Comparison of color image classification accuracy performance

Algorithm	Classification accuracy of UC Merced land	Classification accuracy of architectural heritage elements dataset	Classification accuracy of animal image
Pre-trained GoogleNet	0.97	0.9547	0.95
Pre-trained ResNet18	0.978	0.9557	0.96
[13]	-	0.93	-
[16]	0.98	-	-
Proposed model	0.9810	0.9924	0.9639

7. DISCUSSION

As shown in Table 3, the accuracy ratio of three different dataset using pretrained model GoogleNet and ResNet18 was between 0.95 to 0.97 using GoogleNet, and 0.9557 to 0.978 using ResNet18 for UC Merced land, architectural heritage elements dataset and animal image respectively. In addition, the researcher in [13] tested the proposed model only using one dataset architectural heritage elements dataset and achieved accuracy result was 0.93. as well as in [16] the methods applied and tested on UC Merced land and achieved accuracy result 0.98. all the above-mentioned methods focusing on pretrained model, unlike the proposed model used the design tools in Matlab to complete the CNN design of each layers.

8. CONCLUSION

This work presents a CNN model for color images classification. The suggested model of CNN consists of 13 layers. And have been tested and evaluated on three public and very well-known datasets: UC Merced land, architectural heritage elements dataset and animal image. The CNN make classification using deeper features which extracted from entire color images. The performance of the proposed model has been tasted based on three metrics accuracy, precision and sensitivity. The proposed model achieved a high accuracy 0.9924 of architectural heritage elements dataset, 0.9810 for UC Merced land and 0.9639 of animal image. As well as a precision and sensitivity were calculated to evaluate the performance of proposed model.

REFERENCES

[1] P. Shamsolmoali, D. K. Jain, M. Zareapoor, J. Yang, and M. A. Alam, “High-dimensional multimedia classification using deep CNN and extended residual units,” *Multimedia Tools and Applications,* vol. 78, no. 17, pp. 23867-23882, 2019, doi: 10.1007/s11042-018-6146-7.

[2] J. Llamas, P. M. Lerones, R. Medina, E. Zalama, and J. G. G. Bermejo, “Classification of architectural heritage images using deep learning techniques,” *Applied Sciences (Switzerland)*, vol. 7, no. 10, pp. 1-25, 2017, doi: 10.3390/app7100992.

[3] Z. Wu, C. Shen, and A. V. D. Hengel, “Wider or Deeper: Revisiting the ResNet Model for Visual Recognition,” *Pattern Recognition*, vol. 90, pp. 119-133, 2019, doi: 10.1016/j.patcog.2019.01.006.

[4] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks For Large-Scale Image Recognition,” *arXiv preprint arXiv*, Cornell Universuty, vol. 75, no. 6, pp. 1-14, 2018.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017, doi: 10.1145/3065386.

[6] C. Szegedy *et al*. “Going Deeper with Convolutions,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 91, no. 8, pp. 1-9, 2016.

[7] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks,” *arXiv preprint arXiv:1312.6229*, Cornell University, 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[9] W. Zhiqiang and L. Jun, “A review of object detection based on convolutional neural network,” *36th Chinese Control Conference (CCC)*, 2017, pp. 11104-11109, doi: 10.23919/ChiCC.2017.8029130.

[10] R. O. Ekoputris, “MobileNet: Deteksi Objek pada Platform Mobile,” *Medium*, 2018.

[11] C. J. Lin, C. H. Lin, C. C. Sun, and S. H. Wang, “Evolutionary-fuzzy-integral-based convolutional neural networks for facial image classification,” *Electronics (Switzerland)*, vol. 8, no. 9, pp. 1-16, 2019, doi: 10.3390/electronics8090997.

[12] S. Akcay, M. E. Kundegorski, C. G. Willcocks and T. P. Breckon, “Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2203-2215, 2018, doi: 10.1109/TIFS.2018.2812196.

[13] M. H. Abed, M. Al-Asfoor, and Z. M. Hussain, "Architectural heritage images classification using deep learning with CNN," *CEUR Workshop Proceedings*, 2020, vol. 2602, pp. 1-12.
[14] B. Zhao, J. Feng, X. Wu, and S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *International Journal of Automation and Computing*, vol. 14, no. 2, pp. 119-135, 2017. doi: 10.1007/s11633-017-1053-3.
[15] X. Fang, "Understanding deep learning via backtracking and deconvolution," *Journal of Big Data*, vol. 4, no. 40, pp. 1-14, 2017, doi: 10.1186/s40537-017-0101-8.
[16] M. A. Kadhim and M. H. Abed, "Convolutional Neural Network for Satellite Image Classification," *Asian Conference on Intelligent Information and Database Systems*, vol. 830, no. 1, pp. 165-178, 2019, doi: 10.1007/978-3-030-14132-5_13.
[17] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270-279, doi: 10.1145/1869790.1869829.
[18] S. Fairuz, M. H. Habaebi and E. M. A. Elsheikh, "Pre-trained based CNN model to identify finger vein," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 3, pp. 855-862, 2019, doi: 10.11591/eei.v8i3.1505.
[19] M. S. N. Raju and B. S. Rao, "Colorectal multi-class image classification using deep learning models," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 195-200, 2022, doi: 10.11591/eei.v11i1.3299.
[20] W. Setiawan, M. I. Utoyo, and R. Rulaningtyas, "Reconfiguration layers of convolutional neural network for fundus patches classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 383-389, 2021, doi: 10.11591/eei.v10i1.1974.
[21] M. H. Abed, A. H. Alsaedi, A. D. Alfoudi, A. M. Otebolaku, and Y. S. Razooqi, "Palm Vein Identification Based on Hybrid Feature Selection Model," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 5, pp. 469-478, 2020.
[22] M. H. Abed, L. A. N. Muhamme, and S. H. Toman, "Diabetic Retinopathy Diagnosis based on Convolutional Neural Network," *arXiv preprint arXiv:2008.00148,* Cornell University*,* 2020, doi: 10.48550/arXiv.2008.00148.
[23] H. R. Mohammed and Z. M. Hussain, "Hybrid mamdani fuzzy rules and convolutional neural networks for analysis and identification of animal images," *Computation*, vol. 9, no. 3, pp. 1-17, 2021, doi: 10.3390/computation9030035.
[24] M. A. Obaid and W. M. Jasim, "Pre-convoluted neural networks for fashion classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 2, pp. 750-758, 2021, doi: 10.11591/eei.v10i2.2750.
[25] P. Wang, E. Fan and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognition Letters*, vol. 141, pp. 61-67, 2021, doi: 10.1016/j.patrec.2020.07.042.

BIOGRAPHIES OF AUTHORS

Nora Ahmed Mohammed is currently a researcher at the college of engineering, Al-Qadisiyah University, Iraq. Her main research interest is in the field of natural language processing and computational linguistics. She has worked on mining natural language text for information retrieval, relation extraction, and synonymy discovery. She received her Master degree in Computer Science and Engineering from Osmania University, India. She can be contacted at email: nora.mohammed@qu.edu.iq

Mohammed Hamzah Abed received the B.Sc degree in Computer Science from University of Al-Qadisiyah, Iraq 2008, M.Sc degree in Computer Science from B.A.M. University, India 2011. Currently, he work as an assistant Professor at the Department of Computer Science, University of Al-Qadisiyah. His main research is in the medical image processing and machine learning. His current work is medical image analysis based on deep learning. He can be contacted at email: mohammed.abed@qu.edu.iq.

Alaa Taima Albu-Salih received the B.Sc. degree in Computer Engineering from University of Al-Qadisiyah, Iraq in 2005, the M.Sc degree in Information Technology from Babylon University, Iraq in 2013 and the Ph.D. degree in Computer Engineering at Ferdowsi University of Mashhad, Mashhad, Iran in 2018. His research interests include wireless sensor networks, mobility models, routing, software-defined networking, internet of things, mathematical modeling, optimization, unmanned aerial vehicle and he have several publications in these fields. He can be contacted at email: alaa.taima@qu.edu.iq.