

Diabetic retinopathy grading system based on transfer learning

Eman Abdelmaksoud¹, Sherif Barakat¹ and Mohammed Elmogy^{2*}

Information Systems Department, Faculty of Computers and Information, Mansoura University, Egypt¹

Information Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura, PO 35516, Egypt²

Received: 11-December-2020; Revised: 24-January-2021; Accepted: 26-January-2021

©2021 Eman Abdelmaksoud et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Much effort is being made by the researchers in order to detect and diagnose diabetic retinopathy (DR) automatically and accurately. The disease is very dangerous as it can cause blindness suddenly if it is not continuously screened. Therefore, many computer-aided diagnosis (CAD) systems have been developed to diagnose the various DR grades. Recently, many CAD systems based on deep learning (DL) methods have been adopted to get deep learning merits in diagnosing the pathological abnormalities of DR disease. In this paper, we present a full based-DL CAD system, depending on multi-label classification. In the proposed DL CAD system, we present a customized EfficientNet model in order to diagnose the early and advanced grades of the DR disease based on transfer learning. Transfer learning is very useful in training small datasets. We utilized a multi-label Indian Diabetic Retinopathy Image Dataset (IDRiD) dataset. The experiments manifest that the proposed DL CAD system is robust, reliable, and deigns promising results in detecting and grading DR. The proposed system achieved accuracy (ACC) equals 86%, and the Dice similarity coefficient (DSC) equals 78.45%.

Keywords

Diabetic retinopathy (DR) grades, Deep learning (DL), Computer-aided diagnosis (CAD) systems, Transfer learning, EfficientNet.

1.Introduction

The retina is an essential part of the human eye. It covers the back of the eye. This critical part of the eye detects the lights and sends signals to the brain through the optic nerve (ON). The retina is susceptible to damage due to excess blood sugar caused by diabetes. Generally, diabetes damages the blood vessels (BV) all over the body, not only on the human eye. The retina damage starts when the sugar prohibits the tiny BV that goes with it, causing them to leak fluid or bleed. In this case, the eye begins to grow new weak BV that can leak or bleed easily as they do not work well. The caused damage of the retina is called diabetic retinopathy (DR). Anyone with diabetes can get DR, even with gestational diabetes that can develop during pregnancy [1]. DR, in its advanced stage, can cause blindness. Therefore, the patient with diabetes should constantly control the blood pressure and cholesterol levels by making continuous checking up and auditing.

The leakage and bleeding causes are appearing some DR lesions with different features on the retina and BVs, such as micro and macro aneurysms (MA), dot, blot, and flame hemorrhages (HM), exudates (EX), cotton wool spots (CWS), venous beading (VB) and neovascularization (NV). These lesions specify the DR grades, which are categorized as early or mild, moderate, advanced, or severe. All of these grades are called non-proliferative DR (NPDR). In NPDR, the treatment is possible. On the other hand, the severe grade on NPDR is an indicator of the difficult DR stages that is proliferative DR (PDR). When the patient reaches this category, the vision may be lost suddenly [2].

NPDR is determined by appearing MA, dot HM, scattered EX, CWS, and VB. On the other side, PDR is determined by appearing NV on the optic disc (OD) (NVDs) or elsewhere (NVEs), large blot HM (B-HM), and VB. MA occurs in the superficial of the retina as small red points. The EX is caused by leakage of the plasma. It looks like bright yellowish spots with sharp margins. It appears in the retina's outer layer. In many cases, HM, looks like the MA if

* Author for correspondence

it is a small B-HM or dot HM (D-HM) in the deepest retina's layer. The large flame HM (F-HM) is also called splinter HM (S-HM), is found in the superficial of the retina layer. CWS is a small white with pale and irregular margins. It is caused by nerve fiber swelling [3].

Among the eye imaging modalities, we utilized the most famous one, the fundus modality. It captures the anatomic eye structure, helping detect and diagnosing

DR. Fundus photography is useful in DR screening and documents the retinal disease timely. It includes different types, standard, wide-field, and stereoscopic. Besides, it is cheaper and useful for monitoring the DR progression over time [4, 5]. *Figure 1* shows the different DR signs of the retinal color fundus images and the four DR grades (mild, moderate, severe NPDR, and PDR).

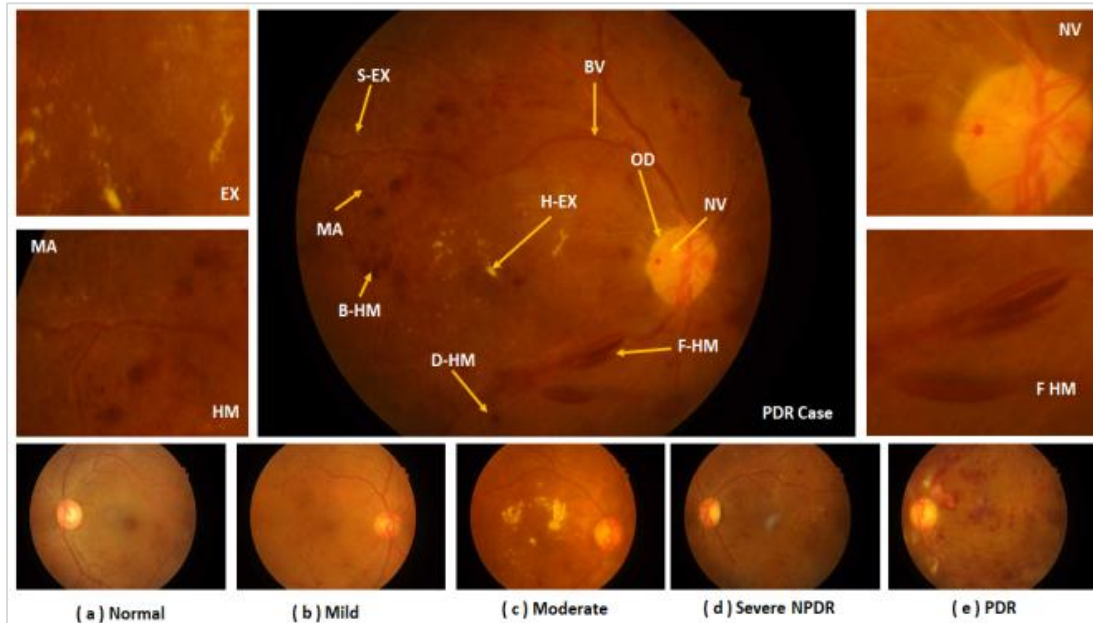


Figure 1 The different DR signs of the color fundus image (EX, MA, HM, and NV), (a) normal case, (b) mild case, (c) moderate case, (d) severe case, and (e) PDR case

Regular retinal screening for diabetes patients is still difficult, especially in poor developing countries. Retinal photography with grading and interpretation by ophthalmologists and trained graders are widely important for DR screening. However, the availability of retina experts and well-trained graders is a significant limitation. Even when they are available, there could be a time delay in graders submitting their DR grading and advice due to their busy schedule. This leads to delayed interpretation and loss to follow-up, miscommunication, and delay in DR severity management.

In addition to the aforementioned challenges, the detection and grading of DR disease are crucial, consumes much time, and burdens the ophthalmologists. Therefore, many systems have been developed to detect the DR grades. These systems may entirely depend on deep learning (DL)

such as [3, 6, 7] or conventional methods such as [8]. Others make only the classification phase by DL, such as [9]. Recently, various DL architectures have been developed, such as Restricted Boltzmann Machines (RBMs), deep neural network (DNN), and Convolutional Neural Networks (CNNs). CNN has achieved great success in many applications, such as medical image analysis and multi-label (ML) classification. That is because CNN can extract features more discriminative for large-scale datasets. We utilized the CNN architecture to classify the healthy and various DR grades from color fundus images.

The main reason for utilizing the DL returns to the various features of the color fundus images is very difficult to be extracted by handcrafted methods. These abnormalities or features are BV, H-EX, S-EX, B-HM, F-HM, D-HM, NV, CWS, VB, and others.

Moreover, there are many signs, as shown in *Figure 1*, take the same features. Small MAs are like D-HM, large MA is like a B-HM in color and shape. On the other hand, H-EXs and OD are the same in the intensity and shape features. HM, MA, and BV take the same color feature. Sometimes, CWS and EXs are the same in the color feature. According to the similarity of the features mentioned above, the traditional (handcrafted) features extraction methods such as scale-invariant feature transform (SIFT), histogram-oriented gradients (HOG), and bag-of-words histograms (BoW) are unsuitable for fundus images. It is very hard for these methods to extract effective features from DR color fundus image datasets, especially for ML ones. That is because fundus images are with small differences. These feature extraction methods are very effective and efficient for images with large differences in features [10].

Besides, handcrafted features are sensitive to fundus photography quality, such as noise, equipment, and exposure intensity. From another perspective, and in addition to the previously stated problems, the traditional classifiers such as SVM and KNN perform very well for small datasets but work poorly with large-scale datasets. Therefore, their generalization is limited. Thus, these methods are universal feature extraction methods, which may not be suitable for DR detection, a more challenging and specific task.

This paper presents a comprehensive CAD system based on a novel customized CNN model (EfficientNet-B0) [11]. To diagnose the health and DR grades from different color fundus images automatically without making handcrafted feature extraction or segmentation phases. It is noteworthy that deep fine-tuned CNNs are very useful in medical image analysis and even outperform the fully trained CNNs, especially in limited training set" [12, 13]. We customized the efficientNet-B0 by optimizing the hyperparameters to increase accuracy. We compared the proposed system and others to ensure the validity of the proposed CAD system. The proposed methodology comprises a series of contributions, which are listed as follows:

- We solved the problem of different resolution, low contrast, illumination, and poor quality of the utilized color fundus images.
- We avoided overfitting by using the different augmentation processes cropping, rotation, and flipping.
- We present a comprehensive system that diagnoses the healthy and DR cases from the processed color

fundus images based on multi-label classification of deep learning.

- We get the benefits of transfer learning ideas on a small dataset IDRiD.
- We customized the efficientNet-B0 model to increase accuracy by optimizing major hyperparameters.
- We validated the proposed system by comparing it with others and using two significant performance measures matrices.

The rest of this manuscript is organized as follows. Section 2 presents all the studies currently conducted and their findings in detecting and grading DR. Section 3 provides a detailed illustration of the proposed framework based on the DL Section 4 shows the experiments and the critical reached results according to applying the proposed CAD system on the IDRiD dataset. Section 5 shows the discussions and the comparative study. Finally, Section 6 presents the conclusion and future work.

2.Literature review

A lot of researchers pay attention to the DR diagnosing. Therefore, they utilized DL in many CAD systems to detect DR. Some of these systems presented in literature depended absolutely on the DL. For example, Junjun et al. [10] processed the color fundus images by standardizing all images to a standard size of 256×256. They removed noise by using a Gaussian filter. Then, they applied some data augmentation procedures such as flipping, rotation, and cropping. They utilized ResNet18 with the regional scoring map. They employed L2 weight with the decay of 5×10^{-4} for all trainable variables except batch-normalization (BN). They trained the model with a 10^{-3} learning rate (lr) for 200 epochs. Although the authors raised the accuracy of detecting the severity grade, it could not predict the MAs signs, which represent the mild grade. Their system could detect H-EX, S-EX, HM, and NV but could not capture MAs. MAs are contributing to the DR diagnosing especially, the mild or early stage.

The work that was conducted by Sahlsten et al. [14] is like what was presented by Junjun et al. [10], while Sahlsten et al. [14] made grading of macular edema (ME). The authors started with the cropping and resizing images. They utilized DNN based on Inception-v3 architecture. They used instance normalization layers rather than BN layers. Although the authors predicted the severity grades accurately in the DR and ME, some limitations were found in their work. The first one was concerned with the image

grading reference. Grader biases negatively affect model performance. Second, DNN learns features that are unknown and ignored by experts. That is done when the network was applied to the image without determining the essential features in diagnoses such as MAs and EXs as well as their numbers, which are essential rules in grading DR.

Unlike the studies of Junjun et al. [10] and Sahlsten et al. [14], who used the color fundus images that were captured by fundus cameras, Rajalakshmi et al. [15] used a portable fundus camera that was a smartphone-based imaging device. They utilized the EyeArt v2.1.0 application. The authors skipped the training phase. Their work's main problems are that the poor image quality according to media opacities like cataract, and they utilized small samples of the utilized datasets.

Doshi et al. [16] aimed to diagnose the DR grades without extracting features manually. Therefore, they utilized the CNN model to learn the features. They made image resizing such as Junjun et al. [10] and Sahlsten et al. [14], but they resized images to 512×512, converted all images to the green channel, and applied histogram equalization (HE). The authors removed noise by using min-max normalization. They built their model based on two consecutive convolutions (CONV) layers followed by max-pooling (MP) and dropout (DO). After that, two of three consecutive CONV layers are followed by MP and DO, then one of two consecutive CONV layers followed by MP and DO. Finally, they added two fully connected layers (FC) and output five classes due to the DR dataset class label numbers. Although the authors extracted only the green channel to save memory space, the two other red and blue channels could give more features in the learning process. Besides, they should utilize a strong filter for removing noise rather than making the normalization only.

Wang and Yang [17] were along the same lines as Doshi et al. [16]. They also built a CNN model to diagnose the grades of DR. They built two CONV layers followed by MP, then three consecutive CONV layers followed by MP. Finally, they added two successive CONV layers followed by global pooling (GP) and the FC or dense layer to output the DR classes. The authors' work recorded a better kappa score than what was achieved by Doshi et al. [16]. Wang and Yang [17] made some data augmentation processes such as translation,

stretching, rotation, and flipping, but it is essential to reduce noise with a significant filter.

Burewar et al. [18] utilized U-Net and CNN model to segment the BV and diagnosed the DR stages. After training U-Net, they got the resulting vasculature network and merged it with the input green channel. They utilized 13 CONV, 5 MP and, 2 FC layers to diagnose the five DR stages in the classification phase. The authors utilized large color fundus datasets KAGGEL. The limitation of their study is that there is no need to merge the segmented BV network with the input green channel to recover the lost vessels from segmentation. They should increase the accuracy of U-Net to give the completed BV network or use a special filter to extract the thick and thin BV.

Zhu et al. [19] segmented the MA sign from the color fundus images to detect the presence/ absence of DR disease. The authors depended on extracting the BV to be eliminated and then detected the MAs. They made some preprocessing steps, including cropping, enhancing images by HE, and extracting the masks from the fundus images using the red channel. They utilized NB and SVM classifiers to detect DR by using the MAs features. The authors extracted MAs features based on color, shape, and diameter. Their work is critical in the early or the mild grade of DR, but it is not enough to determine the disease's stage. The other grades also contain MAs in addition to the EX, HM, NV, and others that need to be detected.

Nikhil and Rose [20] applied three CNN models on the KAGGLE fundus images, which are AlexNet, VGG16, and Inception- V3. The authors used the aforementioned pre-trained models based on transfer learning to classify the DR into five grades. They downsized all images, applied a filter to remove salt and pepper noise, and enhanced the contrast by using HE. Like Zhu et al. [19] work, all of these preprocessing steps share classification accuracy. The authors changed the hyperparameters to increase the accuracy of the utilized models. They concatenated the utilized three models' features to achieve accuracy, but they utilized the transfer learning idea on a large dataset.

Vora and Shrestha [21] detected the DR presence/absence using CNN models (Caffe and Keras) on KAGGLE and EyePACS datasets. Depending on the hyperparameters optimization, the authors used a batch size of 32 and a learning rate of 0.001. They utilized k-fold cross validation by 5, 10,

and 12 and the results that 5-fold cross-validation is optimum. In their work, ACC needs to be improved. Also, the authors ignored the contrast enhancement and filtering.

Similar to the objectives of Zhu et al. [19] and Vora and Shrestha [21], Gadekallu et al. [22] presented their method to detect the occurrence of DR. The authors normalized the color fundus images by StandardScaler. Then, they extracted features by Principal Component Analysis (PCA) and reduced the dimensionality by using the Firefly feature selection technique. Finally, they fed the feature vector to the CNN model with five layers for binary classification. Although their method can be implemented on high dimensional datasets may be in various domains, but the performance may not be achieved on the low dimensional datasets.

Khalifa et al. [23] applied some transfer learning models on the Asia Pacific Tele-Ophthalmology Society (APTOS 2019) dataset. They recorded the results of classifying the five DR grades by VGG16, ResNet18, GoogleNet, AlexNet, VGG19, and Squeeze Net. Before making the aforementioned CNN models' training, they increased the total number of images to avoid overfitting. They increased images four times the original number of some augmentation processes such as reflections around x, y, and x-y axes. Their work's main advantage is utilizing the novel dataset published in the second quadrant of 2019. On the other hand, the authors ignored contrast enhancement and noise removing like Vora and Shrestha [21].

Amalia et al. [24] proposed two CNN models to detect DR. The output of the CNN GoogleNet model is the listing detection entered to Long Short-Term Memory (LSTM) model. The authors applied cropping, rotation, and contrast enhancement by HE. They added three captions for each image in the dataset manually. Then, they applied the two CNN models. Using the LSTM model, the features of the image is mapped in words to constitute a sentence. They utilized Adam optimizer with Lr equals 0.001, the batch size is 30, and epochs equal 50. The resulting sentences from LSTM can help the radiologists to diagnose the DR classes. The authors provided the features in words without diagnosing the DR grade. They left the diagnosing for the ocular specialist. On the other hand, the GoogleNet output is only two classes, normal and DR.

To overcome the current literature's limitations, we processed the images to enhance the contrast and remove noise. After that, we resized images to a standard size of 256×256. Filtering, cropping, rotation, mirroring, and normalization were done to avoid overfitting and ensure quality. Then, we optimized the traditional effecientNet-B0 model by optimizing the hyperparameters. After that, we evaluated the Indian Diabetic Retinopathy Image Dataset (IDRiD) dataset and the performance measures metrics. We made the different comparisons with other systems and models.

3.Methods

This section presents an illustration of the proposed CAD system framework to diagnose the normal and DR grades. The system outputs five classes (normal, mild, moderate, severe NPDR, and PDR). We start our framework by feeding the IDRiD color fundus images to the pre-processing phase. We enhanced contrast and removed the noise. Besides, we normalized the images and resized them to a standard size of 256×256. After that, we made cropping, rotation, mirroring. The normalized images are fed to the customized effecientNet-B0 model to output the healthy and DR grades, as shown in *Figure 2*. The framework consists of the following phases:

3.1The pre-processing

This phase consists of some steps. First, we make filtering by the median filter to remove noise and improve the contrast. Second, we resized the images to a standard size of 256×256. Third, we made some transformation processes on the images, such as cropping, random rotation, and vertical and horizontal flipping. Finally, we normalized the images by using the mean and standard deviation (STD). We make all of these pre-processing steps on the entered color fundus images to enlarge the dataset. The main advantage of learning transfer is that it can train small datasets. Moreover, we utilized the augmentation processes to avoid overfitting. *Figure 3* shows the example of the cropping step in the transformation operations.

3.2Modeling

EffecientNet is one of the CNN architectures. It is a scale model that uniformly scales all dimensions of depth, width, and resolution of the network by utilizing a compound coefficient [11]. Model scaling aims to increase the depth by making the network deeper, adding more convolution layers. These layers provide more detailed features of the image. Of course, the deeper model is more accurate than the

models with few layers. On the other hand, the model with more significant channels in its layer is wider than the others with few channels (R, G, and B). Moreover, the highest image resolution equals higher accuracy and more details. But the increase in the three aforementioned items results in a bigger network. So, it is very necessary to balance all the network dimensions in the scaling process. The next subsection provides the effecientNet-B0 theory.

3.2.1 EffecientNet model

Constant coefficients are (d) for depth, (w) for width, and (r) for image size. If the computations are $2n$, then we increase the network depth by $\mathcal{D} n$, width by ωn , image resolution by $R n$. The model used compound coefficient ϕ to scale network width, depth, and resolution uniformly. Therefore, the model distributes the resources (X) on depth d, width w, and resolution r, as shown in equations (1) and (2).

$$d = \mathcal{D}^\phi, w = \Omega^\phi, r = R^\phi \quad (1)$$

$$X = \mathcal{D} \cdot \Omega^2 \cdot R^2 \approx 2 \quad \text{where } \mathcal{D} \geq 1, \Omega \geq 1, R \geq 1 \quad (2)$$

According to the scaling theory, the EffecientNet-B0 model is scaled to obtain EffecientNet B1, B2, B3,

B4, B5, B6, and B7 models. In this paper, we customized the EffecientNet-Bo model by optimizing the values of some hyperparameters such as the utilized optimizer and its parameter values like lr, regularizations, DO, and others that will be illustrated in the next section. We utilized the model to make a multi-label classification of the entered color fundus images of healthy and various DR grades. The scaling of the utilized EffecientNet-Bo model architecture is shown in *Table 1*. We can notice the changes in the resolution r, channels c, and the number of layers L, which determines the model's depth. MB is the MobileNet model. The model consists of one CONV layer with kernel three and six layers of MB CONV layers. One layer is of MB CONV1, and five are of MB CONV6 between 3 and 5 kernels. The final layer is the CONV 1×1 and MP, then the dense layer or FC. The model started with r of all images 256×256 and ended with r equals 8×8 with increasing C (R, G, and B). We utilized rectified linear unit (Relu) activation function, BN of 512, DO of 0.4.

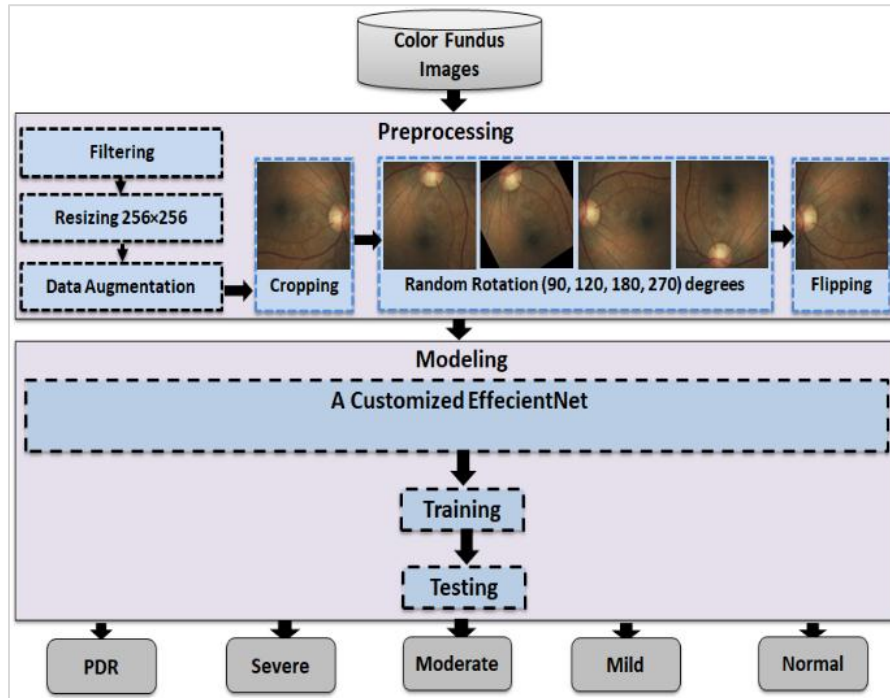
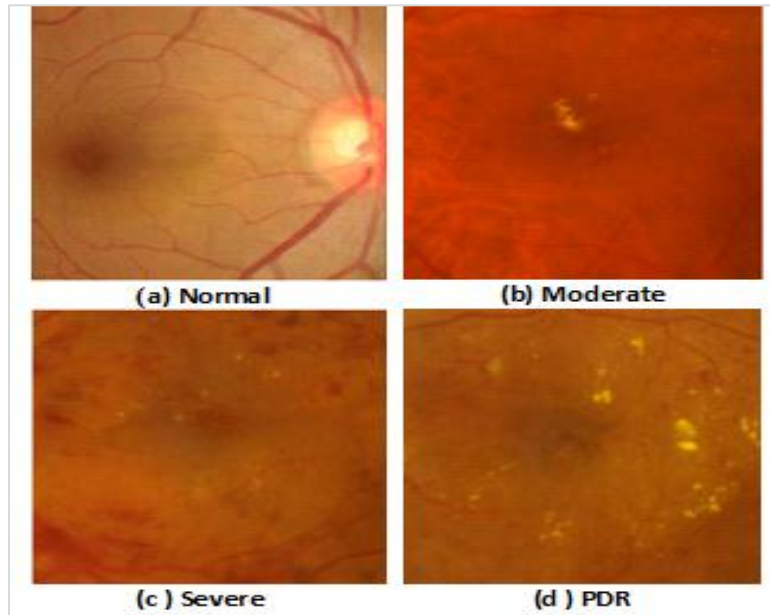


Figure 2 The proposed CAD system framework

Table 1 The scaling parameters of the EffecientNet-B0 model

Operator	R	C	D
CONV 3×3	256×256	32	1
MB CONV1	128×128	16	1
MB CONV6	128×128	24	2
MB CONV6	64×64	40	2
MB CONV6	32×32	80	3
MB CONV6	16×16	112	3
MB CONV6	16×16	192	4
MB CONV6	8×8	320	1
CONV 1×1, MP, dense	8×8	1280	1

**Figure 3** The example of the cropped IDRiD images

4.Results

This section provides a detailed description about the utilized IDRiD [12] dataset, as shown below. We implemented this work by using python 3.7 and cloud computing "Google Colab". This work was implemented on TensorFlow 2.4. Also, for the preprocessing steps, we utilized the open-source Python library OpenCV. For classification (training and validation), we utilized the deep learning Python open-source Library TFLearn. We ran our experiments on a core i5/2.4 GHz machine. It has 8GB RAM and an NVIDIA VGA card with 1GB VRAM.

IDRiD dataset [12]: It is categorized into three parts, which are segmentation, disease grading, and localization. In segmentation, the dataset contains 81 color fundus images in JPG format. It has GT of the four lesions, which are MA, HM, H-EX, and S-EX with TIF format. Eighty-one pixel-level annotated images were separated into 54 for training and 27 for

testing. The EX is found in 80 images, HM is found in 80 images, and MA is found in 81 images. In DR grading, the dataset contains 516 color fundus images in JPG format. The images are divided into 413 training set and 103 testing set. The GT of the disease grading contains two CSV label files. One for training and one for testing. Finally, in the localization, the dataset contains 516 color fundus images in JPG format, which are divided into 413 images for training and 103 images for testing. The GT is for the OD central location and fovea center location. All images are 2848 × 12864. All images are captured by the AKowa VX-10 alpha digital fundus camera with 500 field-of-view and centered near the macula. Six experts served in segmentations and DR grades diagnosing. This dataset is multi-label. It means that it contains many DR lesions. These lesions constitute the DR grades; mild, moderate, and severe. *Table 2* and *Figure 4* show the labels of the IDRiD dataset and their counts.

We measured the customized model's performance with accuracy (ACC) and Dice similarity coefficient (DSC), as shown in equations 3 and 4.

$$DSC = \frac{2TP}{2(TP+FP+FN)} \quad (3)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Where TP is a true positive, TN is a true negative, FN is a false negative, and FP is a false positive. *Table 3* shows the values of hyperparameters of the customized EfficientNet B0 and the DSC results.

From *Table 3*, we can notice that we tried many optimizers such as Adam, Adagrad, Adadelta, AdaMax, ASGD, RmsProp, and Rprop. Moreover, we change the epochs from 10 to 100. From the different values, we can notice that the Adamax optimizer with 100 epochs and learning rate lr 0.002

achieved DSC to be 78.45% in training and 65% in the validation on the IDRiD dataset. The parameters of the AdaMax optimizer are betas=(0.9, 0.999), eps=1e-08. The optimizer Adam comes in the third order as it achieved 56% for DSC with the same parameters. But AdamW comes in the second-order as it achieved 63% for DSC with betas=(0.9,0.999), eps=1e-08, weight_decay=0.01. RmsProp optimizer comes in the fourth-order. It achieved 62% for DSC with 10 epochs and 0.002 lr. RmsProp is with alpha=0.99, and eps=1e-08. Similarly, we can observe that ASGD with lambda=0.0001, alpha=0.75, and t0=1000000 is the worst optimizer in classifying the IDRiD dataset by the EfficientNet-B0 model. It is trained by 0.002 lr and 100 epochs. On the other hand, the model achieved 84% accuracy of multi-label classification. *Figure 5* shows the ACC of the training and validation through 100 epochs.

Table 2 The labels and their counts according to IDRiD images

Label	Count
Normal	134
Mild	20
Moderate	136
Severe	74
PDR	49

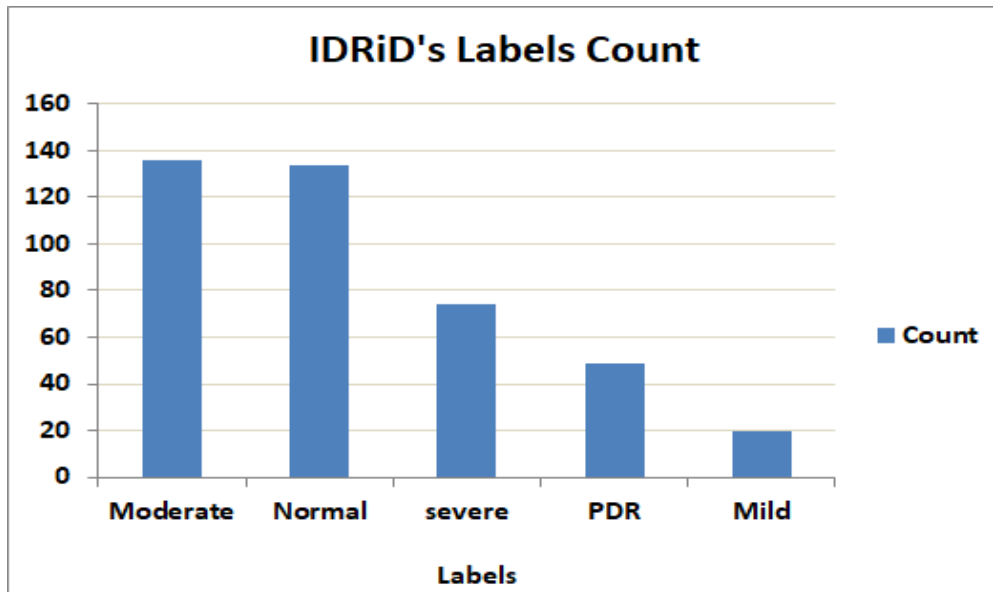


Figure 4 The distribution of the IDRiD dataset labels

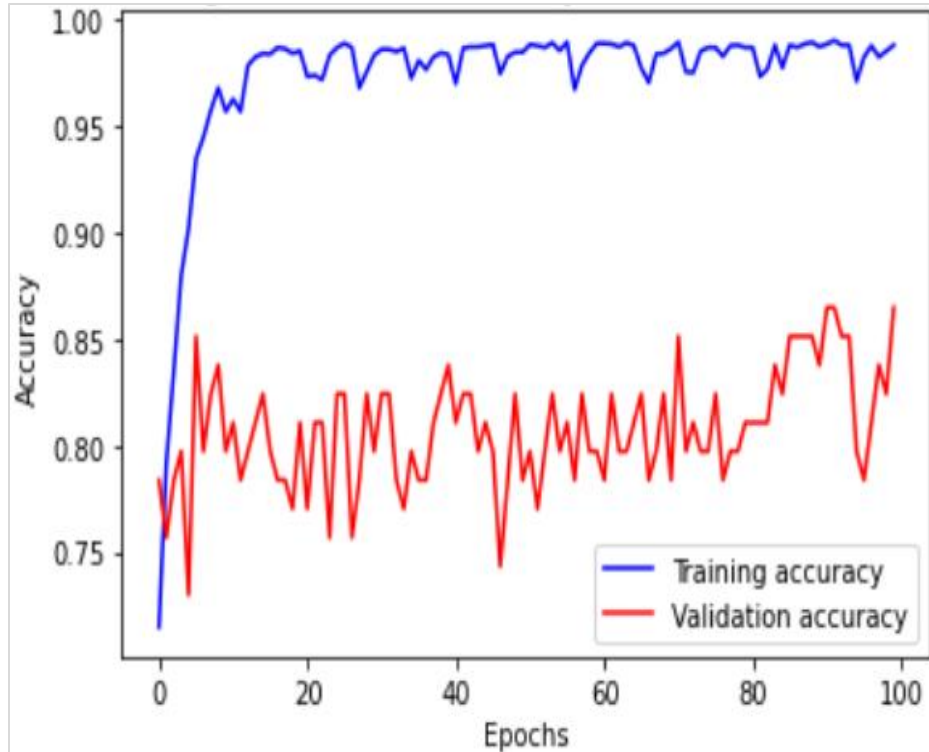


Figure 5 The training and validation ACC

Table 3 The hyper-parameters values of the customized effientNet-B0 model

Optimize r	Optimizer parameters	Ep.	Lr	T loss	T-DSC (%)	V loss	V-DSC (%)
Adadelata	rho=0.9, eps=1e-06	10	0.001	0.74	18	0.68	20
		100	0.001	0.65	24.5	0.73	16.2
Adagrad	eps=1e-10	50	0.001	0.66	24.12	0.73	15.7
		100	0.001	0.57	28.2	0.56	25.8
Adam	betas=(0.9, 0.999), eps=1e-08	50	0.001	0.50	39.4	0.58	26
		100	0.002	0.19	56	0.28	29
		10	0.003	0.20	44	0.27	45
Adamax	betas=(0.9, 0.999), eps=1e-08	10	0.002	0.44	47	0.62	29
		50	0.002	0.14	65.2	0.26	30
		50	1e-3	0.56	28	0.58	18
		100	0.002	0.8	78.45	0.12	65
AdamW	betas=(0.9, 0.999), eps=1e-08, weight decay=0.01	50	0.002	0.48	43	0.50	31
		100	0.002	0.30	63	0.35	36
ASGD	lambd=0.0001, alpha=0.75, t0=1000000	100	0.001	0.72	18	0.69	21
		100	0.002	0.75	18	0.85	24
RmsProp	alpha=0.99, eps=1e-08	10	0.002	0.23	28	0.22	61
		100	0.002	0.42	26	6.7	66
		10	0.002	0.62	25	0.61	24
Rprop	etas=(0.5, 1.2), step_sizes=(1e-06, 50)	10	0.001	0.19	41	0.18	53

5. Discussion

This section provides the comparative study between the proposed system and the other conducted in the literature. We also present a comparison between the customized effcientNet-B0 model and other CNN models based on transfer learning by utilizing the

IDRiD dataset. Although Pan et al. [10] work is very near our system phases, they achieved an accuracy of 78.4% on the EyPacs dataset, which is larger than our utilized dataset IDRiD. In our work, we avoided overfitting by doing some augmentation processes to enlarge the utilized dataset. Besides, we utilized effcientNet-B0 based on the transfer learning, which

is suitable in small datasets. The same thing is for Sahlsten et al. [14]. They achieved 89.6% for sensitivity by applying InceptionV3 on a private dataset. On the other hand, Doshi et al. [17] achieved 0.38 for kappa score on the EyePACS dataset, which means that the graders' agreement is less than 0.50. Burewar et al. [18] achieved 93% for ACC larger than ours, but they utilized a large dataset KAGGEL while utilizing a small dataset IDRiD. Burewar et al. [18] should achieve high ACC of U-Net segmentation rather than merging the segmenting BV images with their input green channel. Of course, the ACC of segmentation affects the results of classification.

Zhu et al. [19] achieved 0.832 for the area under the curve on the Messidor dataset. This result reflects the presence of DR by only segmenting the MAs ignoring the DR grading. Contrary to our work, Nikhil and Rose [20] applied the transfer learning idea on a large dataset KAGGLE. They achieved 80.1% for ACC. Khalifa et al. [23] achieved 97.9% for ACC in grading DR by utilizing transfer learning on a novel APTOS 2019 dataset by applying state of the art models such as AlexNet. If they applied for their work again without preprocessing steps and optimizing the models' hyperparameters as they did, we think the results will be different. From our point of view, they need to solve the two previous issues. Therefore, we put the noise removing and the augmentation steps into account in our work. Alyoubi et al. [3], in their review, proved that 73% of the conducted studies that analyses and detected DR only detected the presence/ absence of DR, but 27% worked in diagnosing the DR grades. It means that our study is one of the few studies that worked on diagnosing the DR grades and can open the way for other researchers to detect the early and advanced stages of the disease.

Vora and Shrestha [21] achieved 76% for average ACC in binary classification to detect only the presence/absence of DR without diagnosing the various DR grades by using CNN on KAGGLE and EyePACS datasets. Unlike Vora and Shrestha [21],

Gadekallu et al. [22] achieved 97% for ACC in DR binary classification on a high-dimensional dataset that was collected from the UCI machine learning repository. On the other hand, the authors concluded that their method might not give the same performance implemented on the low dimensional datasets as it may fall into overfitting. Therefore, it will be practical to utilize data augmentation methods for both high and low dimensional datasets. In this respect, we put standard pre-processing steps that can be utilized in low and high-dimensional datasets. Amalia et al. [24] achieved 90% for ACC on the MESSIDOR dataset by combining the CNN and LSTM. They stated the lesion's names in sentences to help the ophthalmologists to determine the health and DR grade. They achieved 90% accuracy on the MESSIDOR dataset. The authors performed binary classification, while our system provides diagnosing of the healthy and four DR grades

The advantages of the proposed system are its comprehensive style to diagnose the various DR grades, more accurate than others in the literature, utilized real dataset, can be applied to small datasets without falling in overfitting, it can be applied to other ocular diseases diagnosing in the future such as Glaucoma. Because the system depends on a multi-label classification idea, it could diagnose multiple ocular diseases simultaneously. This approach's limitations are that the accuracy needs to be increased somewhat, and it should be applied to other various multi-label datasets to ensure reality and reliability.

Table 4 shows that the proposed system that uses the customized effecientNet-B0 achieved 86% and 65% for ACC and DSC, respectively. It ranks the first order in the presented comparison, but VGG16 and VGG19 come in the second and third-order. Finally, ResNet 50 and Inception V3 come equally in the last order. On the other hand, *Figure 5*. Shown example of confusion matrix of VGG16 and VGG19 models on IDRiD datasets. *Figure 6* present the values of validating each model due to the five classes' true labels (normal, mild, moderate, severe NPDR, and PDR) and the same classes' predicted labels.

Table 4 The comparison between the proposed customized effecientNet-B0 and ResNet50, InceptionV3, VGG16, and VGG19 models in the IDRiD dataset

Model	ACC (%)	DSC (%)
ResNet50	32.5	32.5
Inception V3	32.5	32.5
VGG16	34	33.7
VGG19	33	32.5
The Customized EffecientNet-B0	84	65

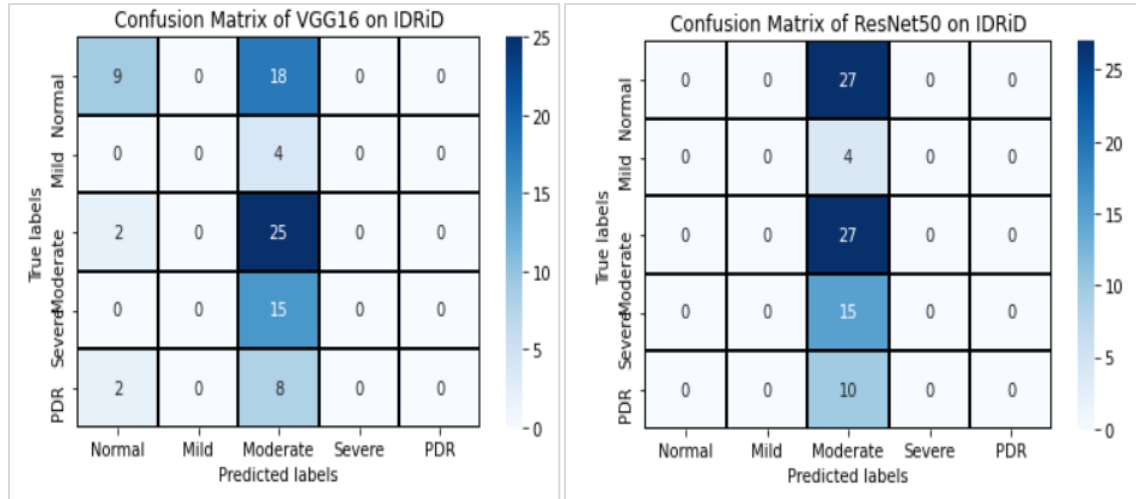


Figure 6 Examples of the confusion matrix VGG16 and ResNet50 on the IDRiD dataset

6. Conclusion and future work

Transfer learning is essential in a small dataset to avoid overfitting. Moreover, it reduces the training time and saves memory space. According to this fact, we produce a comprehensive CAD system based on the CNN efficientNet-B0 model. We utilized the pre-trained EfficientNet-B0 model depending on the multi-label classification idea. We detect the healthy and diabetic retinopathy grades mild, moderate, severe NPDR, and PDR. We get the merits of the label correlation and dependency by customizing the pre-trained efficientNet-B0 model. We customized the model by optimizing the hyperparameters such as optimizer, optimizer's parameter values, learning rate, and the epochs' number. The preprocessing steps such as filtering, normalization, and augmentation are significant in the proposed CAD system. In the future, we intend to detect other retinal diseases such as Glaucoma, diabetic retinopathy, and age-related macular degeneration simultaneously from other imaging modalities such as OCTA.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] At a glance: Diabetic Retinopathy. <https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy>. Accessed 01 November 2020.
- [2] <https://www.webmd.com/diabetes/diabetic-retinopathy>. Accessed 01 January 2020
- [3] Alyoubi WL, Shalash WM, Abulkhair MF. Diabetic retinopathy detection through deep learning

techniques: a review. *Informatics in Medicine Unlocked*. 2020; 20:1-11.

- [4] Dhingra S, Khaw PT. The Moorfields safer surgery system. *Middle East African Journal of Ophthalmology*. 2009; 16(3):112-5.
- [5] Dey N, Borra S, Ashour AS, Shi F, editors. *Machine learning in bio-signal analysis and diagnostic imaging*. Academic Press; 2018.
- [6] Gao J, Leung C, Miao C. Diabetic retinopathy classification using an efficient convolutional neural network. In *international conference on agents 2019* (pp. 80-5). IEEE.
- [7] Tymchenko B, Marchenko P, Spodarets D. Deep learning approach to diabetic retinopathy detection. *arXiv preprint arXiv:2003.02261*. 2020.
- [8] Kirange DK, Chaudhari JP, Rane KP, Bhagat KS, Chaudhri N. Diabetic retinopathy detection and grading using machine learning. *International Journal of Advanced Trends in Computer Science and Engineering*. 2019; 8(6):3570-6.
- [9] Xu K, Feng D, Mi H. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules*. 2017; 22(12):1-7.
- [10] Junjun P, Zhifan Y, Dong S, Hong Q. Diabetic retinopathy detection based on deep convolutional neural networks for localization of discriminative regions. In *international conference on virtual reality and visualization 2018* (pp. 46-52). IEEE.
- [11] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In *international conference on machine learning 2019* (pp. 6105-14). PMLR.
- [12] Porwal P, Pachade S, Kokare M, Deshmukh G, Son J, Bae W, Liu L, Wang J, Liu X, Gao L, Wu T. IDRiD: Diabetic retinopathy-segmentation and grading challenge. *Medical Image Analysis*. 2020; 59:101561.
- [13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014:1-14.

- [14] Sahlsten J, Jaskari J, Kivinen J, Turunen L, Jaanio E, Hietala K, et al. Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. *Scientific Reports*. 2019; 9(1):1-11.
- [15] Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018; 32(6):1138-44.
- [16] Doshi D, Shenoy A, Sidhpura D, Gharpure P. Diabetic retinopathy detection using deep convolutional neural networks. In international conference on computing, analytics and security trends 2016 (pp. 261-6). IEEE.
- [17] Wang Z, Yang J. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *arXiv preprint arXiv:1703.10757*. 2017:1-9.
- [18] Burewar S, Gonde AB, Vipparthi SK. Diabetic retinopathy detection by retinal segmentation with region merging using CNN. In international conference on industrial and information systems 2018 (pp. 136-42). IEEE.
- [19] Zhu CZ, Hu R, Zou BJ, Zhao RC, Chen CL, Xiao YL. Automatic diabetic retinopathy screening via cascaded framework based on image-and lesion-level features fusion. *Journal of Computer Science and Technology*. 2019; 34(6):1307-18.
- [20] Nikhil M, Rose A. Diabetic retinopathy stage classification using CNN. *International Research Journal of Engineering and Technology*. 2019; 6(5):5969-74.
- [21] Vora P, Shrestha S. Detecting diabetic retinopathy using embedded computer vision. *Applied Sciences*. 2020; 10(20):1-10.
- [22] Gadekallu TR, Khare N, Bhattacharya S, Singh S, Reddy Maddikunta PK, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*. 2020; 9(2):1-16.
- [23] Khalifa NE, Loey M, Taha MH, Mohamed HN. Deep transfer learning models for medical diabetic retinopathy detection. *Acta Informatica Medica*. 2019; 27(5):327-32.
- [24] Amalia R, Bustamam A, Sarwinda D. Detection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory. In journal of physics: conference series 2021 (pp.1-6). IOP Publishing.



Eman Abdelmaksoud is a Ph.D. student at the Information Systems Department, Faculty of Computers and Information, Mansoura University. She received her B.Sc. from the Faculty of Computers and Information, Zagazig University, Zagazig, Egypt. She received her M.Sc. in 2016 from the Faculty of Computers and Information, Mansoura University. She has authored approximately 20 research publications in journals, book chapters, and conference proceedings. Her current research interests are Computer Vision, Medical Image Analysis, Artificial Intelligence, Machine Learning, and Biomedical Engineering. Email: eng.eman.te@gmail.com



Sherif Barakat is a Professor at the Information Systems Department, Faculty of Computers and Information, Mansoura University, Egypt. He had received his B.Sc. and M.Sc. from the Faculty of Science, Mansoura University, Mansoura, Egypt. He received his Ph.D. from the Faculty of Science, Helwan University, Helwan, Egypt in 2003. He has authored/coauthored over 50 research publications in peer-reviewed reputed journals, book chapters, and conference proceedings. His current research interests are Machine Learning, Pattern Recognition, and Artificial Intelligence. Email: sherief.barakat@mans.edu.eg



Mohammed Elmogy received the B.Sc. and M.Sc. degrees from the Faculty of Engineering, Mansoura University, Mansoura, Egypt, and the Ph.D. degree from Informatics Department, MIN Faculty, Hamburg University, Hamburg, Germany, in 2010. He is an Associate Professor with the Information Technology Department, Faculty of Computers and Information, Mansoura University, Egypt. He worked as a Visiting Researcher from 7/2016 to 8/2019 with the Bioengineering Department, University of Louisville, Louisville, USA. He has authored/coauthored over 200 research publications in peer-reviewed reputed journals, book chapters, and conference proceedings. He advised more than 30 master's and doctoral graduates. His current research interests are computer vision, medical image analysis, machine learning, pattern recognition, and biomedical engineering. Dr. Elmogy has served as a Reviewer for various international journals. He served as a technical program committee member in many workshops and conferences. He is a Senior IEEE member and a Professional Member of ACM society. Email: melmogy@mans.edu.eg