Diabetic retinopathy detection through deep learning techniques: A review

Wejdan L. Alyoubi [*], Wafaa M. Shalash, Maysoon F. Abulkhair

Information Technology Department, University of King Abdul Aziz, Jeddah, Saudi Arabia

ABSTRACT

Diabetic Retinopathy (DR) is a common complication of diabetes mellitus, which causes lesions on the retina that effect vision. If it is not detected early, it can lead to blindness. Unfortunately, DR is not a reversible process, and treatment only sustains vision. DR early detection and treatment can significantly reduce the risk of vision loss. The manual diagnosis process of DR retina fundus images by ophthalmologists is time-, effort-, and cost-consuming and prone to misdiagnosis unlike computer-aided diagnosis systems. Recently, deep learning has become one of the most common techniques that has achieved better performance in many areas, especially in medical image analysis and classification. Convolutional neural networks are more widely used as a deep learning method in medical image analysis and they are highly effective. For this article, the recent state-of-the-art methods of DR color fundus images detection and classification using deep learning techniques have been reviewed and analyzed. Furthermore, the DR available datasets for the color fundus retina have been reviewed. Difference challenging issues that require more investigation are also discussed.

1. Introduction

In the healthcare field, the treatment of diseases is more effective when detected at an early stage. Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin [1]. It affects 425 million adults worldwide [2]. Diabetes affects the retina, heart, nerves, and kidneys [1,2].

Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood [3]. DR can lead to a loss of vision if it is in an advanced stage. Worldwide, DR causes 2.6% of blindness [4]. The possibility of DR presence increases for diabetes patients who suffer from the disease for a long period. Retina regular screening is essential for diabetes patients to diagnose and to treat DR at an early stage to avoid the risk of blindness [5]. DR is detected by the appearance of different types of lesions on a retina image. These lesions are microaneurysms (MA), haemorrhages (HM), soft and hard exudates (EX) [1,6,7].

- *Microaneurysms* (MA) is the earliest sign of DR that appears as small red round dots on the retina due to the weakness of the vessel's walls. The size is less than 125 μm and there are sharp margins. Michael et al. [8] classified MA into six types, as shown in Fig. 1. The types of MA were seen with AOSLO reflectance and conventional fluorescein imaging.

- *Haemorrhages* (HM) appear as larger spots on the retina, where its size is greater than 125 μm with an irregular margin. There are two types of HM, which are flame (superficial HM) and blot (deeper HM), as shown in Fig. 2.
- *Hard exudates* appear as bright-yellow spots on the retina caused by leakage of plasma. They have sharp margins and can be found in the retina's outer layers.
- *Soft exudates* (also called cotton wool) appear as white spots on the retina caused by the swelling of the nerve fiber. The shape is oval or round.

Red lesions are MA and HM, while bright lesions are soft and hard exudates (EX). There are five stages of DR depending on the presence of these lesions, namely, no DR, mild DR, moderate DR, severe DR and proliferative DR, which are briefly described in Table 1. A sample of DR stages images is provided in Fig. 3.

The automated methods for DR detection are cost and time saving and are more efficient than a manual diagnosis [10]. A manual diagnosis is prone to misdiagnosis and requires more effort than automatic methods. This paper reviews the recent DR automated methods that use deep learning to detect and to classify DR. The current work covered 33 papers which used deep learning techniques to classify DR images. This paper is organized as follows: Section 2 briefly explains deep learning techniques, while Section 3 presents the various fundus retina datasets. Section 4 presents the performance measures while Section 5 reviews

---
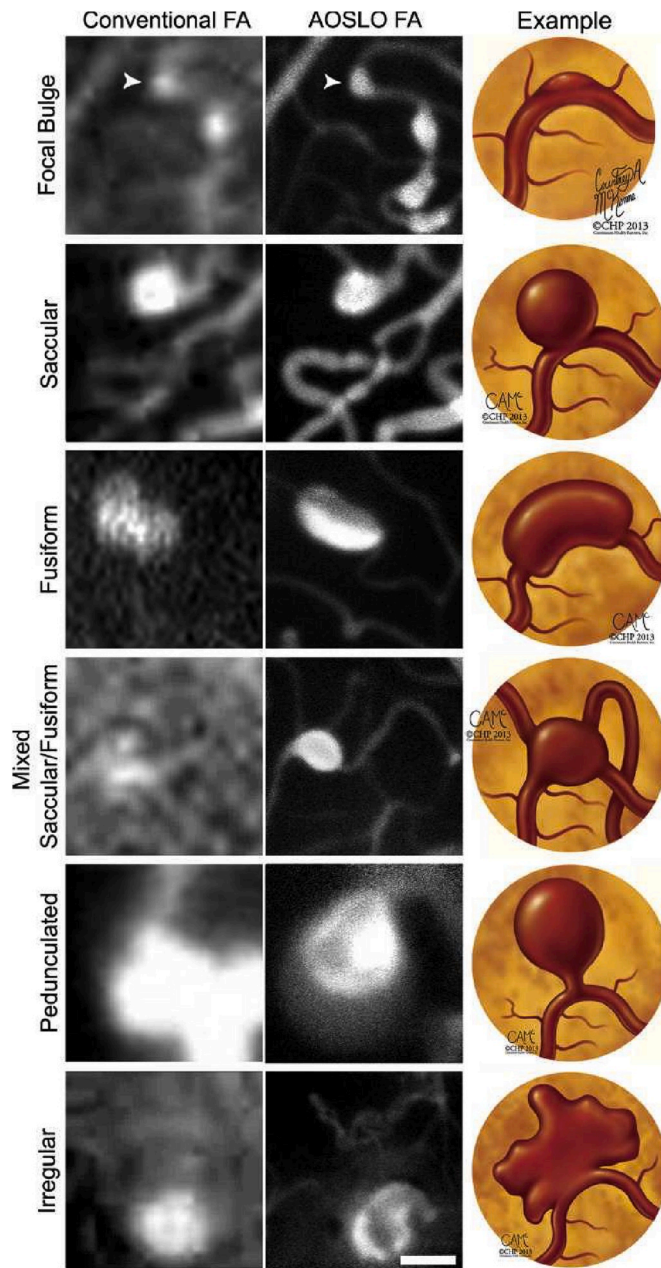
Fig. 1. The different types of MA [8].
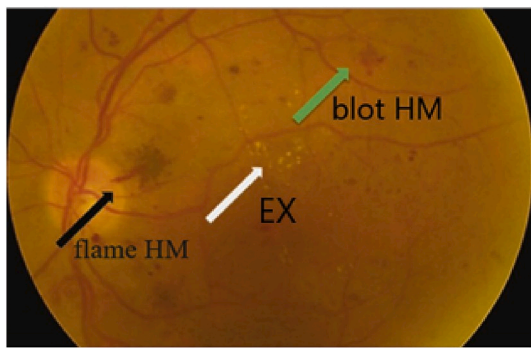
Fig. 2. The different types of HM [9].

Table 1

Levels of DR with its associative lesions [13].

DR Severity Level	Lesions
No DR	Absent of lesions
Mild non-proliferative DR	MA only
Moderate non-proliferative DR	More than just MA but less than severe DR
Severe non-proliferative DR	Any of the following: • more than 20 intraretinal HM in each of 4 quadrants • definite venous beading in 2+quadrants • Prominent intraretinal microvascular abnormalities in 1+ quadrant • no signs of proliferative DR
Proliferative DR	One or more of the following: vitreous/pre-retinal HM, neovascularization

Section 6 describes different DR automated classification methods while a discussion section is presented in Section 7. A summary is provided in Section 8.

2. Deep learning

Deep learning (DL) is a branch of machine learning techniques that involves hierarchical layers of non-linear processing stages for unsupervised features learning as well as for classifying patterns [11]. DL is one computer-aided medical diagnosis method [12]. DL applications to medical image analysis include the classification, segmentation, detection, retrieval, and registration of the images.

Recently, DL has been widely used in DR detection and classification. It can successfully learn the features of input data even when many heterogeneous sources integrated [14]. There are many DL-based methods such as restricted Boltzmann Machines, convolutional neural networks (CNNs), auto encoder, and sparse coding [15]. The performance of these methods increases when the number of training data increase [16] due to the increase in the learned features unlike machine learning methods. Also, DL methods did not require hand-crafted feature extraction. Table 2 summarizes these differences between DL and machine learning methods.

CNNs are more widely used more than the other methods in medical image analysis [17], and it is highly effective [15].

There are three main layers in the CNN architecture, which are convolution layers (CONV), pooling layers, and fully connected layers (FC). The number of layers, size, and the number of filters of the CNN vary according to the author's vision. Each layer in CNN architecture plays a specific role. In the CONV layers, different filters convolve an image to extract the features. Typically, pooling layer follows the CONV layer to reduce the dimensions of feature maps. There are many strategies for pooling but average pooling and max pooling are adopted most [15]. A FC layers are a compact feature to describe the whole input image. SoftMax activation function is the most used classification function. There are different available pretrained CNN architectures on ImageNet dataset such as AlexNet [19], Inception-v3 [20] and ResNet [21]. Some studies like [22,23] transfer learning these pretrained architectures to speed up training while other studies build their own CNN from scratch for classification. The transfer learning strategies of pretrained models include finetuning last FC layer or finetuning multiple layers or training all layers of pretrained model.

Generally, the process used to detect and to classify DR images using DL begins by collecting the dataset and by applying the needed preprocess to improve and enhance the images. Then, this is fed to the DL method to extract the features and to classify the images, as shown in Fig. 4. These steps are explained in the following sections.

3. Retina dataset

There are many publicly available datasets for the retina to detect DR

different image preprocessing methods used with fundus images.

and to detect the vessels. These datasets are often used to train, validate and test the systems and also to compare a system's performance against other systems. Fundus color images and optical coherence tomography (OCT) are types of retinal imaging. OCT images are 2 and 3- dimensional images of the retina taken using low-coherence light and they provide considerable information about retina structure and thickness, while fundus images are 2-dimensional images of the retina taken using reflected light [24]. OCT retinal images have been introduced in past few years. There is a diversity of publicly available fundus image datasets that are commonly used. Fundus image datasets are as follows:

- *DIARETDB1* [25]*:* It contains 89 publicly available retina fundus images with the size of 1500 × 1152 pixels acquired at the 50-degree field of view (FOV). It includes 84 DR images and five normal images annotated by four medical experts.
- *Kaggle* [26]: It contains 88,702 high-resolution images with various resolutions, ranging from 433 × 289 pixels to 5184 × 3456 pixels, collected from different cameras. All images are classified into five DR stages. Only training images ground truths are publicly available. Kaggle contains many images with poor quality and incorrect labeling [23,27].
- *E-ophtha* [28]: This publicly available dataset includes E-ophtha EX and E-ophtha MA. E-ophtha EX includes 47 images with EX and 35 normal images. E-ophtha MA contains 148 images with MA and 233 normal images.
- *DDR* [23]*:* This publicly available dataset contains 13,673 fundus images acquired at a 45-degree FOV annotated to five DR stages. There are 757 images from the dataset annotated to DR lesions.
- *DRIVE* [29]: This publicly available dataset is used for blood vessel segmentation. It contains 40 images acquired at a 45-degree FOV. The images have a size of 565 × 584 pixels. Among them, there are seven mild DR images, and the remaining include images of a normal retina.
- *HRF* [30]: These publicly available images provided for blood vessel segmentation. It contains 45 images with a size of 3504 × 2336 pixels. There are 15 DR images, 15 healthy images and 15 glaucomatous images.
- *Messidor* [31]*:* This publicly available dataset contains 1200 fundus color images acquired at a 45-degree FOV annotated to four DR stages.
- *Messidor-2* [31]*:* This publicly available dataset contains 1748 images acquired at a 45-degree FOV.

Table 2
The differences between DL and machine learning methods.

	DL	Machine learning
Hand-crafted feature extraction	Not required	Required
Training data	Required large data	Not required large data
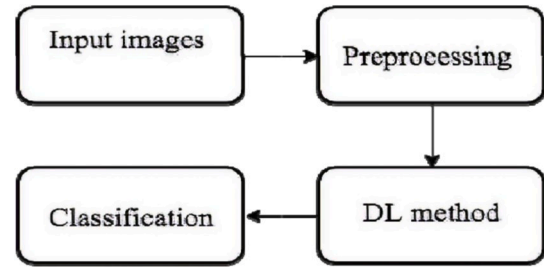
Fig. 4. The process of classifying the DR images using DL.

- *STARE* [32]: This publicly available dataset is used for blood vessel segmentation. It contains 20 images acquired at a 35-degree FOV. The images have a size of 700 × 605 pixels. Among them, there are 10 normal images.
- *CHASE DB1* [33]: This publicly available dataset provided for blood vessel segmentation. It contains 28 images with a size of 1280 × 960 pixels and acquired at a 30-degree FOV.
- *Indian Diabetic Retinopathy Image dataset (IDRiD)* [34]: This publicly available dataset contains 516 fundus images acquired at a 50-degree FOV annotated to five DR stages.
- *ROC* [35]*:* It contains 100 publicly available retina images acquired at the 45-degree FOV. Its size ranging from 768 × 576 to 1389 × 1383 pixels. The images annotated to detect MA. Only training ground truths are available.
- *DR2* [36]*:* It contains 435 publicly available retina images with 857 × 569 pixels. It provides referral annotations for images. There are 98 images were graded as referral.

The study of [37] used DIARETDB1 datasets to detect DR lesions. The study of [38] used DIARETDB1 and E-ophtha to detect red lesion while the study of [39] used these datasets to detect MA. In Ref. [40] DIARETDB1 was used to detect EX. The Kaggle dataset was used in the studies of [22,37,41–45] to classify DR stages. DRIVE, HRF, STARE and
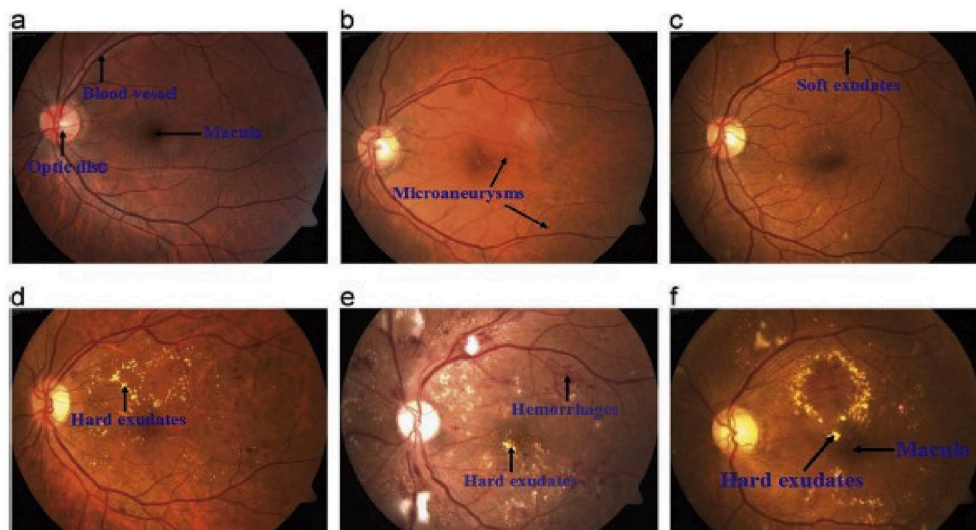
Fig. 3. The DR stages: (a) normal retinal (b) Mild DR, (c) Moderate DR, (d) Severe DR, (e) Proliferative DR,(f) Macular edema [18].

CHASE DB1 were used in the work of [46] to segment the blood vessels, while in Ref. [47] DRIVE dataset was used. The results of these studies are discussed in section 5. Table 3 compares these datasets. Most of the studies processed the datasets before using them for DL methods. The next sections discuss the performance measures and preprocessed methods.

4. Performance measures

There are many performance measurements that applied to DL methods to measure their classification performance. The commonly used measurements in DL are accuracy, sensitivity, specificity and area under the ROC curve (AUC). Sensitivity is the percentage of abnormal images that classified as abnormal, and specificity is the percentage of normal images that classified as normal [65]. AUC is a graph created by plotting sensitivity against specificity. Accuracy is the percentage of images that are classified correctly. The following is the equations of each measurement.

$$Specificity = TN / (TN \, FP) \tag{1}$$

$$Sensitivity = TP/ (TP \, FN) \tag{2}$$

$$Accuracy = TN + TP/(TN + TP + FN + FP) \tag{3}$$

True positive (TP) is the number of disease images that classified as disease. True negative (TN) is the number of normal images that classified as normal while false positive (FP) is the number of normal images that classified as disease. False negative (FN) is the number of disease images that classified as normal. The percentage of performance measures used in the studies, that involved in the current work, shown in Fig. 5.

5. IMAGE PREPROCESSING

Image preprocessing is a necessary step to remove the noise from images, to enhance image features and to ensure the consistency of images [43]. The following paragraph discusses the most common preprocessing techniques that have been used recently in researches.
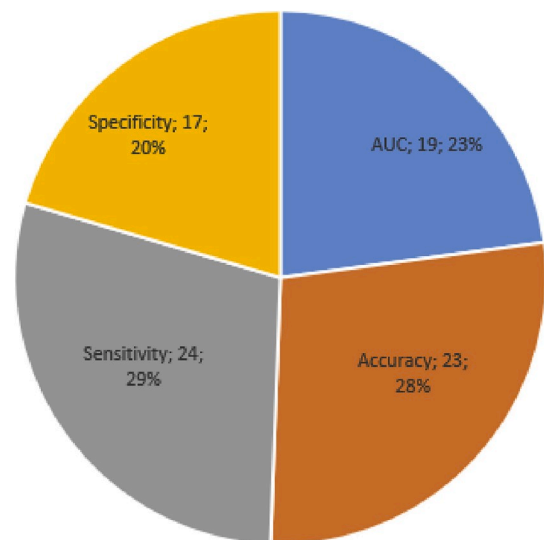

Fig. 5. The percentage of performance measures used in the studies.

Many researchers resized the images to a fixed resolution to be suitable for the used network, as done in Refs. [37,41]. Cropped images were applied to remove the extra regions of the image, while data normalization was used to normalize the images into a similar distribution, as in Ref. [45]. In some works, such as [38], only the green channel of images was extracted due to its high contrast [46], the images were converted into grayscale, such as in Ref. [43].

Noise removal methods include a median filter, Gaussian filter, and NonLocal Means Denoising methods, such as in the works of [38,43,45], respectively. Data augmentation techniques were performed when some image classes were imbalance or to increase the dataset size, such as in Ref. [38,45]. Data augmentation technique include translation, rotation, shearing, flipping, contrast scaling and rescaling. A morphological method was used, such as in Ref. [39], for contrast enhancement. The canny edge method was used for feature extraction in the study of [40].

Table 3
Details of DR datasets.

Dataset	Numbers of images	Normal image	Mild DR	Moderate and severe non-proliferative DR	Proliferative DR	Training Sets	Test Sets	Image Size
DiaretDB1	89 images	27 images	7 images	28 images	27 images	28 images	61 images	1500 × 1152 pixels
Kaggle	88,702 images	–	–	–	–	35,126 images	53,576 images	Different image resolution
DRIVE	40 images	33 images	7 images	–	–	20 images	20 images	565 × 584 pixels
E-ophtha	In e-ophtha EX 82 images and e-ophtha MA 381 images	35 images in e-ophtha EX. 233 images in e-ophtha MA	–	–	–	–	–	Different image resolution
HRF	45 images	15 images	15 images	–	–	–	–	3504 × 2336 pixels
DDR	13,673 images	6266 images	630 images	4713 images	913 images	6835 images	4105 images	Different image resolution
Messidor	1200 images	–	–	–	–	–	–	Different image resolution
Messidor-2	1748 images	–	–	–	–	–	–	Different image resolution
STARE	20 images	10 images	–	–	–	–	–	700 × 605 pixels
CHASE DB1	28 images	–	–	–	–	–	–	1280 × 960 pixels
IDRiD	516 images	–	–	–	–	413 images	103 images	4288 × 2848 pixels
ROC	100 images	–	–	–	–	50 images	50 images	Different image resolution
DR2	435 images	–	–	–	–	–	–	857 × 569 pixels

After preprocessing the images, the images are ready to be used as an input for the DL, which is explained in the next section.

6. Diabetic retinopathy screening systems

Several researches have attempted to automate DR lesions detection and classification using DL. These methods can be categorized according to the classification method used as binary classification, multi-level classification, lesion-based classification, and vessels-based classification. Table 4 summarizes these methods.

6.1. Binary classification

This section summarizes the studies conducted to classify the DR dataset into two classes only. K. Xu et al. [41] automatically classified the images of the Kaggle [26] dataset into normal images or DR images using a CNN. They used 1000 images from the dataset. Data augmentation and resizing to 224*224*3 were performed before feeding the images to the CNN. Data augmentation was used to increase the dataset images by applying several transformations, such as rescaling, rotation, flipping, shearing and translation. The CNN architecture included eight CONV layers, four max-pooling layers and two FC layers. The SoftMax function was applied at the last layer of CNN for classification. This method had an accuracy of 94.5%.

In the study performed by G. Quellec et al. [37], each image was classified as referable DR (refer to moderate stage or more) or non-referable DR (No DR or mild stage) by training three CNNs. The images were taken from three datasets, namely, Kaggle (88,702 image) [26], DiaretDB1 (89 image) [25] and private E-ophtha (107,799 image) [28]. During the preprocessing stage, the images were resized, cropped to 448×448 pixels, normalized, and eroded the FOV by 5%. A large Gaussian filter was used and the augmented data were applied. The used CNNs architectures were pretrained AlexNet [19] and the two networks of o_O solution [48]. MA, HM, soft and hard EX were detected by the CNNs. This study had an area under the ROC curve of 0.954 in Kaggle and 0.949 in E-ophtha.

M. T. Esfahan et al. [22] used a known CNN, which is ResNet34 [49] in their study to classify DR images of the Kaggle dataset [26] into normal or DR image. ResNet34 is one available pretrained CNN architecture on ImageNet database. They applied a set of image preprocessing techniques to improve the quality of images. The image preprocessing included the Gaussian filter, weighted addition and image normalization. The image number was 35000 images and its size was 512×512 pixels. They reported an accuracy of 85% and a sensitivity of 86%.

R. Pires et al. [50] built their own CNN architecture to determine whether an image was referable DR. The proposed CNN contains 16 layers, which is similar to pretrained VGG-16 [51] and o_O team [48]. Two-fold cross-validation and multi-image resolution were used during training. The CNN of the 512×512 image input was trained after initializing the weights by the trained CNN on a smaller image resolution. The drop-out and L2 regularization techniques were applied to the CNN to reduce overfitting. The CNN was trained on the Kaggle dataset [26] and was tested by the Messidor-2 [31] and DR2 dataset. The classes of the training dataset were balanced using data augmentation. The work achieved an area under the ROC curve of 98.2% when testing the Messidor-2.

The study of H. Jiang et al. [52] integrated three pretrained CNN models, namely, Inception V3 [20], Inception-Resnet-V2 [53] and Resnet152 [21] to classify their own dataset as referable DR or non-referable DR. In CNNs training, Adam optimizer was used to update their weights. These models were integrated using the Adaboost algorithm. The dataset of 30,244 images was resized to 520×520 pixels, enhanced and augmented before being fed to the CNNs. The work obtained an accuracy of 88.21% and area under the curve (AUC) of 0.946.

Y. Liu et al. [54] built a weighted paths CNN (WP-CNN) to detect referable DR images. They collected over 60,000 images labeled as referable or non-referable DR and augmented them many times to balance the classes. These images were resized to 299×299 pixels and were normalized before being fed to the CNN. The WP-CNN includes many CONV layers with different kernel sizes in different weighted paths that merged by taking the average. The WP-CNN of 105 layers had a better accuracy than pretrained ResNet [21], SeNet [55] and DenseNet [56] architectures with 94.23% in their dataset and 90.84% in the STARE dataset.

G. Zago et al. [57] detected DR red lesions and DR images based on augmented 65*65 patches using two CNN models. The CNNs used were pretrained VGG16 [51] and a custom CNN, which contains five CONV, five max-polling layers and a FC layer. These models were trained on the DIARETDB1 [25] dataset and tested on the DDR [23], IDRiD [34], Messidor-2, Messidor [58], Kaggle [26], and DIARETDB0 [59] datasets to classify patches into red lesions or non-red lesions. After that, the image with DR or non-DR were classified based on a lesion probability map of test images. The results of this work achieved the best sensitivity of 0.94 and an AUC of 0.912 for the Messidor dataset.

Unfortunately, the researchers who classified DR images into two classes did not consider the five DR stages. The DR stages are important to determine the exact stage of DR to treat the retina with the suitable process and to prevent the deterioration and blindness.

6.2. Multi-level classification

This section reviews the studies in which the DR dataset was classified into many classes. The work by V. Gulshan et al. [60] introduced a method to detect DR and diabetic macular edema (DME) using CNN model. They used Messidor-2 [31] and eyepacs-1 datasets which contain 1748 images and 9963 images, respectively to test the model. These images were first normalized, and the diameter was resized to 299 pixels wide before feeding them to the CNN. They trained 10 CNNs with the pretrained Inception-v3 [20] architecture with a various number of images, and the final result was computed by a linear average function. The images were classified into referable diabetic macular edema, moderate or worse DR, severe or worse DR, or fully gradable. They obtained a specificity of 93% in both datasets and 96.1% and 97.5% in sensitivity for the Messidor-2 and eyepacs-1 datasets, respectively; however, they did not explicitly detect non-DR or the five DR stage images.

M. Abramoff et al. [61] integrated a CNN with an IDX-DR device to detect and to classify DR images. They applied data augmentation to the Messidor-2 [31] dataset, which contains 1748 images. Their various CNNs were integrated using a Random Forest classifier to detect DR lesions as well as retina normal anatomy. The images in this work were classified as no DR, referable DR, or vision threatening DR. They reported an area under the curve of 0.980, a sensitivity of 96.8%, and a specificity of 87.0%. Unfortunately, they considered images of the mild DR stage as no DR, and the five DR stages were not considered.

H. Pratt et al. [42] proposed a method based on a CNN to classify images from the Kaggle dataset [26] into five DR stages. In the preprocessing stage, color normalization and image resizing to 512×512 pixels were performed. Their custom CNN architecture contained 10 CONV layers, eight max-pooling layers, and three FC layers. The Soft-Max function was used as a classifier for 80,000 test images. L2 regularization and dropout methods was used in CNN to reduce overfitting. Their results had a specificity of 95%, an accuracy of 75% and a sensitivity of 30%. Unfortunately, CNN does not detect the lesions in the images, and only one dataset was used to evaluate their CNN.

S. Dutta et al. [43] detected and classified DR images from the Kaggle dataset [26] into five DR stages. They investigated the performance of three networks, the back propagation neural network (BNN), the deep neural network (DNN), and the CNN, using 2000 images. The images were resized to 300×300 pixels and converted into grayscale, and the statistical features were extracted from the RGB images. Furthermore, a set of filters were applied, namely, edge detection, median filter,

Table 4
The methods used for DR detection/classification.

Ref.	DL method	Lesion detection	Dataset (Dataset size)	Performance measure			
				AUC	Accuracy	Sensitivity	Specificity
[60]	CNN (Inception-v3)	No	Messidor-2 (1748) and EyePACS-1 (9963)	–	–	96.1% 97.5%	93.9% 93.4%
[61]	CNN	yes	Messidor-2 (1748)	0.980	–	96.8%	87.0%
[42]	CNN	No	Kaggle (80,000)	–	75%	30%	95%
[41]	CNN	No	Kaggle (1000)	–	94.5%	–	–
[37]	CNN	yes	Kaggle (88,702), DiaretDB1 (89) and E-ophtha (107,799).	0.954 – 0.949	–	–	–
[38]	CNN	Red lesion only	DIARETDB1 (89), E-Ophtha (381) and MESSIDOR (1200)	CPM = 0.4874 for DIARETDB1 and CPM = 0.3683 for e-ophtha		0.4883 0.3680	–
[22]	CNN-ResNet34	No	Kaggle (35000)	–	85%	86%	–
[43]	DNN, CNN (VGGNET architecture), BNN	No	Kaggle (2000)	–	BNN = 42% DNN = 86.3% CNN = 78.3%	–	–
[44]	CNN (InceptionNet V3, AlexNet and VGG16)	No	Kaggle (166)	–	AlexNet = 37.43%, VGG16 = 50.03%, and InceptionNet V3 = 63.23%	–	–
[45]	CNN (AlexNet, VggNet, GoogleNet and ResNet)	No	Kaggle (35,126)	The higher is VggNet-s (0.9786)	The higher is VggNet-s (95.68%)	VggNet-16 achieved higher result (90.78%).	The higher is VggNet-s (97.43%)
[39]	CNN	MA only	E- Ophtha (381), ROC (100) and DIARETDB1 (89)	0.562 0.193 0.392	–	–	–
[40]	CNN	EX Only	DiaretDB0 (130), DiaretDB1 (89), and DrimDB (125).	–	99.17 98.53 99.18	100 99.2 100	98.41 97.97 98.44
[46]	Fully CNN	No	STARE (20), HRF (45), DRIVE (40) and CHASE DB1 (28).	0.9801 0.9701 0.9787 0.9752	0.9628 0.9608 0.9634 0.9664	0.8090 0.7762 0.7941 0.7571	0.9770 0.9760 0.9870 0.9823
[47]	Fully CNN	No	STARE (20), DRIVE (40) and CHASE_DB1 (28).	0.9905 0.9821 0.9855	0.9694 0.9576 0.9653	0.8315 0.8039 0.7779	0.9858 0.9804 0.9864
[63]	CNN (AlexNet, VggNet16, custom CNN)	No	MESSIDOR (1200)	–	98.15%	98.94%	97.87%
[65]	CNN (ResNet50, InceptionV3, InceptionResNetV2, Xception and DenseNets)	No	Their Own dataset (13767)	–	96.5%	98.1%	98.9%
[50]	CNN	No	Messidor-2 (1748), Kaggle (88,702) and DR2 (520)	98.2% – 98%	–	–	–
[68]	CNN (AlexNet)	No	Kaggle (22,700) and IDRiD (516)	–	90.07%	–	–
[52]	CNN (Inception V3, Inception-Resnet-V2 and Resnet152)	No	Their Own dataset (30244)	0.946	88.21%	85.57%	90.85%
[54]	CNN (WP-CNN, ResNet, SeNet and DenseNet)	No	Their Own dataset (60,000), and STARE (131)	0.9823 0.951	94.23% 90.84%	90.94%	95.7–4%
[74]	CNN (improved LeNet, U-net)	Red lesion only	DIARETDB1 (89)	CPM = 0.4823		48.71%	–
[57]	CNN (VGG16, custom CNN)	Red lesion only	DIARETDB1 (89), DIARETDB0 (130), Kaggle (15,919), Messidor (1200), Messidor-2 (874), IDRiD (103) and DDR (4105)	– 0.786 0.764 0.912 – 0.818 0.848	– 0.821 0.911 0.94 – 0.841 0.891	–	
[23]	CNN (GoogLeNet, ResNet-18, DenseNet-121, VGG-16 and SE-BN-Inception)	No	DDR (13,673)	–	0.8284	–	–
[69]	CNN (modified Alexnet)	No	Messidor (1190)	–	96.35	92.35	97.45
[78]	CNN	No	HRF (45) and DRIVE (40)	0.894	93.94%	–	–
[81]	CNN (ResNet-101)	No	DRIVE (40)	0.9732	0.951	0.793	0.974
[75]	CNN	EX only	E-optha (82) and HEI-MED (169)	0.9644 0.9323	–	0.8990 0.9477	–
[77]	Deep residual network	EX only	E-optha (82) and HEI-MED (169)	0.9647 0.9709	–	0.9227 0.9255	–
[82]	CNN	No	DRIVE (40) and STARE (20)				

(*continued on next page*)

6

Table 4 (*continued*)

Ref.	DL method	Lesion detection	Dataset (Dataset size)	Performance measure			
				AUC	Accuracy	Sensitivity	Specificity
				0.9822	0.9685	0.7439	0.99
				0.9868	0.9735	0.8196	0.9871
[83]	CNN	No	DRIVE (40), STARE (20) and CHASE (28).	98.30%	95.82%	79.96% 79.63%	98.13%
				98.75%	96.72%	80.03%	98.63%
				98.94%	96.88%		98.80%
[84]	CNN	No	DRIVE (40) and CHASE_DB1 (28).	0.9560	0.9580	0.8639	0.9665
				0.9577	0.9601	0.8778	0.9680
[70]	CNN	Red lesion only	Their dataset (9194) and Messidor (1200)	–	92.95	99.39%	99.93%
				0.972	–	92.59%	96.20%
[72]	CNN (ResNet50)	No	Messidor (1200) and IDRiD (516)	96.3%	92.6%	92%	–
				–	65.1%	–	

morphological processing, and binary conversion, before being fed into the networks. Pretrained VGG16 [51] was used as the CNN architecture, which includes 16 CONV layers and 4 max pooling layers and three FC layers while the DNN includes three FC layers. Their results shown that the DNN outperforms the CNN and the BNN. Unfortunately, few images were used for networks training, and thus the networks could not learn more features. Also, only one dataset was used to evaluate their study.

X. Wang et al. [44] studied the performance of the three available pretrained architectures of CNN, VGG16 [51], AlexNet [19] and InceptionNet V3 [20], to detect the five DR stages in the Kaggle [26] dataset. The images were resized to 224×224 pixels for VGG16, 227×227 pixels for AlexNet, and 299×299 pixels for InceptionNet V3 at the preprocessing stage. The dataset only contains 166 images. They reported an average accuracy of 50.03% in VGG16, 37.43% in AlexNet and 63.23% in InceptionNet V3; however, they trained the networks with limited number of images, which could prevent the CNN from learning more features and the images required more preprocessing functions to improve them. Also, only one dataset was used to evaluate their study.

The performance of four available pretrained architectures of the CNN was investigated in Ref. [45]: AlexNet [19], ResNet [21], GoogleNet [62] and VggNet [51]. These architectures were trained to detect the five DR stages from the Kaggle [26] dataset, which contains 35,126 images. Transfer learning these CNNs was done by fine tuning the last FC layer and hyperparameter. During the preprocessing stage, the images were augmented, cropped, normalized and the NonLocal Means Denoising function was applied. This study achieved an accuracy of 95.68%, AUC of 0.9786 and a specificity of 97.43% for VggNet-s, which had a higher accuracy, specificity, and an AUC than the other architectures. The use of more than one dataset makes a system more reliable and able to generalize [83]. Unfortunately, the study only included one dataset and their method does not detect the DR lesions.

Mobeen-ur-Rehman et al. [63] detected the DR levels of the MESSIDOR dataset [31] using their custom CNN architecture and pretrained models, including AlexNet [19], VGG-16 [51] and SqueezeNet [64]. This dataset contains 1200 images classified into four DR stages. The images were cropped, resized to a fixed size, which was 244×244 pixel, and enhanced by applying the histogram equalization (HE) method at the pre-processing stage. The custom CNN includes five layers: two CONV layers, two max-pooling layers, and three FC layers. They reported the best accuracy of 98.15%, specificity of 97.87% and sensitivity of 98.94% by their custom CNN. Unfortunately, only one dataset was used to evaluate their CNN and does not detect the DR lesions.

W. Zhang et al. [65] proposed a system to detect the DR of their own dataset. The dataset includes 13,767 images, which are grouped into four classes. These images were cropped, resized to the required size of each network, and improved by applying HE and adaptive HE. In addition, the size of the training images was enlarged by data augmentation, and the contrast was improved by a contrast stretching algorithm that is used for dark images. They finetuned pretrained CNN architectures: ResNet50 [66], InceptionV3 [20], InceptionResNetV2 [53], Xception [67], and DenseNets [56] to detect the DR. Their approach involved training the added new FC layers on top of these CNNs. After that, they finetuned some layers of the CNNs to retrain it. Lastly, the strong models were integrated. This approach achieved an accuracy of 96.5%, a specificity of 98.9% and a sensitivity of 98.1%. Unfortunately, CNNs do not detect the lesions in the images and only one private dataset was used to evaluate their method.

B. Harangi et al. [68] integrated the available pretrained AlexNet [19] and the hand-crafted features to classify the five DR stages. The CNN was trained by the Kaggle dataset [26] and tested by the IDRiD [34]. The obtained accuracy for this study was 90.07%. Unfortunately, the work does not detect the lesions in the images and only one dataset was used to test their method.

T. Li et al. [23] detected DR stages in their dataset (DDR) by fine-tuning the GoogLeNet [62], ResNet-18 [21], DenseNet-121 [56], VGG-16 [51], and SE-BN-Inception [55] available pretrained networks. Their dataset includes 13,673 fundus images. During preprocessing, the images were cropped, resized to 224×224 pixels, augmented and resampled to balance the classes. The SE-BN-Inception network obtained the best accuracy at 0.8284. Unfortunately, the work does not detect the lesions in the images and only one dataset was used to test their method.

T. Shanthi and R. Sabeenian [69] detected the DR stages of the Messidor dataset [31] using a pretrained architecture Alexnet [19]. The images were resized, and the green channel was extracted before being fed into the CNN. This CNN achieved an accuracy of 96.35. Unfortunately, the work does not detect the lesions in the images and only one dataset and architecture were used to test their method.

J. Wang et al. [70] modified a R-FCN method [71] to detected DR stages in their private dataset and the public Messidor dataset [58]. Moreover, they detected MA and HM in their dataset. They modified the R-FCN by adding a feature pyramid network and also adding five region proposal networks rather than one to the method. The lesion images were augmented for training. The obtained sensitivity for detecting DR stages were 99.39% and 92.59% in their dataset and the Messidor dataset, respectively. They reported a PASCAL-VOC AP of 92.15 in lesion detection. Unfortunately, the study only evaluated the method on one public dataset and only detected HM and MA without detecting EX.

X. Li et al. [72] classified the public Messidor [58] dataset into referable or non-referable images and classified the public IDRiD dataset [34] into five DR stages and three DME stages by using the ResNet50 [21] and four attention modules. The features extracted by ResNet50 used as the inputs for the first two attention modules to select one disease features. The first two attention modules contain average pooling layers, max-pooling layers, multiplication layers, concatenation layer, CONV layer and FC layers while the next two attention modules contain FC and multiplication layers. Data augmentation, normalization and resizing were performed before feeding the images to the CNN. This work achieved a sensitivity of 92%, an AUC of 96.3% and an accuracy of 92.6% for the Messidor dataset and an accuracy of 65.1% for the IDRiD.

Unfortunately, the study does not detect the lesions in the images.

6.3. Lesion-based classification

This section summarizes the works performed to detect and to classify certain types of DR lesions. For example, J. Orlando et al. [38] detected only red lesions in DR images by incorporating DL methods with domain knowledge for feature learning. Then, the images were classified by applying the Random Forest method. The images of the MESSIDOR [58], E-ophtha [73] and DIARETDB1 [25] datasets were processed by extracting the green band and expanding the FOV, and applying a Gaussian filter, r-polynomial transformation, thresholding operation and, many morphological closing functions. Next, red lesion patches were resized to 32*32 pixels and were augmented for CNN training. The datasets contain 89 images, 381 images and 1200 images in DIARETDB1, E-ophtha and MESSIDOR, respectively. Their custom CNN contains four CONV layers, three pooling layers and one FC layer. They achieved a Competition Metric (CPM) of 0.4874 and 0.3683 for the DIARETDB1 and the E-ophtha datasets, respectively.

P. Chudzik et al. [39] used custom CNN architecture to detect MA from DR images. Three datasets were used in this study: ROC [35] (100 images), E-ophtha [73] (381 images), and DIARETDB1 [25] (89 images). These datasets were processed by extracting the green plane and then performing cropping, resizing, applying Otsu thresholding to generate a mask, and utilizing a weighted sum and morphological functions. Next, MA patches were extracted, and random transformations were applied. The used CNN includes 18 CONV layers, and each CONV layer is followed by a batch normalization layer, three max-pooling layers, three simple up-sampling layers, and four skip connections between both paths. They reported a ROC score of 0.355.

The system proposed by Refs. [40], detected the exudates from DR images using the custom CNN with Circular Hough Transformation (CHT). They used three public datasets: the DiaretDB0 dataset includes 130 images, the DiaretDB1 dataset contains 89 images and the DrimDB dataset has 125 images. All the datasets were converted into grayscale. Then, Canny edge detection and adaptive histogram equalization functions were applied. Next, the optical disc was detected by CHT and then removed from the images. The 1152*1152 pixels of the images were fed into the custom CNN, which contains three CONV layers, three max pooling layers, and an FC layer that uses SoftMax as a classifier. The accuracies of detecting exudates were 99.17, 98.53, and 99.18 for DiaretDB0, DiaretDB1, and DrimDB, respectively.

Y. Yan et al. [74] detected DR red lesions in the DIARETDB1 [25] dataset by integrating the features of a handcrafted and improved pretrained LeNet architecture using a Random Forest classifier. The green channel of the images was cropped, and they were enhanced by CLAHE. Also, noise was removed by the Gaussian filter, and a morphological method was used. After that, the blood vessels were segmented from images by applying the U-net CNN architecture. The improved LeNet architecture includes four CONV layers, three max-pooling layers, and one FC layer. This work achieved a sensitivity of 48.71% in red lesions detection.

H. Wang et al. [75] detected hard exudate lesion in the E-ophtha dataset [28] and the HEI-MED dataset [76] by integrating the features of a handcrafted and custom CNN using a Random Forest classifier. These datasets were processed by performing cropping, color normalizing, modifying a camera aperture and detecting the candidates by using morphological construction and dynamic thresholding. After that, patches of size 32*32 are collected and augmented. The custom CNN includes three CONV and three pooling layers and a FC layer to detect the patches features. This work achieved a sensitivity of 0.8990 and 0.9477 and an AUC of 0.9644 and 0.9323 for the E-ophtha and HEI-MED datasets, respectively.

J. Mo et al. [77] detected exudate lesions in the public available E-optha [28] and the HEI-MED [76] datasets by segmenting and classifying the exudates using deep residual network. The exudates were segmented using a fully convolutional residual network which contains up-sampling and down-sampling modules. After that, the exudates were classified using a deep residual network which includes one CONV layer, one max-pooling layer and 5 residual blocks. The down-sampling module includes CONV layer followed by a max pooling layer and 12 residual blocks while the up-sampling module comprises CONV and deconvolutional layers to enlarge the image as the input image. The residual block includes three CONV layers and three batch normalization layers. This work achieved a sensitivity of 0.9227 and 0.9255 and an AUC of 0.9647 and 0.9709 for the E-optha and HEI-MED datasets, respectively.

Unfortunately, these studies detected only some DR lesions without considering the five DR stages. Furthermore, they used a limited number of images for DL methods.

6.4. Vessel-based classification

Vessel segmentation is used to diagnosis and to evaluate the progress of retinal diseases, such as glaucoma, DR and hypertension. Many studies have been conducted to investigate vessel segmentation as part of DR detection. DR lesions remain in the image after the vessels have been extracted. Therefore, detecting the remaining lesions in the images lead to detect and classify DR images. The study in Ref. [74] detected the red lesions after vessels were extracted. Some studies on vessel segmentation used DL methods, which is reviewed in this section.

Sunil et al. [78] used a modified CNN of pretrained DEEPLAB-COCO-LARGEFOV [79] to extract the retinal blood vessels from RGB retina images. They extracted 512×512 image patches from the dataset and then fed them to the CNN. After that, they applied a threshold to binarize the images. The CNN includes eight CONV layers and three max-pooling layers. The HRF [30], DRIVE [29] datasets were used to evaluate the method. They reported an accuracy of 93.94% and an area under the ROC of 0.894.

The study conducted by Refs. [46] used fully CNN to segments the blood vessels in RGB retina images. The images of the STARE [32], HRF [30], DRIVE [29] and CHASE DB1 [33] datasets were preprocessed by applying morphological methods, flipped horizontally, adjusted to different intensities, and cropped into patches. Then they were fed to the CNN for segmentation and to condition random field model [80] to consider the non-local correlations during segmentation. After that, the vessel map was rebuilt, and morphological operations were applied. Their CNN contains 16 CONV layers and five dilated CONV layers. The STARE, HRF, DRIVE and CHASE DB1 datasets contain 20, 45, 40, and 28 images, respectively were used. An accuracy of 0.9634, 0.9628, 0.9608 and 0.9664 was achieved for the DRIVE, STARE, HRF and CHASE DB1, respectively.

The work conducted by Refs. [47] included the Stationary Wavelet Transform (SWT) with a fully CNN to extract the vessels from the images. The STARE (20 images) [32], DRIVE (40 images) [29] and CHASE_DB1 (28 images) [33] datasets were preprocessed by extracting the green channel and normalizing images, and SWT was applied. Next, the patches were extracted and augmented. Then, the patches were fed to the CNN which includes CONV layers, max-pooling layers, crop layer, and SoftMax classifier and up-sampling layer that return the feature maps to the previous dimensions. The results of this study reached an AUC of 0.9905, 0.9821 and 0.9855 for the STARE, DRIVE and CHASE_DB1 datasets, respectively.

Cam-Hao et al. [81] extracted retinal vessels from the DRIVE dataset [29]. They selected four feature maps from the pretrained ResNet-101 [21] network and then combined each feature map with its neighbor. After that, the feature map outputs were also combined until one feature map was obtained. Next, each round of the best resolution feature maps was concatenated. They augmented the training images before being fed to the network. They achieved a sensitivity of 0.793, an accuracy of 0.951, a specificity of 0.9741 and an AUC of 0.9732.

Ü. Budak et al. [82] extracted retinal vessels from the DRIVE [29]

and STARE [32] public datasets using custom CNN architecture. The custom CNN includes three blocks of concatenated encoder-decoder and two CONV layers between them. Each block contains eight CONV layers, eight Batch Normalization layers, two max-unpooling layers and two max-pooling layers. They cropped images, extracted and augmented patches before feeding them into the CNN for training. They reported an accuracy of 0.9685, 0.9735 an AUC of 0.9822 and 0.9868 for the DRIVE and STARE datasets, respectively.

Y. Wu et al. [83] used a custom CNN to extract the retinal blood vessels from the DRIVE [29], STARE [32] and CHASE [33] public datasets. They converted RGB images to grayscale, normalized and enhanced them by CLAHE. After that, they extracted and augmented 48 × 48 patches from the datasets and fed them to the CNN. The CNN includes of two networks that have encoder–decoder structure and skip connections. The encoder–decoder structure has CONV layers, Batch Normalization layers, concatenation layers and dropout layers. They achieved an accuracy of 95.82%, 96.72% and 96.88%, and an AUC of 98.30%, 98.75% and 98.94% for the DRIVE, STARE and CHASE, respectively.

C. Tian et al. [84] extracted retinal vessels from the DRIVE [29] and CHASE_DB1 [33] public datasets using custom multi-path CNN architecture and Gaussian matched filter. They obtained a high-frequency and a low-frequency images from Gaussian filter. Then they constructed CNN path for low-frequency images that composed of CONV layers, down-sampling and up-sampling modules and constructed other CNN path for high-frequency images that composed of two CONV layers and seven encoder–decoder blocks. The encoder–decoder structure has dilated CONV layers, down-sampling and up-sampling modules. The segmentation maps were extracted from the two paths and then merged them for final segmentation results. They achieved an accuracy of 0.9580 and 0.9601 and an AUC of 0.9560 and 0.9577 for the DRIVE and CHASE_DB1, respectively.

Unfortunately, these works only considered segmented vessels and did not detect DR stages or DR lesions.

7. Discussion section

The current study reviewed 33 papers. All of the studies mentioned in the current work manipulated the diabetic retinopathy screening system using deep learning techniques. The need for reliable diabetic retinopathy screening systems became a critical issue recently due to the increase in the number of diabetic patients. Using DL in DR detection and classification overcomes the problem of selecting reliable featured for ML; on the other hand, it needs a huge data size for training. Most studies used data augmentation to increase the number of images and overcoming overfitting on training stage. The studies covered on the current work 94% of them used public datasets, 59% of them used a combination of two or more public datasets to overcome the problem of data size and to evaluate the DL methods on many datasets as shown in Fig. 6.
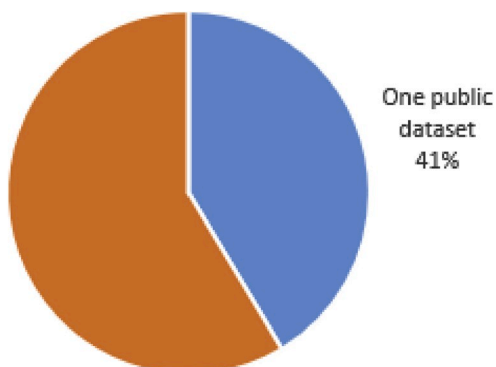
One of the limitations of the usage of deep learning with medical field faces is the size of the datasets needed to train the DL systems, as DL is required large amount of data. The results of DL systems depend heavily on the size of the training data as much as its quality and balance its classes. So, the current public datasets sizes need to be increased, while the big size one like public Kaggle dataset needs to be refined to eliminates miss labeled and low-quality data.

The covered studies here varied in using DL techniques. They differ in the number of studies that built their own CNN structure, and those who preferred to use the existing structures, such as VGG, ResNet, or AlexNet, with transfer learning is slightly small. Building a new CNN architecture from scratch needs a lot of effort and time consuming while using transfer learning is much easier and speed up the process of structuring and developing new architecture. On the other hand, it is notable that the accuracy of the system which built their own CNN structure is higher than those using the existing structures. This point needs to be focused by the researchers and more studies should be conducted to judge among the two trends.

Most of the studies covered here (73%) only classified the fundus input image to DR non-DR, while 27% classified input to one or more stages as shown in Fig. 7. On the other hands 70% of the current studies didn't detect the affected lesions while, 30% of them detected the affected lesions. Among them, only 6% of the studies succeeded in classifying images and detecting the type of the affected lesion on the retain image as shown in Fig. 8. The existence of a reliable DR screening system capable of detecting different lesions types and DR stages leads to an effective follow up system for DR patients, which averted the danger of losing sight. The gap that needed to be covered is the existence of systems that could determine the five DR stages with high accuracy as well as detecting DR lesions. This point could be considered as the current challenge for researchers for further investigations.

8. Conclusion

Automated screening systems significantly reduce the time required to determine diagnoses, saving effort and costs for ophthalmologists and result in the timely treatment of patients. Automated systems for DR detection play an important role in detecting DR at an early stage. The DR stages are based on the type of lesions that appear on the retina. This article has reviewed the most recent automated systems of diabetic retinopathy detection and classification that used deep learning techniques. The common fundus DR datasets that are publicly available have been described, and deep-learning techniques have been briefly explained. Most researchers have used the CNN for the classification and the detection of the DR images due to its efficiency. This review has also discussed the useful techniques that can be utilized to detect and to classify DR using DL.
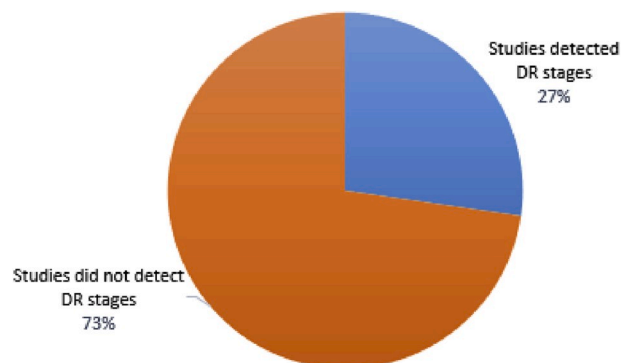
Fig. 6. The percentage of studies that used one or more public datasets.

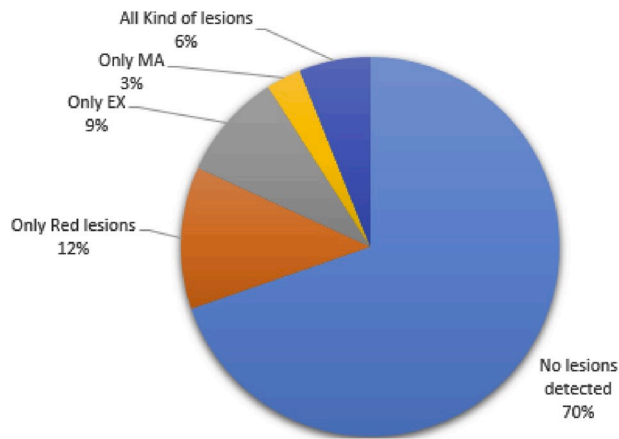**Fig. 7.** The percentage of studies that detected DR stages.

Fig. 8. The percentage of studies that detected DR lesions.

Ethical statement

This work did not receive any grant from funding agencies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank King Abdul-Aziz University for encouraging us to do this work.

References

[1] Taylor R, Batey D. Handbook of retinal screening in diabetes:diagnosis and management. second ed. John Wiley & Sons, Ltd Wiley-Blackwell; 2012.

[2] International diabetes federation - what is diabetes [Online]. Available, https://www.idf.org/aboutdiabetes/what-is-diabetes.html.

[3] American academy of ophthalmology-what is diabetic retinopathy? [Online]. Available, https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopathy.

[4] Bourne RR, et al. Causes of vision loss worldwide, 1990-2010: a systematic analysis. Lancet Global Health 2013;1(6):339–49.

[5] Harper CA, Keeffe JE. Diabetic retinopathy management guidelines. Expet Rev Ophthalmol 2012;7(5):417–39.

[6] E. T. D. R. S. R. GROUP. Grading diabetic retinopa thy from stereoscopic color fundus photographs- an extension of the modified Airlie House classification. Ophthalmology 1991;98(5):786–806.

[7] Scanlon PH, Wilkinson CP, Aldington SJ, Matthews DR. A Practical manual of diabetic retinopathy management. first ed. Wiley-Blackwell; 2009.

[8] Dubow M, et al. Classification of human retinal microaneurysms using adaptive optics scanning light ophthalmoscope fluorescein angiography. Investig Ophthalmol Vis Sci 2014;55(3):1299–309.

[9] Bandello F, Zarbin MA, Lattanzio R, Zucchiatti I. Clinical strategies in the management of diabetic retinopathy. second ed. Springer; 2019.

[10] Scotland GS, et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol 2010; 94(6):712–9.

[11] deng Li. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf Process 2014;3(2):1–29.

[12] V Vasilakos A, Tang Y, Yao Y. Neural networks for computer-aided diagnosis in medicine : a review. Neurocomputing 2016;216:700–8.

[13] Wilkinson CP, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Am Acad Ophthalmol 2003;110(9): 1677–82.

[14] Chen XW, Lin X. Big data deep learning: challenges and perspectives. IEEE Access 2014;2:514–25.

[15] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: a review. Neurocomputing 2016;187:27–48.

[16] Deng L, Yu D. Deep learning: methods and applications. Found Trends® Signal Process 2014;7(3–4):197–387.

[17] Bakator M, Radosav D. Deep learning and medical diagnosis: a review of literature. Multimodal Technol Interact 2018;2(3):47.

[18] Mookiah M, Acharya UR, Kuang C, Min C, Ng EYK, Laude A. Computer-aided diagnosis of diabetic retinopathy : a review. Comput Biol Med 2013;43:2136–55.

[19] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6).

[20] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: IEEE Conference on computer Vision and pattern recognition; 2016. p. 2818–26.

[21] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on computer Vision and pattern recognition; 2016. p. 770–8.

[22] Esfahani MT, Ghaderi M, Kafiyeh R. Classification of diabetic and normal fundus images using new deep learning method. Leonardo Electron J Pract Technol 2018; 17(32):233–48.

[23] Li T, Gao Y, Wang K, Guo S, Liu H, Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Inf Sci 2019;501:511–22.

[24] Abràmoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. IEEE Rev Biomed Eng 2010;3:169–208.

[25] Kauppi T, et al. The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: Proceedings of the British machine vision conference 2007; 2007. p. 1–10.

[26] Kaggle dataset [Online]. Available, https://kaggle.com/c/diabetic-retinopathy-detection.

[27] Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using Deep Learning. AMIA Joint Summits on Translational Science 2018;2017:147–55.

[28] Decencière E, et al. TeleOphta : machine learning and image processing methods for teleophthalmology. IRBM 2013;34(2):196–203.

[29] Staal JJ, Abràmoff MD, Niemeijer M, Viergever MA, van Ginneken B. Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imag 2004;23 (4):501–9.

[30] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," Int J Biomed Imag, vol. 2013, 2013.

[31] Decenciere E, et al. Feedback on a publicly distributed image database: the messidor database. Image Anal Stereol 2014;33(3):231–4.

[32] Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. IEEE Trans Med Imag 2000;19(3):203–10.

[33] Owen CG, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. Investig Ophthalmol Vis Sci 2009;50(5):2004–10.

[34] Porwal P, et al. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data 2018;3(3).

[35] ROC dataset [Online]. Available, http://roc.healthcare.uiowa.edu.

[36] DR2 [Online]. Available, https://figshare.com/articles/Advancing_Bag_of_Visual_Words_Representations_for_Lesion_Classification_in_Retinal_Images/953671.

[37] Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. Med Image Anal 2017;39:178–93.

[38] Orlando JI, Prokofyeva E, del Fresno M, Blaschko MB. An ensemble deep learning based approach for red lesion detection in fundus images. Comput Methods Progr Biomed 2018;153:115–27.

[39] Chudzik P, Majumdar S, Calivá F, Al-Diri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. Comput Methods Progr Biomed 2018; 158:185–92.

[40] Adem K. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. Expert Syst Appl 2018;114: 289–95.

[41] Xu K, Feng D, Mi H. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. Molecules 2017;22(12): 2054.

[42] Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. Procedia Comput Sci 2016;90:200–5.

[43] Dutta S, Manideep BC, Basha SM, Caytiles RD, Iyengar NCSN. Classification of diabetic retinopathy images by using deep learning models. Int J Grid Distr Comput 2018;11(1):99–106.

[44] Wang X, Lu Y, Wang Y, Chen WB. Diabetic retinopathy stage classification using convolutional neural networks. In: International Conference on information Reuse and Integration for data science; 2018. p. 465–71.

[45] Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electr Eng 2018;72:274–82.

[46] Lu J, Xu Y, Chen M, Luo Y. A coarse-to-fine fully convolutional neural network for fundus vessel segmentation. Symmetry 2018;10(11):607.

[47] Oliveira A, Pereira S, Silva CA. Retinal vessel segmentation based on fully convolutional neural networks. Expert Syst Appl 2018;112:229–42.

[48] Team o-O solution [Online]. Available, https://www.kaggle.com/c/diabetic-retinopathydetection/discussion/15617.

[49] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: The thirty-first AAAI conference on artificial intelligence; 2016. p. 4278–84.

[50] Pires R, Avila S, Wainer J, Valle E, Abramoff MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. Artif Intell Med 2019;96:93–106.

[51] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: International Conference on learning representations; 2015.

[52] Jiang H, Yang K, Gao M, Zhang D, Ma H, Qian W. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: 2019 41st Annual International conference of the IEEE engineering in medicine and biology society (EMBC); 2019. p. 2045–8.

[53] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on artificial intelligence; 2017. p. 4278–84.

[54] Liu YP, Li Z, Xu C, Li J, Liang R. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. Artif Intell Med 2019;99:101694.

[55] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7132–41.

[56] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: The IEEE Conference on computer Vision and pattern recognition (CVPR); 2017. p. 4700–8.

[57] Zago GT, Andreão RV, Dorizzi B, Teatini Salles EO. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. Comput Biol Med 2019;116:103537.

[58] Messidor dataset [Online]. Available, http://messidor.crihan.fr.

[59] Kauppi T, et al. DIARETDB0 : evaluation database and methodology for diabetic retinopathy algorithms. In: Machine Vision and pattern recognition research group; 2006. p. 1–17.

[60] Gulshan V, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Am Med Assoc 2016;316:2402–10.

[61] Abràmoff MD, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investig Ophthalmol Vis Sci 2016;57(13):5200–6.

[62] Szegedy C, et al. Going deeper with convolutions. In: IEEE Conference on computer Vision and pattern recognition (CVPR); 2015. p. 1–9.

[63] Mobeen-Ur-Rehman, Khan SH, Abbas Z, Danish Rizvi SM. Classification of diabetic retinopathy images based on customised CNN architecture. In: Proceedings - 2019 Amity International Conference on artificial intelligence, AICAI 2019; 2019. p. 244–8.

[64] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. 2016. p. 1–13. arXiv:1602.07360.

[65] Zhang W, et al. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowl Base Syst 2019;175:12–25.

[66] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 2015;37(9):1904–16.

[67] Chollet F. Xception: deep learning with depthwise separable convolutions. In: The IEEE Conference on computer Vision and pattern recognition (CVPR); 2017. p. 1251–8.

[68] Harangi B, Toth J, Baran A, Hajdu A. Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. In: 2019 41st annual International Conference of the IEEE Engineering in Medicine and biology society (EMBC); 2019. p. 2699–702.

[69] Shanthi T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. Comput Electr Eng 2019;76:56–64.

[70] Wang J, Luo J, Liu B, Feng R, Lu L, Zou H. Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm. IET Comput Vis 2020;14(1):1–8.

[71] Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. In: Advances in neural information processing systems; 2016. p. 379–87.

[72] Li X, Hu X, Yu L, Zhu L, Fu C-W, Heng P-A. CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE Trans Med Imag 2020;39(5):1483–93.

[73] E-Ophtha dataset [Online]. Available, http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html.

[74] Yan Y, Gong J, Liu Y. A novel deep learning method for red lesions detection using hybrid feature. In: Proceedings of the 31st Chinese Control and decision conference, CCDC 2019; 2019. p. 2287–92.

[75] Wang H, et al. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening. Comput Methods Progr Biomed 2020;191:105398.

[76] Giancardo L, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal 2012;16(1):216–26.

[77] Mo J, Zhang L, Feng Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. Neurocomputing 2018;290:161–71.

[78] Vengalil SK, Sinha N, Kruthiventi SSS, Babu RV. Customizing CNNs for blood vessel segmentation from fundus images. In: International Conference on signal Processing and communications, SPCOM 2016; 2016. p. 1–4.

[79] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected crfs. 2014. arXiv preprint arXiv:1412.7062.

[80] Zheng S, et al. Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1529–37.

[81] Hua CH, Huynh-The T, Lee S. Retinal vessel segmentation using round-wise features aggregation on bracket-shaped convolutional neural networks. In: Proceedings of the annual International Conference of the IEEE Engineering in Medicine and biology society, EMBS; 2019. p. 36–9.

[82] Budak Ü, Cömert Z, Çıbuk M, Şengür A. DCCMED-Net: densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images. Med Hypotheses 2020;134.

[83] Wu Y, Xia Y, Song Y, Zhang Y, Cai W. NFN + : a novel network followed network for retinal vessel segmentation. Neural Network 2020;126:153–62.

[84] Tian C, Fang T, Fan Y, Wu W. Multi-path convolutional neural network in fundus segmentation of blood vessels. Biocybern Biomed Eng 2020;40(2):583–95.