Review Article GU J Sci 36(3): 1140-1157 (2023) DOI: 10.35378/gujs.1081546

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

A Systematic Review of Transfer Learning-Based Approaches for Diabetic Retinopathy Detection

Burcu OLTU[1,*], Busra Kubra KARACA[1], Hamit ERDEM[2], Atilla OZGUR[3]

[1]Baskent University, Department of Biomedical Engineering, 06790, Ankara, Turkey
[2]Baskent University, Department of Electrical and Electronics Engineering, 06790, Ankara, Turkey
[3]SMS digital GmbH, 40237, Düsseldorf, Germany

Highlights

- This study reviews 43 articles published in the period between 2016-2021.
- 29 CNN based architectures and the most used 13 DR data sets have been evaluated.
- Obtained result summarized in 10 tables and 4 figures.
- Tables and figures are interpreted in detail.
- The findings help researchers to select databases, algorithms, performance metrics for DR detection.

Article Info

Received: 2 Mar 2022
Accepted: 20 June 2022

Keywords

Deep learning
Convolutional neural networks
Diabetic retinopathy
Transfer learning

Abstract

Diabetic retinopathy, which is extreme visual blindness due to diabetes, has become an alarming issue worldwide. Early and accurate detection of DR is necessary to prevent the progression and reduce the risk of blindness. Recently, many approaches for DR detection have been proposed in the literature. Among them, deep neural networks (DNNs), especially Convolutional Neural Network (CNN) models, have become the most offered approach. However, designing and training new CNN architectures from scratch is a troublesome and labor-intensive task, particularly for medical images. Moreover, it requires training tremendous amounts of parameters. Therefore, transfer learning approaches as pre-trained models have become more prevalent in the last few years. Accordingly, in this study, 43 publications based on DNN and Transfer Learning approaches for DR detection between 2016 and 2021 are reviewed. The reviewed papers are summarized in 4 figures and 10 tables that present detailed information about 29 pre-trained CNN models, 13 DR data sets, and standard performance metrics.

1. INTRODUCTION

Diabetic Mellitus, commonly known as diabetes, is a metabolic disorder that occurs when the insulin hormone cannot be produced enough or cannot be used effectively [1]. It is one of the most common diseases in the modern world. According to a report by World Health Organization [2], more than 400 million people have diabetes in the world. By 2024, this number is expected to increase to 552 million [3]. Diabetes is an important cause for death, blindness, and amputation, according to the same report [2]. Diabetic retinopathy (DR), on the other hand is among the most common eye diseases caused by diabetes. DR, that is due to chronic diabetes which cause retinal capillary damage, is the leading cause of blindness [4]. Early and accurate detection is mandatory for preventing disease progression and reducing the risk of vision loss, also it plays an important role in optimal treatment [1]. Traditionally, the diagnosis of DR is based on manual examinations of fundus images or optical coherence tomography [5]. However, these manual examinations are time consuming and highly dependent on the experience of the clinicians. Therefore, this procedure is challenging and is open to misdiagnosis [4].

In recent years, applications of Artificial intelligence (AI), machine learning (ML), and Deep Neural Networks (DNNs) have become prominent in DR detection and classification [5].

*Corresponding author, e-mail: boltu@baskent.edu.tr

Many ML based approaches have been applied for detecting, analyzing, and classifying DR images [4, 6-8]. DNNs are a subset of ML methods, and the convolutional neural networks (CNNs) are a subset of DNN methods. CNNs, which simulate the human visual system and are capable of extracting features automatically, have been successfully used for image-based classification and pattern recognition problems [9, 10]. DNNs and CNNs have also been applied to many different problems in health care [10–12]. CNNs as a supervised learning architecture is the most applied method in image-based classification [6, 13]. CNNs were also used to tackle the DR detection task during the last decade [13–15].

CNNs have superior performance, but in order to train them, a lot of time and huge datasets are required [16]. Since only a limited number of images are available in the medical image classification tasks, training DNN algorithms are a challenge. In order to overcome this shortcoming, applying transfer learning (TL) methods have been proposed. Transfer Learning can be defined as learning a new task through the transfer of knowledge from an already learned related task. Recent studies have shown that TL approaches do not need large datasets. Additionally, the required training times are also reduced since models are already somewhat pre-trained [6].

Considering the popularity of the DR and DNNs, many review studies have been published [4, 6, 8, 13–20]. Details of these studies could be found in Section 2. The only existing work on the intersection of DR, DNN, and TL is the study of Kandel and Castelli [16]. This fast-moving field requires more review studies for DR, DNNs, and TL approaches; therefore, the current review study is prepared. Considering previously published review articles in the same field, the present study differs from the aforementioned studies due to the following aspects:

- Considering three main keywords, DR, DNN, and TL, this study reviews 43 articles which are published in the period between 2016-2021. The number of studies selected year-wise is as follows: 2, 3, 11, 14, 8, and 5 articles from 2016, 2017, 2018, 2019, 2020 and 2021, respectively.
- These details of reviewed studies are summarized in 4 figures and 10 tables.
- These 43 articles are categorized into 29 CNN based architectures in Table 7.
- Used datasets (13 of them) have been analyzed considering their size and number of classes, in reviewed studies, Tables 3 and 2.
- Most of the reviewed articles compared their proposed performances using similar performance metrics. These metrics have been shown in Table 8.
- Additionally, the achieved accuracy and AUC values due to these 13 datasets have been presented in Table 9.
- Some of the reviewed studies have reported their used DNN optimization algorithms. These optimization algorithms have been given and discussed in Table 10.
- Pre-processing and data augmentation techniques applied in the reviewed articles have been evaluated and summarized in Tables 5 and 6.

The findings of this systematic review would enable the researchers to find more easily which databases, algorithms and performance metrics used in DR studies. This study is organized as the following; section 2 provides related works, then section 3 presents an overview of the methodology, then section 4 gives background materials, followed by section 5 which demonstrates DR datasets. Therewithal, section 6 addresses pre-trained CNN architectures and section 7 presents the performance metrics used in DR studies. Then, section 8 offers the parameter optimization for pre-trained CNN, and finally, section 9 provides a conclusion.

2. RELATED WORKS

Due to the popularity of the DR and DNNs many review studies have been published in recent years.

In 2019, Asiri et al. [13] reviewed studies that applied various DNN methods employed for DR diagnosis. In their study, detailed information about DR is given, and 18 DR datasets that can be employed for 4 different DR detection tasks are introduced. Nielsen et al. [15] reviewed 11 studies to evaluate the diagnostic accuracy of DNN to classify DR images. They summarized the characteristics and results of the reviewed studies, datasets used, and participants included in the studies using 2 tables. Ishtiaq et al. [8] published a detailed review on the application of AI techniques for the detection of DR, reviewing 74 papers from eight academic databases. Reviewed studies are discussed in various views such as datasets, image pre-processing methods, ML approaches, DNN based approaches, and evaluation of performance.

In 2020, Alyoubi et al. [6] published a review that analyzed the recent state-of-the-art DNN techniques for DR image detection and classification. They reviewed 33 studies due to used image pre-processing techniques, used screening systems (binary, multi-level, lesion-based, vessel-based), and 13 publicly available datasets. Stolte et al. [18] performed a comprehensive survey on medical image analysis in DR that provides a description of the currently used technology for DR detection in extensive detail. In their study, a detailed introduction about DR, current technologies, and available resources is given beside a discussion of frameworks used for DR detection and classification.

Also, in 2020, Badar et al. [19] presented a review study considering the application of DNN for the detection of retinal image impairments like DR. The published studies were analyzed due to publicly available datasets contains fundus and optical coherence tomography retinal images and DNN architectures. Lastly, they showed that DNN capable of replacing traditional classification methods. Another review study that combines DNN and DR published by Chu et al. [17] researched 15 high-quality papers. These papers are analyzed according to the selected dataset and performance criteria. Moreover, different from other review studies, this article focuses on the potential algorithm limitations of each study. Sarki et al. [14] published a review article for analyzing detection of Diabetic Eye Disease (DED) using DNN. DED includes DR and three various eye diseases, and 65 articles are reviewed from 10 academic databases in their study. Datasets, image processing methods, and detailed classification approaches (Tranfer learning, DNN, and combined DNN and ML classification) used for each eye disease are presented, respectively. Consequently, they show that DNN provides valuable results for DED detection.

In 2021, Tsiknakis et al. [20] published a review that provides an analysis of DL-based methods for DR detection and classification published after 2016. Their review drew attention to publicly available datasets, commonly used preprocessing methods, and frequently used segmentation and classification methods. However, their study only gave a tiny part for transfer learning approaches. Again in 2021, Lakshminarayanan et al. [4] presented a review that summarizes AI approaches for DR detection, severity grading, and segmentation using both fundus and OCT images. In this study, open literature sources published between 2016 and 2021 were presented. Though they gave a detailed summary of DR detection studies, they primarily focused on ML and DNN approaches rather than specific TL approaches. In 2021, Wu et al. [21] presented an overview that examines the overall diagnostic accuracy of ML-based studies in diagnosing DR of different severities based on color fundus photographs. In their review they concluded that ML-based DR detection and screening algorithms can be used for clinical applications, though the earlier developed algorithms had methodology flaws.

Among the published review studies, the only study that focuses on DR classification by transfer learning is published by Kandel and Castelli [16] in 2020. In their study, 18 papers were analyzed in accordance with architectures used, datasets used, optimizers used, and fine-tuning technique used. Ultimately, they showed that transfer learning can provide a substantial contribution to DR detection and classification.

Consequently, it is readily apparent that in recent years review papers on DR detection, severity grading, and segmentation methods using AI or DNN methods have been published. Although some review papers have been published, very few focus on the TL approaches; even in the study that focuses on TL [16], a detailed account of papers is not included. Therefore, this study presents a detailed systematic review of 43 publications based on DNN and Transfer Learning approaches for DR detection and classification.

3. OVERVIEW OF METHODOLOGY

This study is a systematic review that presents details about TL based DR detection and classification papers. Systematic review implicates the stages of a detailed and comprehensive screening of all studies on a specific issue, evaluating and synthesizing the findings based on certain research criteria [22].

This study aims to answer the following research questions (RQ):

- RQ1 What are the TL strategies used in DR studies?
- RQ2 Which CNN architectures are preferred in TL strategies?
- RQ3 What are the most used publicly available DR datasets?
- RQ4 What are the most used performance metrics in DR studies?
- RQ5 What are the optimization methods used?
- RQ6 What are the preprocessing and data augmentation techniques applied in DR studies?

To find the answers to these questions; initially, the review keywords were identified as "Diabetic Retinopathy, Deep Learning, and Transfer Learning". Then search keywords were used to find the most relevant studies from various academic publications between 2016 and 2021. As shown in Figure 1, the main filter was applied to select the related articles in two main groups: similar review studies and CNN-based DR detection studies. In the next step, as given in Figure 1, due to 3 inputs, 2 outputs, 10 tables and interpretations of these tables, have been generated. The details and discussion of the provided tables and figure have been presented under related titles.

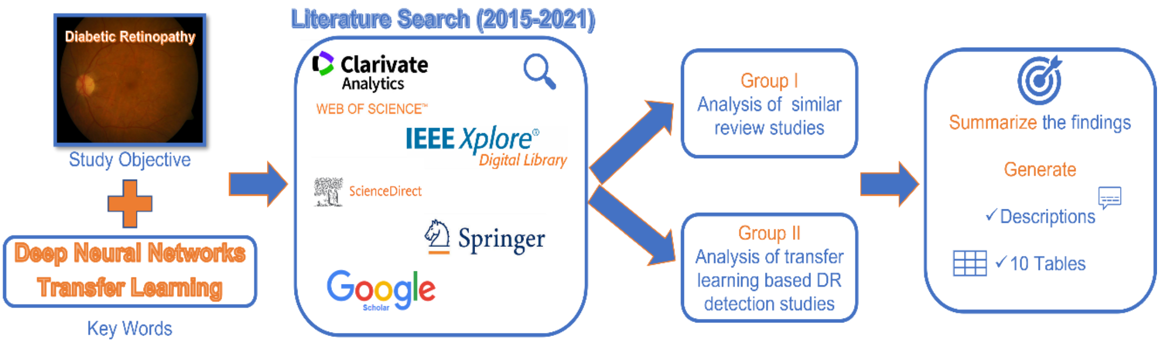

Figure 1. *Steps of research*

4. BACKGROUND MATERIALS

4.1. Diabetic Retinopathy

Diabetic retinopathy (DR) is a complication of diabetes that causes vision problems and blindness [2]. DR related blindness is preventable when it is early and correctly detected. According to World Health Organization [2], the global incidence of diabetes is projected to reach 552 million by 2024.

DR is classified into 5 stages according to the severity level (Figure 2). Stage 0 refers to non-apparent retinopathy, stage 1 is mild None-Proliferative DR (NPDR), stage 2 is Moderate NPDR, stage 3 is Severe NPDR, and stage 5 is Proliferative DR [23]. NPDR is the stage where the proliferative process has not yet begun, and patients are asymptomatic. At the NPDR stage, the visual acuity of the patients continues. Therefore, determining the NPDR stage of the patient is essential in predicting the risk of progression to proliferative retinopathy. Early and accurate detection of DR is necessary to prevent the progression of the disease and reduce the risk of vision loss. Early detection plays a vital role in optimal treatment [5, 24]. The traditional diagnosis of DR is made by manual examination of retinal scanning [23, 24]. This diagnostic method is a time-consuming task and the result obtained depends on the experience of the clinician [5, 24].

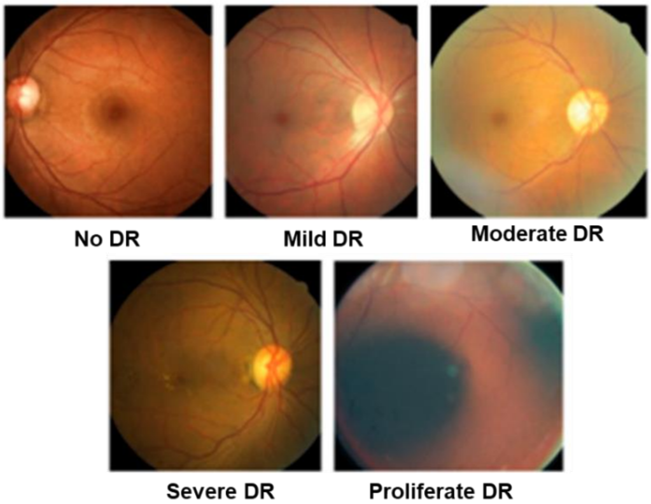

Figure 2. *Images from the five different severity levels*

4.2. Deep Neural Networks and Transfer Learning

Terms of deep learning, deep neural networks, and deep neural nets are used interchangeably in the literature [9]. DNNs are an advanced version of neural networks that belong to the family of ML and AI methods. Unlike traditional neural networks, DNNs have many hidden layers for low-level feature extraction [16]. Over the years, in the context of DNNs, various architectures have been introduced. Among these architectures, CNNs, which are feed-forward, multi-layered neural networks composed of feature extractor and classifier parts, have high achievements, especially on medical image classification [25]. Basic CNN architecture consists of four main parts: convolutional layers, pooling layers, fully connected (FC) layers, and a decision layer (output) [26].

The most crucial step of CNN applications is the training process. In order to avoid under-fitting and over fitting, and expand generalization, CNNs need to be trained appropriately. CNN can be trained with 2 different approaches: training from scratch and transfer learning. The transfer learning approach has three strategies. In the first three strategies, original FC layers are removed. Accordingly, the first strategy is to use the CNN pre-trained layers as feature extractor without any change and add a new classifier layer instead of the original FC layers. The second strategy is to fine-tune the entire network and add a suitable classifier layer for the task. The third strategy is to fine-tune only the selected amount of top layers where the bottoms layers are frozen and then to add a new, suitable classifier layer. An additional training strategy is sometimes mentioned as transfer learning approach. This last strategy is to select the state-of-the-art pre-trained structure and to train the architecture without any addition or removal (Figure 3). This strategy is also same as training from scratch. Except for the last strategy, the transfer learning approaches do not require a large amount of training data. Since most medical image datasets are not very large, applying transfer learning for image-based medical diagnosis is beneficial.

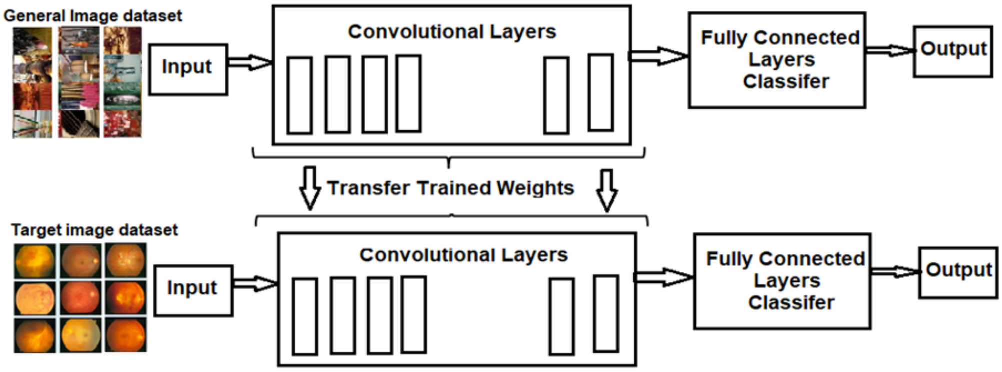

Figure 3. *Transfer learning approach*

Many of the popular pre-trained CNN models are trained using ImageNet [27]. ImageNet is a large-scale image classification dataset that contains 14.197.122 annotated images with 1000 object classes. Annually, an object classification and detection competition called The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) uses a subset of ImageNet is run [28]. Over the years, distinct approaches and architectures were presented in this competition. However, CNN architectures have been continuous winners of this competition since 2012. Even the dominance of CNN structures for image classification has been so well recognized that almost all of the structures participating in the competition in recent years are CNN structures. Although the basic components (e.g., convolutional and pooling layers) of the models are almost identical in newly presented architectures, some topological differences are presented in modern architectures [29]. Accordingly, since the introduction of AlexNet many highly thriving architectures, such as VGG, ResNet, DenseNet, GoogleNet, were presented for image classification [16]. Therefore, many of these well-known pre-trained CNN architectures have also been used for DR detection.

In the reviewed studies, various transfer learning approaches were preferred as can be seen in Table 1. In some of the studies, the preferred approach was given in detail, however in most of the studies the preferred strategy is just stated as Transfer Learning and details were not given. Note that if the study mentioned the used strategy as transfer learning or training a pre-trained network it is evaluated as fine-tuning the entire network in Table 1.

Table 1. *Transfer learning strategies used in the review studies*

TL strategy	Study
Using pre-trained network without fine-tuning	[23, 29]
Fine-tuning entire pre-trained network	[23, 24, 30–58]
Fine-tuning a part of the pre-trained network	[54, 55, 59]
Training a state-of-art architecture from scratch	[31, 54, 57, 60]
Modifying a pre-trained network	[33, 34, 36, 41, 48, 58, 61–64]
Not stated	[65–67]

As can be seen from Table 1, fine-tuning the whole network is the most preferred approach, where using the pre-trained network without fine-tuning is the least preferred one. Besides, training the state-of-the-art architecture from scratch is the least preferred approach due to the long training time. Additionally, in some of the studies, higher performances were achieved by modifying the state-of-the-art architecture trained by ImageNet.

5. DIABETIC RETINOPATHY DATASETS

There are various publicly available datasets, including images for classification or detection of DR. Table 2 shows the used datasets among the reviewed 43 articles. Properties of datasets, such as their sizes, the number of classes they contain, and their download links, are given.

Table 2. *DR dataset properties and access links*

Name	Data Size	Class	Link
DR Detection Kaggle	88702	5	https://www.kaggle.com/c/diabetic-retinopathy-detection/data
Messidor	1200	4	https://www.adcis.net/en/third-party/messidor/
Messidor-2	1748	5	https://www.adcis.net/en/third-party/messidor2/
APTOS 2019	3662	5	https://www.kaggle.com/c/aptos2019-blindness-detection
DIARETDB1	89	2	http://www2.it.lut.fi/project/imageret/diaretdb1/
IDRiD	516	5	https://idrid.grand-challenge.org/Data_Download/
eOphtha	463	3	https://www.adcis.net/en/third-party/e-ophtha/
DIABRET	1331	5	https://www.kaggle.com/lrasmy/sample-diab-retn
DR1	1014	2	http://www.record.ic.unicamp.br/site/asdr

University of Auckland Diabetic Retinopathy (UoA-DR)	200	3	https://researchspace.auckland.ac.nz/handle/2292/46737
STARE	397	14	https://cecas.clemson.edu/ ahoover/stare/
ODIR 2019	5814	8	https://odir2019.grand-challenge.org/dataset/
DIARETDB0	130	2	https://www.it.lut.fi/project/imageret/diaretdb0

Table 3 demonstrates the used datasets, and the frequencies of these used datasets in the reviewed articles. As shown in Table 3, many publicly available DR datasets were utilized for DR classification. Among them, the Diabetic Retinopathy Detection Dataset on Kaggle was the most preferred because of its size and accessibility. Besides, it is noteworthy that many non-public datasets were also used in most studies [5, 35, 36, 38, 44, 47–49, 58, 60]. Deep neural networks work better on large datasets, and the size of the data set is a very important parameter in the network's performance. For this reason, most studies have used more than one dataset to improve classification [30, 32, 34, 36, 40, 43, 46, 49, 50, 52, 53, 55, 56, 63, 65, 67].

Table 3. *Data sets used in the reviewed papers*

Dataset	Count	Study
Diabetic Retinopathy Detection Dataset on Kaggle	24	[23, 24, 30, 32, 34, 37, 40–43, 47, 49, 50, 52, 53, 56, 57, 59, 61–63, 65, 67, 68]
Messidor Dataset	11	[30, 40, 48–50, 52, 55, 56, 63, 66, 69]
Non-public Dataset	10	[5, 35, 36, 38, 44, 47–49, 58, 60]
Messidor-2 Dataset	5	[5, 36, 43, 49, 65]
APTOS 2019 Kaggle Dataset	4	[33, 39, 45, 64]
DIARETDB1 Dataset	4	[46, 50, 56, 67]
eOptha Dataset	3	[46, 50, 53]
IDRiD Dataset	2	[32, 49]
Used only once	1	DIABRET Dataset [34], DR1 Dataset [55], UoA-DR Dataset [49], STARE Dataset [51], ODIR 2019 [54], DIARETDB0 [50], Not given [31]

5.1. Datasets: The Number of Classes Used

Table 4 presents the number of classes for DR classification used in the reviewed papers. As mentioned before, the number of normal and abnormal images included in each dataset is different. In particular, the number of abnormal images graded according to DR severity level is inconsistent. The imbalance of the datasets and the quality of the images make it difficult to evaluate for more than two classes. For this reason, binary classification was used more than multi-class classification in the many reviewed studies [5, 30, 32, 37, 38, 40, 42–51, 53, 55, 56, 59, 64–69].

Table 4. *Number of classes for DR classification*

Number of class	Count	Study
2	26	[5, 30, 32, 37, 38, 40, 42–51, 53, 55, 56, 59, 64–69]
3	3	[30, 36, 51]
4	4	[30, 44, 58, 60]
5	19	[5, 23, 24, 31, 33, 34, 39–41, 45, 50, 52, 53, 57, 61–64, 68]
6	1	[35]
8	1	[54]
10	1	[51]

5.2. Preprocessing and Data Augmentation

Deep learning algorithms are used in order to eliminate the need for handcrafted features and, therefore pre-processing steps. However, in some DNN studies, some pre-processing steps for enhancing and improving the image quality are applicable. Additionally, several pre-processing techniques for identifying the region of interest (ROI) in images are used commonly. Applied pre-processing methods in the reviewed studies are presented in Table 5. Resizing is the most utilized pre-processing step. The basis for this is to make the images appropriate for the input size of CNN. Resizing is followed by normalization and cropping, as expected. Normalizing is performed to promote the DNN learning process, make the training faster, and avoid overfitting. Further, cropping improves success by reducing the contribution of the background to the training process.

Table 5. *Applied pre-processsing methods in the review studies*

Pre-Processing Method	Count	Study
Resize/Reshape/Rescale	32	[5, 23, 32–37, 40–43, 45–47, 49–52, 54–57, 59, 61, 62, 64–69]
Crop	17	[24, 30, 37, 39, 40, 42, 48–50, 53–56, 61, 62, 64, 69]
Normalization	14	[24, 30, 33, 41, 48, 50, 53, 56–58, 62, 65, 67, 68]
Contrast adjustment/enhancement/ improvement	8	[5, 30, 41, 44, 48, 52, 56, 67]
Nonstandard normalization	5	[32, 42, 47, 55, 68]
Removal of unwanted background, cropped out all backgrounds	5	[34, 35, 44, 57, 68]
Not given	3	[31, 38, 60]
Clipped	3	[40, 47, 59]
Scaling	2	[37, 61]
Downsampling	2	[39, 49]
Mask	2	[49, 56]
Histogram Equilization	2	[44, 54]
Nonlocal means denoising	2	[5, 24]
Used only once	1	Size standardization [46] Color enhancement [57], Upsampling Pixel Mapping [59], Sharpness unification [59], Min-pooling [33], Downsize [5], Resampling [35], Unsharp masking [41], Brightness adjustment [48], Noise reduction [48], Oversampling [51], Illumination [67], Illumination equalization [56], Gaussian smoothing [37], Gamma correction [36], Convert to binary [50]

Data augmentation is a commonly performed technique in deep learning applications to reduce overfitting due to the lack of training data set and increase the algorithm's performance by increasing the amount of data. Traditional data augmentation techniques are translation, stretching, flipping, zooming, contrast adjustment, and rotation. As shown in Table 6, in most of the studies, traditional data augmentation techniques are used, however, in a minority of the studies, different approaches are used. At the same time, as can be seen from Table 6 in almost half of the studies data augmentation process is not mentioned nor the used technique is explained. Nevertheless, the effect of data augmentation in DR classification is still a question mark and needs further investigation.

Table 6. *Applied data augmentation methods in the reviewed studies*

Data Augmentation Method	Count	Study
Not given	17	[5, 23, 34, 37–40, 42, 43, 45, 46, 57, 60, 65–67, 69]
Rotation	23	[24, 30–32, 35, 36, 41, 44, 47, 48, 50–52, 54–56, 58, 59, 61–64,

		68]
Flipping	18	[24, 32, 35, 36, 41, 44, 47, 49, 52, 55, 56, 58, 59, 61–64, 68]
Translation	7	[35, 50, 51, 53, 54, 56, 68]
Zoom-in and zoom-out	5	[30, 32, 36, 53, 58]
Scaling	5	[50, 56, 61, 62, 68]
Brightness	4	[41, 51, 53, 64]
Contrast	4	[41, 53, 61, 64]
Stretching	3	[24, 31, 35]
Shifting	3	[47, 52, 59]
Shearing	3	[32, 62, 68]
Cropped	2	[61, 68]
Mirror	2	[48, 49]
Rolling of the image	2	[30, 53]
Used only once	1	Additive Gaussian noise [51], Distort [58], Padding [30], Color augmentation [35], Saturation [41], Shuffle [41], Inverting [68], Not applied [33]

6. PRE-TRAINED CNN ARCHITECTURES

A pre-trained CNN model is a model which uses parameters of a model that is trained on another problem to solve a similar task rather than training a new model from scratch. Many of the popular pre-trained CNN models are trained using ImageNet [27]. Generally, though some topological differences are presented in modern architectures, the basic components (e.g., convolutional and pooling layers) of the models are almost identical [29]. Accordingly, since the introduction of AlexNet many highly thriving architectures, such as VGG, ResNet, DenseNet, GoogleNet, were presented for image classification [16]. Therefore, many of these well-known pre-trained CNN architectures have also been used for DR detection.

In most of the studies reviewed, different architectures were compared to achieve the best performance [23, 24, 30, 32–36, 38, 40, 44–46, 48, 49, 51–55, 57–59, 63, 64, 69, 70]. Whereas only one architecture is analyzed in some of the studies [5, 31, 37, 39, 42, 43, 47, 50, 56, 60–62, 65–68].

Table 7 illustrates the architectures used in the reviewed studies. As can be seen from Table 7, Inception-v3 is the most commonly used architecture (20 studies) among the many state-of-the-art architectures in the literature. The selection of Inception-V3 architecture may be due to the obtained optimized results using this architecture [41]. Subsequent to Inception-v3, VGG-16 and AlexNet are the architectures that are mostly preferred in the reviewed articles.

In the studies that investigated the performances of different architectures, AlexNet produced the lowest performances most of the time [23, 24, 30, 49, 51–53, 55, 57, 69]. According to [16], this situation occurs since AlexNet is a shallow model, and the sufficient number of convolutional layers makes the classification a challenge. The architecture that yields the best performance among the investigated studies is Inception-V3 [23, 41, 53, 59, 69]. The reason of this performance may lie behind the inception modules that are used by Inception-V3. Inception modules are consist of different sized filters on the same convolution layer; hence they extract features in different aspect ratios. Since features of DR differ in size, the inception module can be beneficial for DR classification [42].

Table 7. *CNN architectures used in the review studies*

CNN architecture	Count	Study
InceptionV3	20	[5, 23, 33, 34, 36, 41–44, 46–48, 53, 54, 58, 59, 61, 65, 68, 69]
VGG16	15	[23, 24, 34–36, 38–40, 45, 52–55, 57, 69]
AlexNet	12	[23, 24, 30, 49, 51–53, 55–57, 66, 69]
GoogLeNet	10	[24, 30, 35, 40, 49, 52, 53, 55, 60, 69]
VGG19	9	[24, 36, 46, 51, 54, 55, 58, 63, 69]

ResNet50	9	[33, 36, 38, 44, 46, 49, 54, 57, 69]
InceptionResNetV2	9	[32, 34, 36, 44, 45, 48, 54, 64, 69]
Xception	9	[32–34, 36, 44, 45, 54, 59, 64]
ResNet18	3	[34, 35, 58]
NASNet	3	[36, 45, 50]
Densenet	3	[31, 44, 54]
DenseNet121	2	[35, 36]
DenseNet201	2	[34, 69]
VGG-s	2	[24, 55]
MobileNet	2	[33, 54]
Resnet	3	[24, 53, 60]
Resnet152	2	[52, 63]
Used only once	1	ResNet101 [58], SE-BN-Inception [5], VGG-f [55], VGG-m [55], EfficientNetB3 [54], Inception@4 [58], Unknown [37], DPN107 [63], ResNetGB [63], VGG-NiN [62], ZFNet [52], MobileNetV2 [64], RA-EfficientNet.[64]

7. PERFORMANCE METRICS

Common metrics for measuring the performance of classification and detection include accuracy, sensitivity (recall), specificity, precision, F-score. Similar to other classification problems, common performance metrics used for DR detection. The formulation of these metrics has been given in Figure 4.

Performance metrics		
$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	$Sensitivity\ (Recall) = \frac{TP}{TP + FN}$	$Specificity = \frac{TN}{TN + FP}$
$Precision = \frac{TP}{TP + FP}$	$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	

Figure 4. *Common performance metrics*

Additionally, a receiver operating characteristic (ROC), which is a graphical technique used for visualizing and comparing the classifiers performance that presents sensitivity versus (1-specificity) in 2-dimensional graphs, and area under the ROC curve (AUC) are two of the most commonly used techniques for performance measurements [71].

Table 8 provides the performance metrics used in the reviewed studies. As stated above, accuracy, precision, recall, AUC metrics are typically used to measure the performance of medical classification problems. Accordingly, in the reviewed studies, accuracy is the most commonly used metric, followed by sensitivity/recall, AUC, and specificity. Moreover, in some of the mentioned studies, [31, 35, 41, 42, 44, 45, 49, 51–54, 61] metrics that are not so common are used, which inhibits the performance comparison between similar studies. Additionally, time is a significant metric for deep learning structures; however, only one study [41] presents this metric.

Table 8. *Performance metrics used in the review studies*

Performance metric	Count	Study
Accuracy	34	[5, 23, 24, 30, 33–37, 39–42, 44–52, 54, 55, 57–60, 62–64, 66, 69]
Sensivity (Recall)	29	[5, 24, 30, 32, 33, 36, 38, 40, 41, 43–50, 52–56, 58, 61–63, 65, 67, 68]
AUC	20	[5, 24, 33, 37, 38, 40, 43, 44, 48, 49, 51, 53–57, 62, 65, 67, 68]
Specifity	19	[5, 24, 30, 32, 33, 38, 40, 43, 44, 48–50, 52, 55, 62, 63, 65, 67, 68]
Precision	16	[35, 36, 41, 44, 45, 47, 50, 52–54, 56, 58, 61, 62, 64, 72]

F1-score	13	[34, 36, 41, 44-47, 54, 56, 57, 61, 62, 64]
Kappa Statistic (weighted and adjusted)	12	[32, 35, 39, 44, 45, 51, 54, 60, 61, 63, 64, 68]
Loss	4	[41, 42, 44, 45]
Support	3	[41, 44, 61]
Used only once	3	Equal Error Rate [49], Intersection Over Union (IoU) [35], Time [41], Error rate [31], Relative classifier information (RCI) [51], Area under the precision recall curve [53], Final score [54], Youden's index [44], Gmean [52]

As stated above, accuracy and AUC are two of the most preferred performance metrics in the reviewed articles. Therefore, these two metrics are selected in order to compare the performances of the studies according to data sets. Table 9 provides the best performance results achieved in the studies according to used datasets. As seen in the table, Using Diabetic Retinopathy Detection Dataset on Kaggle, the most preferred data set, [63] achieved the highest accuracy whereas the highest AUC is achieved by [65]. In the Messidor dataset, which is the second most used set, [49] obtained the highest accuracy of 99.1% yet, the highest AUC of 0,9834 is obtained by [55]. Besides, [64] achieved the highest accuracy using the APTOS data set.

Table 9. *Accuracy and AUC*

Position & Study	Accuracy (%)	AUC	Position & Study	Accuracy (%)	AUC	Position & Study	Accuracy (%)	AUC
1 [39]	77	-	21 [65]	-	0.991	41 [43]	-	0.853
2 [33]	83.09	-	22 [63]	99.73	0.98	42 [65]	-	0.99
3 [45]	97.41	-	23 [62]	85	0.95	43 [36]	95	-
4 [64]	98.36	-	24 [52]	88.68	-	44 [5]	93.49	0.9905
5 [56]	-	0.925	25 [34]	98.69	-	45 [35]	82.84	-
6 [24]	95.68	0.9786	26 [72]	98.91	-	46 [38]	-	0.987
7 [42]	90.9	-	27 [56]	-	-	47 [60]	96	-
8 [61]	80	-	28 [50]	84.27	-	48 [48]	97.07	0.997
9 [53]	98	0.99	29 [55]	94.12	0.9786	49 [58]	88.72	-
10 [40]	83.63	-	30 [32]	-	-	50 [47]	70.30	-
11 [23]	63.23	-	31 [66]	88.3	-	51 [49]	98.6	0.998
12 [43]	-	0.951	32 [40]	89.7	0.891	52 [44]	97.67	0.986
13 [59]	87.12	-	33 [48]	87.18	0.926	53 [51]	80.8	0.903
14 [57]	79.04	0.82	34 [69]	95.83	-	54 [46]	98.43	-
15 [34]	98.63	-	35 [56]	-	0.960	55 [53]	-	0.95
16 [68]	-	0.951	36 [55]	92.01	0.9834	56 [50]	85.75	-
17 [41]	94.3	-	37 [49]	99.1	-	57 [54]	89	0.73
18 [37]	98.46	0.985	38 [63]	98.88	0.98	58 [30]	74.5	-
19 [49]	92.1	-	39 [50]	86.73	-	59 [50]	85.38	-
20 [67]	-	0.963	40 [52]	97.65	-			

- Studies 1-4 achieved the results using APTOS 2019 Kaggle Dataset.
- Studies 5-24 achieved the results using DR Detection Dataset on Kaggle.
- Study 25 achieved the result using DIABRET Dataset.
- Studies 26-28 used achieved the results using DIARETDB1 Dataset.
- Studies 41-43 used achieved the results using Messidor-2 Dataset.
- Studies 44-52 used achieved the results using Non-public Dataset.
- Study 53 achieved the result using STARE Dataset.
- Studies 54-56 used achieved the results using eOPTHA Dataset.

• Study 29 achieved the result using DR1 Dataset. • Study 30 achieved the result using IDRiD Dataset. • Studies 31-40 used achieved the results using Messidor Dataset.	• Study 57 achieved the result using ODIR 2019 Dataset. • Study 58 achieved the result using Merged Dataset. • Study 59 achieved the result using DIARETDB0 Dataset.

8. PARAMETER OPTIMIZATION FOR PRE-TRAINED CNN

The learning process in multilayer artificial neural networks is an optimization problem. During the learning process, many classical and intelligent approaches can be applied. Additionally, some hyper parameter tuning can be included in the learning error minimization process. For example, using adaptive learning rate or including moment parameter in weight update formulation helps convergence and reduces the training time. All of these approaches can be evaluated as hyper parameter tuning. According to this subject, some of the evaluated papers, during the learning process, used adaptive learning rate and momentum parameter.

Table 10 illustrates the optimizers used in the reviewed studies. The optimizer is a significant technique used to change the weights with the learning rate in order to minimize the loss function during the neural network training. As shown in Table 10, the stochastic gradient descent (SGD) is the most commonly used optimizer, followed by the Adaptive Moment Estimation (Adam). SGD performs a fast training process as it updates a training example on each iteration. Some studies like [59] and [47] propose the Adam algorithm to avoid local minimums and to converge faster than SGD. Furthermore, the deficiency in some of the mentioned studies [35, 37–40, 46, 53, 57, 66, 69] is that they do not specify the optimizers used during the training. However, few studies used different optimizer algorithms such as Root Mean Square Error Probability (RMSProp), Numerical Algorithms Group (NAG), Adaptive Gradient Algorithm (AdaGrad), Stochastic Gradient Descent with Momentum (SGDM), Adamax and Gradient Descent.

Table 10. *Optimization methods used in the review studies*

Optimizer	Count	Study
Stochastic Gradient Descent (SGD)	16	[5, 24, 30, 31, 34, 42, 49, 51, 54–56, 59, 61, 62, 65, 67]
Adaptive Moment Estimation (Adam)	15	[32–34, 36, 45, 47, 48, 50, 53, 54, 58, 59, 63, 64, 68]
Not Given	10	[35, 37, 38, 40, 46, 53, 57, 60, 66, 69]
Root Mean Square Error Propability (RMSProp)	3	[30, 43, 44]
Stochastic Gradient Descent with Momentum (SGDM)	3	[23, 50, 52]
Used only once	1	Numerical Algorithms Group (NAG) [30], Adaptive Gradient Algorithm (AdaGrad) [30], Adamax [41], Gradient descent [39]

9. CONCLUSION

DR is one of the complications of diabetes that occurs when the blood vessels in the retina are damaged due to diabetes. DR, if left untreated, could cause blindness. Early diagnosis of DR can effectively prevent visual impairment. Automatic classification of DR images can effectively assist physicians in diagnosing process and improves diagnostic efficiency accordingly. In recent years, DNN algorithms, mostly CNN architectures, have been frequently used in studies related to the analysis and classification of retinal images. However, since CNN training requires large datasets for high performance, transfer learning approaches are preferred due to insufficient sized medical data. Accordingly, in this review, 43 papers that performed transfer learning approaches for DR detection have been analyzed in terms of the used dataset, used architecture, used performance metrics, used optimizers, and applied preprocessing and augmentation methods. Findings and statistics have been summarized in the aforementioned sections.

The findings of this review are as follows:

- The most studied dataset in the reviewed studies is Diabetic Retinopathy Detection Dataset on Kaggle.
- The most preferred pre-trained CNN architecture is Inception-v3.
- Through all performance metrics, most performance measurements are made using accuracy and sensitivity (recall).
- Among the parameter optimization techniques, SGD is the most preferred one, followed by Adam.
- Additionally, in many of the reviewed studies, the used parameter optimization technique is not mentioned.
- Since pre-trained CNN architectures have specific input shapes, resize/reshape/rescale pre-processing method is the most implemented technique.
- Though data augmentation techniques are not applied (or given) in most studies, rotation is the most used technique among the given ones.

According to these findings in this study, we present 2 groups of suggestions as future work to help improve this research field.

Researchers that are eager to detect or classify DR;

- can use more than one dataset for both training and testing the DNNs. Since data size is an important parameter for DNN training and achieving a robust model, combining the datasets and/or collecting more images and creating a mixed- voluminous dataset would improve the performance. Thence, a robust and successful model would have been built and it can be used in decision support systems.
- can apply hyperparameter tuning (optimization) to decide the parameters of CNNs. In most of the studies, the importance of the correct hyperparameter selection is mentioned. Therefore, rather than selecting the hyperparameters according to trial-and-error method or based on the former studies, hyperparameters should be selected for the aforementioned task and dataset. As a consequence, hyperparameter tuning using the appropriate optimization algorithms would both increase the performance as well as decrease the training time.
- can prefer Python or Matlab for building the models. But according to the literature, Python is preferable to Matlab. The reason for selecting Python is mostly its extensive libraries and ease of programming. Moreover, since python is an open-source environment it would be more advantageous to use Python rather than Matlab though in recent years Matlab has improved its libraries on deep learning.
- would prefer DR classification (5 class) rather than DR detection (healthy-DR). since in clinical applications the severity of the disease is important for both the treatment and medicine selection, it would be more beneficial to create a CAD that both detects and classifies DR rather than only perform DR detection.

Researchers that are enthusiastic to review studies about DR detection and classification can take into account the following suggestions.

- In our study, we have presented the findings (applied TL strategies, DR datasets and their properties, the number of classes of DR classes, applied pre-processing and data augmentation techniques, used CNN architectures, performance metrics, and used optimization methods). However, we have not mentioned the comparisons i.e. which augmentation methods improve the performance of CNN for the same dataset or which optimizer would be useful for building successful models. Therefore, a comparison review would set light to the researchers for selecting the optimal methods.

- Additionally, in our study, the used environments, libraries and the training time is not mentioned since they are not given in many of the reviewed studies. However, an analysis of these approaches would be guiding for researchers that are new to this research area.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Karaca, B. K., Oltu B., Ozgur A., and Erdem H., “Comparison of Transfer Learning Strategies for Diabetic Retinopathy Detection”, 2021 Innovations in Intelligent Systems and Applications Conference (ASYU), 1–5, (2021).

[2] https://www.who.int/features/factfiles/diabetes/en/ Access date: 21.02.2021

[3] Guariguata, L., Whiting, D. R., Hambleton, I., Beagley, J., Linnenkamp, U., and Shaw, J. E., “Global estimates of diabetes prevalence for 2013 and projections for 2035”, Diabetes Research and Clinical Practice, 103(2): 137–149, (2014).

[4] Lakshminarayanan, V., Kheradfallah, H., Sarkar, A., and Jothi Balaji, J., “Automated detection and diagnosis of diabetic retinopathy: A comprehensive survey”, Journal of Imaging, 7(9): 165, (2021).

[5] Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X., and Wu, Z., “Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm”, Translational Vision Science and Technology, 8(6), (2019).

[6] Alyoubi, W. L., Shalash, W. M., and Abulkhair, M. F., “Diabetic retinopathy detection through deep learning techniques: A review”, Informatics in Medicine Unlocked, 20: 100377, (2020).

[7] Safi, H., Safi, S., Hafezi-Moghadam, A., and Ahmadieh, H., “Early detection of diabetic retinopathy”, Survey of Ophthalmology, 63(5): 601–608, (2018).

[8] Ishtiaq, U., Abdul Kareem, S., Abdullah, E. R. M. F., Mujtaba, G., Jahangir, R., and Ghafoor, H. Y., “Diabetic retinopathy detection through artificial intelligent techniques: a review and open issues”, Multimedia Tools and Applications, 79(21–22): 15209–15252, (2020).

[9] Lecun, Y., Bengio, Y., and Hinton, G., “Deep learning”, Nature, 521(7553): 436–444, (2015).

[10] Tombaloğlu, B., and Erdem, H., “Turkish Speech Recognition Techniques and Applications of Recurrent Units (LSTM and GRU)”, Gazi University Journal of Science, 34(4): 1035-1049, (2021).

[11] Shamshirband S., Fathi M., Dehzangi A., Chronopoulos A. T., and Alinejad-Rokny H., “A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues”, Journal of Biomedical Informatics, 113: 103627, (2021).

[12] Lin, D., Vasilakos, A. V., Tang, Y., and Yao, Y., “Neural networks for computer-aided diagnosis in medicine: A review”, Neurocomputing, 216: 700–708, (2016).

[13] Asiri, N., Hussain, M., Al Adel, F., and Alzaidi, N., “Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey”, Artificial Intelligence in Medicine, 99: 101701, (2019).

[14] Sarki, R., Ahmed, K., Wang, H., and Zhang, Y., “Automatic detection of diabetic eye disease

through deep learning using fundus images: A survey", IEEE Access, 8: 151133–151149, (2020).

[15] Nielsen, K. B., Lautrup, M. L., Andersen, J. K. H., Savarimuthu, T. R., and Grauslund, J., "Deep learning–based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance", Ophthalmology Retina, 3(4): 294–304, (2019).

[16] Kandel, I., and Castelli, M., "Transfer learning with convolutional neural networks for diabetic retinopathy image classification, A review", Applied Sciences, 10(6): 2021, (2020).

[17] Chu, A., Squirrell, D., Phillips, A. M., and Vaghefi, E., "Essentials of a robust deep learning system for diabetic retinopathy screening: A systematic literature review", Journal of Ophthalmology, (2020).

[18] Stolte, S., and Fang, R., "A survey on medical image analysis in diabetic retinopathy", Medical Image Analysis, 64: 101742, (2020).

[19] Badar, M., Haris, M., and Fatima, A., "Application of deep learning for retinal image analysis: A review", Computer Science Review, 35: 100203, (2020).

[20] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D. I., and Marias, K., "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review", Computers in Biology and Medicine, 135: 104599, (2021).

[21] Wu J.-H., Liu T. Y. A., Hsu W.-T., Ho J. H.-C., and Lee C.-C., "Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis", Journal of Medical Internet Research, 23(7): e23863, (2021).

[22] Uman, L. S., "Systematic reviews and meta-analyses", Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20(1): 57, (2007).

[23] Wang, X., Lu, Y., Wang, Y., AND Chen, W. B., "Diabetic retinopathy stage classification using convolutional neural networks", 2018 IEEE International Conference on Information Reuse and Integration (IRI), 465-471, (2018).

[24] Wan, S., Liang, Y., Zhang, Y., "Deep convolutional neural networks for diabetic retinopathy detection by image classification", Computers and Electrical Engineering, 72: 274–282, (2018).

[25] Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., and de Albuquerque, V. H. C., "A novel transfer learning based approach for pneumonia detection in chest X-ray images", Applied Sciences (Switzerland), 10(2), (2020).

[26] Ul Abideen, Z., Ghafoor, M., Munir, K., Saqib, M., Ullah, A., Zia, T., Tariq, S. A., Ahmed, G., and Zahra, A., "Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks", IEEE Access, 8:22812–22825, (2020).

[27] Kieu, S. T. H., Bade, A., Hijazi, M. H. A., and Kolivand, H., "A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions", Journal of Imaging, 6(12): 131, (2020).

[28] Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg Alexander C. and Li Fei-Fei, "Imagenet large scale visual recognition challenge", International Journal of Computer Vision, 115: 211–252, (2015).

[29] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van

Essen, B. C., Awwal, A. A. S., and Asari, V. K., “A state-of-the-art survey on deep learning theory and architectures”, Electronics (Switzerland), 8(3): 1–67, (2019).

[30] Lam, C., Yi, D., Guo, M., and Lindsey, T., “Automated detection of diabetic retinopathy using deep learning”, AMIA summits on translational science proceedings, 2018: 147, (2018).

[31] Xu, X., Lin, J., Tao, Y., & Wang, X., “An improved DenseNet method based on transfer learning for fundus medical images”, 2018 7th International Conference on Digital Home (ICDH), 137–140, (2018).

[32] Hathwar, S. B., and Srinivasa, G., “Automated grading of diabetic retinopathy in retinal fundus images using deep learning”, 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 73–77, (2019).

[33] Kassani, S. H., “Diabetic retinopathy classification using a modified Xception architecture”, IEEE international symposium on signal processing and information technology (ISSPIT), 1-6, (2019).

[34] Wijesinghe, I., Gamage, C., and Chitraranjan, C., “Transfer learning with ensemble feature extraction and low-rank matrix factorization for severity stage classification of diabetic retinopathy”, IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 931–938, (2019).

[35] Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., and Kang, H., “Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening”, Information Sciences, 501: 511–522, (2019).

[36] Ahmad, M., Kasukurthi, N., and Pande, H., “Deep learning for weak supervision of diabetic retinopathy abnormalities”, IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 573–577, (2019).

[37] Math, L. and Fatima, R., “Identification of diabetic retinopathy from fundus images using CNNs”, International Journal of Innovative Technology and Exploring Engineering, 9(1): 3439–3443, (2019).

[38] Yip, M. Y. T., Lim, Z. W., Lim, G., Quang, N. D., Hamzah, H., Ho, J., Bellemo, V., Xie, Y., Lee, X. Q., Lee, M. L., Hsu, W., Wong, T. Y., and Ting, D. S. W., “Enhanced detection of referable diabetic retinopathy via DCNNs and transfer learning”, Asian Conference on Computer Vision, Springer, Cham., 282-288, (2019).

[39] Dekhil, O., Naglah A., Shaban, M., Ghazal, M., Taher, F., and Elbaz, A., “Deep learning based method for computer aided diagnosis of diabetic retinopathy”, IEEE International Conference on Imaging Systems and Techniques (IST), 1–4, (2019).

[40] Vo, H. H., and Verma, A., “New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space”, IEEE International Symposium on Multimedia (ISM), 209-215, (2016).

[41] Islam, N., Saeed, U., Naz, R., Tanveer, J., Kumar, K., and Shaikh, A. A., “DeepDR: An image guided diabetic retinopathy detection technique using attention-based deep learning scheme”, 2nd International Conference on new Trends in Computing Sciences (ICTCS) 2019, 1-6, (2019).

[42] Hagos, M. T. and Kant ,S., “Transfer learning based detection of diabetic retinopathy from small dataset”, arXiv preprint arXiv: 1905.07203, (2019).

[43] Voets, M., Møllersen, K., and Bongo, L. A., “Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs”, PLoS One, 14(6): 1–11, (2019).

[44] Zhang, W., Zhong, J., Yang, S., Gao, Z., Hu, J., Chen, Y., and Yi, Z., "Automated identification and grading system of diabetic retinopathy using deep neural networks", Knowledge-Based Systems, 175: 12–25, (2019).

[45] Bodapati, J. D., Veeranjaneyulu, N., Shareef, S. N., Hakak, S., Bilal, M., Maddikunta, P. K. R., and Jo, O., "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction", Electronics (Switzerland), 9(6): 1–16, (2020).

[46] Mateen, M., Wen, J., Nasrullah, N., Sun, S., and Hayat, S., "Exudate detection for diabetic retinopathy using pretrained convolutional neural networks", Complexity, 1–11, (2020).

[47] Roshan, S. M., Karsaz, A., Vejdani, A. H., and Roshan, Y. M., "Fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening: a clinical study", International Journal of Computational Science and Engineering, 21(4): 564–573, (2020).

[48] Yu, Y., Chen, X., Zhu, X., Zhang, P., Hou, Y., Zhang, R., and Wu, C., "Performance of deep transfer learning for detecting abnormal fundus images", Journal of Current Ophthalmology, 32(4): 368, (2020).

[49] Hacisoftaoglu, R. E., Karakaya, M., and Sallam, A. B., "Deep learning frameworks for diabetic retinopathy detection with smartphone-based retinal imaging systems", Pattern recognition letters, 135: 409–417, (2020).

[50] Chen, P. N., Lee, C. C., Liang, C. M., Pao, S. I., Huang, K. H., and Lin, K. F., "General deep learning model for detecting diabetic retinopathy", BMC Bioinformatics, 22(5): 1-15, (2021).

[51] Choi, J. Y., Yoo, T. K., Seo, J. G., Kwak, J., Um, T. T., and Rim, T. H., "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database", PLoS One, 12(11): 1–16, (2017).

[52] Somasundaram, K., Sivakumar, P., and Suresh, D., "Classification of diabetic retinopathy disease with transfer learning using deep convolutional neural networks", Advances in Electrical and Computer Engineering, 21(3): 49-56, (2021).

[53] Lam, C., Yu, C., Huang, L., and Rubin, D., "Retinal lesion detection with deep learning using image patches", Investigative ophthalmology & visual science, 59(1): 590-596, (2018).

[54] Wang, J., Yang, L., Huo, Z., He, W., and Luo, J., "Multi-Label classification of fundus images with EfficientNet", IEEE Access, 8: 212499–212508, (2020).

[55] Li, X., Pang, T., Xiong, B., Liu, W., Liang, P., and Wang, T., "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification", 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI), 1–11, (2018).

[56] Zhou, L., Zhao, Y., Yang, J., Yu, Q., and Xu, X., "Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images", IET Image Processing, 12(4): 563-571, (2018).

[57] Lian, C., Liang, Y., Kang, R., and Xiang, Y., "Deep convolutional neural networks for diabetic retinopathy classification", Proceedings of the 2nd International Conference on Advances in Image Processing, 68–72, (2018).

[58] Gao, Z., Li, J., Guo, J., Chen, Y., Yi, Z., and Zhong, J., "Diagnosis of Diabetic Retinopathy Using Deep Neural Networks", IEEE Access, 7: 3360–3370, (2019).

[59] Mohammadian, S., Karsaz, A., and Roshan, Y. M., “Comparative study of fine-tuning of pre-trained convolutional neural networks for diabetic retinopathy screening”, 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME), 1–6, (2018).

[60] Takahashi, H., Tampo, H., Arai, Y., Inoue, Y., and Kawashima, H., “Applying artificial intelligence to disease staging : Deep learning for improved staging of diabetic retinopathy”, PloS one, 12(6): 1–11, (2017).

[61] Chen, H., Zeng, X., Luo, Y., and Ye, W., “Detection of diabetic retinopathy using deep neural network”, IEEE 23rd International Conference on Digital Signal Processing (DSP), 1-5, (2019).

[62] Khan, Z., Khan, F. G., Khan, A., Rehman, Z. U., Shah, S., Qummar, S., Ali, F., and Pack, S., “Diabetic retinopathy detection using VGG-NIN a deep learning architecture”, IEEE Access, 9: 61408–61416, (2021).

[63] Saeed, F., Hussain, M., and Aboalsamh, H. A., “Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network”, IEEE Access, 9: 41344–41359, (2021).

[64] Yi, S. L., Yang, X. L., Wang, T. W., She, F. R., Xiong, X., and He, J. F., “Diabetic retinopathy diagnosis based on RA-Efficientnet”, Applied Sciences, 11(22): 11035, (2021).

[65] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., and Webster, D. R., “Development and validation of a deep learning algorithm detection of diabetic retinopathy in retinal fundus photographs”, Jama, 316(22): 2402-2410, (2016).

[66] Hazim, J. M., Hassan, H. A., Yassin, A. I. M., Tahir, N. M., Zabidi, A., Rizman, Z. I., Baharom, R., and Wahab, N. A., “Early detection of diabetic retinopathy by using deep learning neural network”, International Journal of Engineering & Technology, 7(411): 1997-2004, (2018)

[67] Math, L., and Fatima, R., “Adaptive machine learning classification for diabetic retinopathy”, Multimedia Tools and Applications, 80(4): 5173-5186, (2020).

[68] Zeng, X., Chen, H., Luo, Y., and Ye, W., “Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network”, IEEE Access, 7: 30744–30753, (2019).

[69] Qomariah, D. U. N., Tjandrasa, H., and Fatichah, C., “Classification of diabetic retinopathy and normal retinal images using CNN and SVM”, 2019 12th International Conference on Information & Communication Technology and System (ICTS), 152-157, (2019).

[70] Islam, K. T., Wijewickrema, S., and O'Leary, S., “Identifying diabetic retinopathy from OCT images using deep transfer learning with artificial neural networks”, IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), 281–286, (2019).

[71] Fawcett, T., “An introduction to ROC analysis”, Pattern Recognition Letters, 27(8): 861–874, (2006).

[72] Mateen, M., Wen, J., Hassan, M., Nasrullah, N., Sun, S., and Hayat, S., “Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics”, IEEE Access, 8: 48784–48811, (2020).