

Segmentation and Classification of Retinal Fundus Images using Convolutional Neural Networks: A systematic review of methods and latest trends

Ademola Ilesanmi (✉ demoranky00@yahoo.com)

University of Pennsylvania

Taiwo Ilesanmi

National Population Commission of Nigeria

Oluwagbenga P. Idowu

Shenzhen Institutes of Advanced Technology Chinese Academy of Sciences

Drew A. Torigian

University of Pennsylvania

Jayaram K. Udupa

University of Pennsylvania

Research Article

Keywords: Convolutional neural network, Retinal fundus images, Segmentation and classification, Eye-related disease

Posted Date: February 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1151947/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The retina fundus images are critical diagnostic tools for the early detection of ophthalmic problems. Prompt diagnosis and treatment of eye-related problems are important to prevent blindness or vision loss. Recently, CNN algorithms have become effective in tasks relating to recognition, delineation, and classification. Therefore we propose a review to summarize different CNN algorithms for segmenting and classifying retinal fundus images. This review systematically searches different repositories for methods that use CNN for the segmentation and classification of retinal fundus images. A thorough screening of abstracts and titles to ascertain their relevance was done. A total of fifty-two studies were included in our review, with content such as database usage, disadvantages, and advantages. A comparison of two accuracies was also carried out with a graph depicting database usage. Important insights, limitations, observations, and future directions were elucidated. Finally, findings suggest that CNN algorithms produce good accuracies despite their limitations.

1 Introduction

Globally, eye-related diseases have increased with several individuals suffering from acute cases. The WHO estimates that about 2.2 billion people suffer from visual impairment. Out of this number, about 1 billion cases would have been prevented if detected on time. In the United States, at least 40 million people suffer from acute to worse eye-related diseases. The majority of these eye issues involve the retina, Glaucoma, and others [1, 2]. In Africa, approximately 26.3 million people have a form of visual impairment. Out of which 20.4 million have low vision, 5.9 million are blind (resulting in 15.3% of the world's blind population). Uncorrected refractive error and cataracts account for the major cause of blindness and visual impairment. People over the age of 50 are prone to visual impairment and blindness, however, vision loss and blindness can affect people of all ages. According to the International Classification of Diseases 11(2018), visual impairment is classified into two groups: 1) Distance vision impairment (Mild to Blindness) and 2) Near vision impairment (acuity worse than N6 or M.08 at 40cm).

The state of visual impairment depends on factors such as the availability of preventives and treatment, access to vision rehabilitation, and access to basic amenities (such as building, transport, and information). The leading causes of vision impairment are: 1) Cataract, 2) Glaucoma, 3) Diabetic retinopathy 4) Corneal Opacity, 5) Trachoma, 6) Uncorrected refractive error, and 7) age-related macular degeneration. The causes of visual impairment among children vary across countries and social statuses. In countries with low income, cataract is the leading cause, whereas, premature retinopathy is the leading cause in middle-income countries. The uncorrected refractive error remains a leading cause of visual impairment in children of all countries. In the adult population, Glaucoma will affect almost 80 million people worldwide, and it is estimated as one of the major causes of blindness. Open-angle and closure glaucoma are the two main types of glaucoma. Clinician estimates that about 90% of affected people suffer from open-angle glaucoma. To diagnose glaucoma, procedures such as Neuroretinal rim loss, visual fields, and nerve fiber are performed [3].

The white cup-like display at the center of the eye is the optic cup. This cup is located at the center of the optic disk and it serves as one of the measures used in the diagnosis of glaucoma. Another common cause of visual impairment is diabetic retinopathy (DR). DR is caused by damage to tissues located in the blood vessel. This damage occurs at the back of the eye. The WHO estimates DR as a major cause of blindness (fourth cause of blindness globally) [4] (see Figs. 2 and 3). In the early stages, there are no symptoms or changes to eyesight, however, if not treated timely it can permanently damage eyesight. A study conducted by the center for disease control and prevention in the United States found that almost one-third of adults over 40 years have DR. More than one-third of African-Americans and Mexican-American have diabetes [5]. The retinal fundus image is a photograph that captures the direct optical process of the eye. Structures like blood vessels, macula, fovea, optic disk, hemorrhages, arterioles, venules, exudates, and micronaneurysms are morphological and pathological components of the eye (see details in [2]). A diagram of the retinal fundus images showing different signs of the DR (exudates, micronaneurysms, hemorrhages, and neovascularization) is available in Fig. 1.

To avoid visual loss it is recommended to do the following: 1) Maintain blood sugar levels, 2) Know family eye health history, 3) Eat right, 4) Maintain a healthy weight, 5) Wear protective eyewear, 6) Make exercise part of your routine, 7) Reduce tobacco usage, 8) Rest the eye, and properly clean the hands and eye lenses [8 - 10]. Early detection and diagnosis remain the best form of protection against visual impairment and blindness. The recommended screening for DR type 1 is three to five years after onset, while type 2 is 1 year after onset [6]. After the initial screening, subsequent examinations should be performed every six months or 1 year. Depending on the severity of DR, the clinician may recommend earlier examination. Diagnosis of visual impairment involves examination and screening of high-quality retinal photography of patients combined with a timely follow-up process. Direct and indirect ophthalmoscopy, stereoscopic retinal fundus photography, mydriatic and nonmydriatic photographs are used to detect and classify visual impairment.

The gold standard is the stereoscopic photograph in 7 standard fields (30 degrees) [11 - 13]. This gold standard is accurate and effective and requires skilled officials to operate, however, they are laborious and time-consuming. In most countries around the world, clinicians are overburdened (low doctors to patient ratio), hence diagnosis and treatment may become error-prone and time-intensive. Different medical photographs are taken daily in hospitals and clinics, these photographs are domicile in the hospitals without usage. Fortunately, these photographs will be useful for Computer-Aided Detection systems (CADs). CADs assist doctors to interpret medical imaging by processing digital images and highlighting conspicuous sections for possible disease spots [112].

Convolutional neural networks (CNN) have been developed to automatically detect and diagnose ophthalmic problems in retinal fundus images. A major advantage of CNN is the ability to automatically learn complex features and translate such features to meaningful results. The CNN system can improve health care by reducing manual procedures, hence acting as a second interpreter. CNN can automate eye screening and detect procedures, and it is flexible and available for healthcare practitioners. The goal of this study is to report the results of different CNN algorithms that were used to investigate the automatic segmentation, classification, and detection of diseases in retinal fundus images. This review will examine recent trends between the years 2015 and beyond. Analysis and performance of different CNN methods, databases, and validation metrics are examined. The relevant road maps for CNN detection, diagnosis is elucidated. From what we know, only reference [2] has examined some deep learning for retinal fundus images. This review proposes to examine the latest trends by reviewing CNN methods for both the pathology and morphologies of retinal fundus images. We will also highlight the different validation metrics, limitations, and future directions for the segmentation and classification of retinal fundus images. Our contribution is summarized as follows:

1. Analyzes of different CNN methods for segmentation and classification of retinal fundus images, different databases, and validation metrics.
2. Potential challenges and road maps used by CNN for segmentation and classification of retinal fundus images.

The rest of the paper is organized as follows. Section 2 gives the background and the review of methods for the classification and segmentation of fundus images using CNN. The different evaluation metrics used for the classification and segmentation of fundus images are discussed in section 3. Section 4 gives a brief explanation of layers used in CNN, while section 5, examines the databases, and section 6 gives statistical analyses of different algorithms. Finally, the paper is concluded in section 6.

2 Background

As early as the 1860s doctors have started looking for possible solutions to taking pictures of the eye. By the 1880s, a partial solution was developed. Doctors can now take the picture of the eye by placing a camera on the head of the patient but still need to wait for about 3 minutes to get the film developed. This was a great development, although the procedure looks simple and basic it was the first time the picture of the eye could be captured by anyone. In 1926, the first fundus camera was invented. This camera could only take photography of a portion of the eye. In 1997 (70 years later), the retinal fundus camera that could take pictures of 130 degrees was developed. By the 21st century, a non-invasive camera that could capture 200 degrees view was developed [14]. In terms of computers, the first algorithm that was executed on a machine was created by Ada Lovelace in 1843. Since then several algorithms have been created to perform specific tasks. In 1966, computers were attached to objects to see if they could identify the object (this was the beginning of computer vision). By the next decade, mathematical analysis and quantitative applications were introduced to computer vision. Scale-space representation [15], contour models [16], and Markov random fields [17] became part of computer vision algorithm.

Marvin Minisky was the first to attempt to mimic the human brain. His research opened the way into the abilities of computers to process information for decision-making. In 1959, Russell Kirsch invented a digital image scanner that transforms images into good numbers. In 1963, Lawrence Robert processed the 3D information about a solid object from 2D photographs. In 1980 Kunishiko Fukushima built the precursor of modern CNN. In 1999, David Lowes, describe a visual recognition system that uses local invariant features. In 2001, the first face detection framework with a real-time procedure was introduced. The breakthrough moment in computer vision came up in 2012 when AlexNet won the ImageNet completion. Since then, several researchers have used the CNN method to segment and classify medical images (especially retinal fundus images).

2.1 Review of methods

Several existing CNN methods to segment and classify retinal fundus images have been proposed by researchers. These methods use combinations of network architectures resulting in sophisticated AI platforms. To give the readers a grasp of the latest trends, we have categorized these CNN methods into three: (1) CNN methods to classify and segment optic disc and cup (2) CNN methods to classify

and segment arteries and veins (3) CNN methods to classify and segment retinal blood vessels. These three categories have a significant number of CNN methods that have shown good results. A block diagram depicting the different CNN approaches is shown in Fig. 4.

2.2 CNN used for Segmentation and Classification

2.2.1 CNN used for Optic Disk and Optic Cup

Automatic segmentation of the optic disk and optic cup can help remove problems encountered or envisaged in the manual procedure. However, this segmentation technique is faced with challenges such as 1) unclear boundaries 2) big variability, 3) Interference from other components in the image, and 4) mixed pathologies. To solve these problems, researchers have proposed different CNN methods. Reference [18] is an encoder-decoder network that involves two components. The first component is the feature detection (FDS) while the second is the cross-correction sub-network (CCS). The FDS preserves features by stacking two convolutional layers (3x3), a batch normalization (BN), and a rectified linear unit (ReLU). The CCS uses the subnetwork to reduce multiple pooling operations. The second layer of the network is the decoding block, this block improves the contrast and combines multiple encoding features. Overall, reference [18] used the FDS and CCS as the encoding layer and uses a decoding layer to upsample the segmentation. Fu et al [21] used a combination of UNet and probability bubbles for segmentation of optic disc (OD). The images were preprocessed with an iterative robust homomorphic surface filtering (IRHSF) method [22]. Then the UNet detects the OD and the blood vessel. The position constraint model was introduced to avoid the bright lesion which could act as a distraction. The hough transform segments the images and creates a technique to avoid DR lesion distraction. Finally, probability bubbles were modeled and used to fuse segmented components. Reference [23] segment OD with RGV generated CNN images. First, the data was prepared and augmented. Then RGV images were converted to RGB images. A two-stage technique was used to alleviate the class imbalance constraint problem. The candidate location was determined with the guided search procedure. Finally, a weighted neighborhood voting was conducted and the OD produced the localized portable position. The CNN architecture consists of a convolutional layer, max-pooling layer, fully connected layer, and a ReLU. The softmax was used for logistic regression which produced the output of the image.

Yuna et al. [25], proposed a multiscale CNN method for OD and optic cup (OC) segmentation. First, the images are cropped to get the region of interest (ROI). Then, the contrast limited adaptive histogram equalization (CLAHE) (see [26] for the usage of CLAHE for medical images) was applied to the image for enhancement. The ROI image was again transformed into polar coordinates [27], then the CLAHE and polar coordinate images were concatenated. The concatenated image was then passed to the W-Net architecture. The W-Net is an end-to-end CNN architecture that consists of feature extractor and context extractor modules. The W-Net uses these modules to create a segmentation map for the output procedure (see Fig 5 for better understanding).

Reference [28] proposed the coarse-to-fine deep learning architecture. This method consists of the input, the segmentation architecture, and output. The vessel is extracted from the original image, then a vessel density map is generated to highlight the location of the OD. Next, the UNet segments the retinal image and the vessel density map. The outcome produces two different segmentation. An overlapping strategy with the disc patch is used to execute false segmentation. Finally, the UNet is used to segment candidate regions for the final OD segmentation. Reference [29] segment and classify three components (vasculature, OD, and Fovea) with the simple CNN method. The color image was normalized from RGB to LUV color space. Thereafter the LUV color space was converted back to RGB. The normalized image was fed into the CNN which has 6 layers (2 convolution, 2 max pooling, and 2 fully connected layers). At the end of the CNN procedure, four outputs were produced.

Reference [30] proposed the deep learning enhanced CNN. The first process was to preprocess the image with the Gaussian filter and image normalization. The filter removes noise while the normalization balances the color boost region and the illumination of the image. Next, the color-texture morphological approach (combination of erosion and dilation) was used to capture the global distribution features. Then, the edge histogram texture descriptor (CLAHE and Sobel edge detector[32]) was applied to analyze and detect structures. The Watershed algorithm [31] was applied to extract contour shape and localize the disc part on the OD. A cropping method was used to Curson the image to the required size. Finally, the image was fed to the proposed deep learning CNN algorithm. This algorithm is an end-to-end decoder-encoder method with 39 layers (19 convolutions, 4 max-pooling, 4 upsampling, 4 dropout, and 11 merge layers). In both testing and training procedures, the complete RGB image was used for segmentation. Reference [33] proposed a semantic method for the segmentation and classification of OD. The input image was augmented based on the label training data. The augmentation was artificially generated based on the image. The augmentation was used for training the network and the dataset was divided into 70% training and 30% testing. Rotation (horizontal and vertical flip) of retinal fundus images was effective and enables effective segmentation results. The network uses the encoder and decoder architecture. The encoder block downsamples to keep the required

classes in the image and to remove unwanted pixels. It consists of 18 layers with 13 convolutions and 5 pooling layers. The decoder block upsamples the image with operations to generate the same size. This block has 20 layers consisting of 14 convolutions, 5 pooling, and 1 softmax layer. Finally, the pixels are classified and marked according to the classes and network. The pixel layer gives probabilities of each class with the combination of the loss function. The output of the procedure is a segmentation mask (see Fig. 6 for more details).

Reference [115] developed a coarse-to-fine segmentation process that uses UNet to obtain a rough segmentation boundary and cropping to secure the boundary area from a contour-centered image. Secondly, SU-Net (a fully convolutional network) was combined with the Viterbi algorithm to segment boundaries. Then, a data augmentation method was introduced to avoid overfitting. The network uses the UNet and the centered contour coordinates for segmentation. The SU-Net consists of three blocks: the Encoder, decoder, and sequence decoder. The first and the second blocks contain the traditional encoder and decoder (segmentation mask) of UNet with skip connection, convolution, ReLU, upsampling, and downsampling. Sequence encoder layers consist of two parts: gateway module and cascaded gate units. The gateway module gets its inputs from each layer of the decoder, these inputs are processed and concatenated before sending to the gates unit. The gateway module consists of 3 upsampling and 3 convolutions + sigmoid layers, while the gate unit consists of ReLU and softmax layer. To model, the interaction of prediction and spatial constraints the Viterbi algorithm [117, 118] was used to decode the output of the sequence decoder. Reference [119] proposed the attention-based fully connected CNN (AFCNN). The images were resized to 512 x 512 pixels, then morphological operations (opening, closing, and erosion) were performed on the resized image. A cropping procedure was then placed on the image with a bounding box of the mask. The AFCNN has 19 layers consisting of 3 attention blocks, 12 convolution layers, 2 dropout layers, and 1 softmax layer. Reference [120] and [121] both used the CNN for segmentation of OD in RF images. In [120] image patches were first extracted then the global contrast normalization was used to brighten the image. The zero-phase component analysis was then applied to the image before it was augmented and fed to the CNN. The CNN is a simple traditional method with convolution and pooling layers arranged sequentially. Images segmented with the CNN were again segmented with the fuzzy c means [122] for object detection procedure. Meanwhile, in [121] a 9 layer CNN was used to segment the image. This network consists of 3 convolutional layers, 4 ReLU, 2 max-pooling, 1 fully connected layer, and 1 softmax. **The Comparison of different methods in this section is available in Tables 1 and 2.**

Table 1: CNN used for Optic Disk and Optic Cup

Author	Ref	Year	CNN Name	Inspiration for research	procedure	Additional comments
Wan et al	[18]	2021	Asymmetric deep learning network	UNet [20] M-Net [19]	Segmentation	Uses the FDS and CCS for encoding while the decoder upsamples
Fu et al.	[21]	2021	Fusing UNet with probability bubbles	UNet	Segmentation	Preprocessing and a combination of UNet and probability hough bubbles
Meng et al	[23]	2018	RGV generated CNN model	LeNet-5 [24]	Segmentation	Simple CNN model that works with RGV images
Yuan et al	[25]	2021	Multi-scale W-Net	M-NET and UNet	Segmentation	Pyramid W- shaped backbone network for OD and OC
Wang et al	[28]	2019	Coarse-to-fine deep learning	UNet	Segmentation	Vessel extraction and vessel density map with the UNet for OD segmentation
Tan et al	[29]	2017	Single convolutional neural network	Multiple segmentation	Segmentation	Single CNN to segment vasculature, OD, and Fovea
Veena et al.	[30]	2021	Deep learning enhanced CNN	Encoder-decoder CNN	Segmentation	Combined preprocessing, enhancement, and deep CNN model for OD segmentation
Imtiaz et al.	[33]	2021	Label based encoder and decoder semantic segmentation	Encoder-decoder CNN	Segmentation	Augmentation based label semantic segmentation for OD
Xie et al.	[115]	2020	SU-Net and Viterbi algorithm	UNet, dilated CNN [116], Viterbi algorithm [117, 118]	Segmentation	UNet, S-UNet, and Viterbi algorithm connected to segment vessels in RF images.
Sadhukhan et al.	[119]	2020	AFCNN	FCNN	Segmentation	Attention mechanism combined with FCNN
Priyanka et al.	[120]	2017	Patches CNN	CNN, Fuzzy C Means	Segmentation	Combined the CNN with the Fuzzy C means for segmentation
Raja et al.	[121]	2020	Traditional CNN	CNN	Segmentation	Used the tradition CNN for segmentaation

Table 2: Pros and cons of CNN used for Optic Disk and Optic Cup

No of the Databases used	Ref	Advantages	Disadvantages	Accuracies (Accuracy)	Accuracies (AUC)
3	[18]	Can train a model directly using the end-to-end procedure on source and target. Easily converts data from one form to the other.	May become too lossy. If not properly configured, can produce improper decoding output. Easy to miss important features from the encoder	0.937	-
4	[21]	Can easily detect noise because of the excellent pixel capture procedure. Images with less symmetric components are captured faster.	Easy to produce misleading output Symmetric components produce terrible results	0.99	0.99
4	[23]	Easy to use with less dependent on computational space Easy to detect and extract important feature	The process is too cumbersome. May produce poor accuracies for certain images because the architecture is not deep	0.98	-
3	[25]	Gives a more flexible and detailed images representation. May produce very good accuracies since it uses a very deep network	Require large computational space. Difficult to implement since it requires some little human interaction.	0.95	0.99
6	[28]	Allow usage of global location and context at the same time. Does not need multiple runs to get acceptable segmentation	Accuracies may be poor because the network is not deep and features are not robust	0.93	0.97
1	[29]	Segmentation and classification may be effective since the image was already reduced between 0 and 1. Noise is reduced effectively.	Produce unstable procedures during training. Highly dependent on data for effectiveness.	0.96	0.95
1	[30]	Can execute and correlate features that enable faster training and learning	Uses several deep networks that require large computation Produces several weight parameters that make training slow	0.98	
2	[33]	Prevents data scarcity by adding more training data to the model. Resolves class imbalance and increase generalization ability	Provides data bias that leads to suboptimal performance. Accuracies may reduce in the very noisy image due to lack of preprocessing method.	0.86	0.99
3	[115]	Avoid overfitting with the data augmentation procedure. Highly satisfactory bit error rate performance with the high-speed operation and ease of implementation.	Accuracies may reduce in the very noisy image due to lack of preprocessing method.	-	0.97
6	[119]	Resized images help to decrease computational time. Effective identification of information in input to accomplish a task.	Difficult to parallelize the system and could be time-consuming.	-	-
1	[120]	Allows gradual membership and can cluster at points measured at pixel degrees May give good accuracy for overlapping datasets.	Involve so many iterations to give good accuracy. This could involve a lot of time.	0.95	-
1	[121]	In the recognition, framework accuracy may be very high due to the added ReLU	Do not encode the orientation object which can cause a vanishing	0.90	-

layers.

gradient.

Does not require spatial input data, and requires so much training data for a good result.

2.2.2 CNN used for Arteries and Vein

Reference [34] proposed the encoder-decoder CNN model for the segmentation of arteries and veins. The median filter (with kernel size equal to one-tenth) was applied to correct illumination from the retinal fundus image. Then the image was passed to the CNN model. The encoding layer encodes inputs into smaller vectors while the decoding layer upsamples. Each encoder layer has 3 stacked convolutions followed by a max-pooling layer. The decoding layer has upsampling layer followed by a convolutional layer. In the encoding block, there are 32 feature maps for the first, 64 for the second, 128 for the third, 256 for the fourth, and 512 for the fifth. In the decoder, the number of features map is reduced downward. A final convolution layer reduces the map from 16 to 3 classes (background, arteries, and vein). Like previous researches, Morano et al. [35] proposed a simultaneous segmentation module inspired by the UNet method. This research preprocessed the fundus images with a local intensity normalization and channel-wise global contrast enhancement [35]. Thereafter, the image was passed to the CNN network. This Network uses UNet to predict the mask and was channeled to three parts (arteries, veins, and vessels). The binary cross-entropy and manually annotated segmentation mask of the structure were combined to produce the final mask. The Comparison of different methods in this section is available in Tables 3 and 4.

Table 3: CNN used for Arteries and Vein

Author	Ref	Year	CNN Name	Inspiration for research	procedure	Additional comments
Girard et al	[34]	2019	Joint segmentation model	UNET	Segmentation	Uses the median filter and encoder and decoder semantic segmentation.
Morano et al.	[35]	2021	Simultaneous segmentation	UNET	Segmentation	Preprocessed and UNET multichannel for segmentation

Table 4: Pros and cons of CNN used for Arteries and Vein

No of the Databases used	Ref	Advantages	Disadvantages	Accuracies (Accuracy)	Accuracies (AUC)
2	[34]	Effective in edge preservation. Require a fewer number of images to produce accurate results.	The small noise ratio can break up image edge and produce false noise on the edge which could affect the accuracy	0.96	0.98
2	[35]	Do not need a lot of data to perform optimally	May produce errors from the normalization process. Accuracy may not be excellent since the network is not deep.	0.96	0.97

2.2.3 CNN used for Retinal Vessel

Retinal vessel segmentation is a problem that has been long-standing in medical image analysis [148, 149]. Several challenges characterize the vessel segmentation, they include:

1. Presence of several abnormalities of varying sizes and shapes: Several abnormalities surround the vessel, these abnormalities can affect the effective segmentation of the vessels in RF images.

2. The fewer annotated data: The limited numbers of elucidated data can result in overfitting, hence this is a major challenge when segmenting vessels in FR images.
3. Vessel structural differences: retinal vessels are characterized by thick and thin structures, hence it is difficult to specify a particular model or network that is suitable for all kinds of vessels when segmenting.
4. Unstructured prediction: Pixel classification is different from vessel segmentation, therefore it has become difficult to predict the structure.

In the light of the above challenges, several authors have used CNN methods to segment vessels in RF images. Budak et al [36] proposed the densely connected and concatenated multi-encoder-decoder CNN (DCCMED-CNN). The DCCMED uses the patch-based learning network and consists of a training and testing phase. For the training, the inputs are color patches extracted from raw retina images without preprocessing. The DCCMED was utilized as the network for the model and segmented binary masks were produced as the output of the training phase. The training phase also has weights, this weight were trained by stochastic gradient descent methods [37]. The DCCMED consists of three encoder-decoder blocks. The first block has 2 max-pooling layers, 2 max unpooling layers, 8 concatenated convolution, Batch normalization, and ReLU layers. The second and third blocks have the same configurations as the first. Finally, a softmax layer was used for prediction.

Tang et al. [38] proposed the multi-proportion channel ensemble model (MPC-EM) for the segmentation of retina vessels. The MPC-EM consists of 5 submodel networks. The Green and Red channels were divided into 5, with the green channel taking the highest chunk ($0.69G + 0.4R$, $0.7G + 0.3R$, and so on.). These channels are preprocessed and passed to the sub-models. Each submodel has a Net-like structure of encoder-center-decoder. The encoder (consisting of convolutional, max-pooling, and ReLU layers) converts the image into the feature vector representation. The decoder (consisting of ConvTranspose, convolutional, and ReLU layers) converts the feature vectors into a probabilistic map. A center architecture was used as a transitional region for the shape of feature vector adjustment. Each submodel jointly used shallow localization to classify pixels into images. To optimize the subnetworks the triple convolutional residual block was used to enhance, ease and avoid vanishing gradient [39].

Reference [40] proposed the RCNN-based junction proposal network. This network takes an input 128×128 image patches and outputs the bounding boxes of potential junction location. The network consists of four parts: 1) backbone for feature extraction 2) region proposal network 3) head module for bounding-box regression, and 4) classification for mask generation. For the backbone feature extraction, the ResNet [50] was used and a pyramid structure [43] was adopted to consider multiple scales. The region proposal network takes image patches as input and outputs rectangular junction proposal location with the fully convolutional network (kernel size 3 and stride 1). The resulting features are a 1×1 convolutional layer producing the region regression and region classification as output. The junction classification network uses the pixel-wise difference of segmentation task loss and classification task loss. The proposed network is a combined multi-task CNN with 27 layers. The output of the ResNet50 is processed for classification, while the output of the convolutional layer is used for segmentation. Overall, the model has 15 convolutional, 4 duplicated feature maps, and 2 fully connected layers.

Reference [44] used the shallow UNet to segment retinal fundus images. This method has six stages: 1) Phase image registration, 2) Vessel probability map generator, 3) Postprocessing, 4) Second segmentation, 5) Region of interest selection and 6) Vessel diameter measurement. The image registration stage aligns the fundus image spatial domain to the moving spatial image domain [45]. The vessel probability stage uses the encoder-decoder CNN (shallow UNet model) for extracting vessels from fundus images. The Shallow UNet model consists of 11 convolutional layers, 5 dropout layers, 2 max-pooling layers, 2 upsampling layers, and 1 softmax layer. In the postprocessing stage outputs of the shallow UNet were thresholded, binarized, and cropped to a region of interest determined by an optic disc size. Then a manual delineation of veins and arteries (marked in red or blue) are processed to the shallow UNet again for segmentation. This time, the segmentation output is the optic disc. The shallow UNet was assigned for segmenting both the vessel and the optic disc (at different stages). Finally, the retinal diameters were measured using a similar method proposed by [46] (central venular equivalent (CRVE)).

Reference [47] proposed the multipath cascaded UNet (MCU-Net) for segmentation and classification. The MCU-Net takes as input three data (raw FFA, small-scale FFA, and large-scale FFA) and fuses vessel features from these inputs to generate a vascular probability map as output. The MCU-Net consists of an attention gate [48] and a residual recurrent unit [49]. The processed inputs are cascaded with the UNet architecture to produce the final output. The MCU-Net has two blocks 1) the refinement block, and 2) the FFA image fusion block. The FFA image fusion block consists of three fusion strategies (early fusion strategy, late fusion strategy, and intermediate fusion

strategy). These fusion strategies fused the inputs for further processing. The refinement blocks accept the output of the fusion process to produce the final mask (for more on this method see fig 7).

Reference [50] proposed the Nested U-shaped attention network (NUA-Net) for the segmentation and classification of retinal images. The images were first transferred from RGB color space to LAB space, then CLAHE was applied to the images. Thereafter, the images were transferred back to RGB color space. Similar to the experiment conducted in [51], the green channel images were used as the network inputs. The NUA-Net extracts patches as inputs and predicts pixel-wise soft segmentation. This network consists of an encoding stage and 4 decoding stages. The resolution of feature maps was halved once the scale increases. A 3 x 3 convolution layer was used to extract shallow features, then downsampling of the feature vector was performed. Each encoding stage has a 2x2 max-pooling followed by convolution with batch normalization, ReLU, and dropout. A simple bottom-up approach is used to derive features at a larger scale. The multiscale upsampling attention (MSUA) model was developed to harness mutual relationships among blocks. A joint loss was adopted to put supervision on each decoding stage.

Guo et al. [52] proposed the multiscale deeply supervised network with short connection (BTS-DSN). This network uses short connections to transfer semantic information between side-output layers. Two approaches were considered: the bottom-top short connections and the top-bottom short connections. The RF images were divided into 4 combined pairs of convolution+ReLU+max-pooling. The short connection approach then carries the signal to the upsampling layer which is also divided into 4 pairs. The process of upsampling further moves the connection to the sigmoid to translate the image into a mask. The four masked images are then fused to form a single image mask. The key element of this network is the top-bottom, and bottom-top short connection approaches. A switch of connectivity within layers gives the BTS-DSN a flexible procedure. Reference [53] proposed the multiple deep convolutional neural network (MDCNN) for a formulated classification and segmentation task. The MDCNN was constructed by cascading multiple networks with the same structure. The training procedure was performed by using the incremental learning strategy which improves the network performance. Training with the same procedure was introduced to overcome the poor performance of the previous CNN. The final output was determined using a majority voting method on the MDCNNs results.

Noh et al. [35] proposed the scale-space approximation for multi-scale representation in CNN (SSANet). A 1-dimensional signal was used as an example for analyzing convolution and upsampling in the frequency domain. The SSANet consists of 3 blocks, (feature generation, feature aggregation, and inference) with 33 layers. The feature generation block has 21 layers consisting of 1 convolution layer, 3 upsampling layers, and 17 ResBlock layers. The feature generation block gets the input and extracts features from the inputs. At every upsampling interval, there is a connection to the feature aggregation stage. The feature aggregation block is the intermediary stage that connects outputs of the generation stage to the inference stage. The aggregation stage perform two key procedure: it moves input before each upsampling in the generation block. Secondly, it accepts inputs from the final block of the generation using 9 layers (5 Convolutional layers, and 4 upsampling layers) for this procedure. The inference block collects inputs from the aggregation and transforms these inputs to mask. The aggregation concatenates upsampling procedures and sends them to the inference. The inference has 3 layers (2 convolutions, and 1 sigmoid layer).

Reference [57] combined the size-invariant feature maps [58] with the dense connectivity [59] (SID²Net) for the segmentation and classification of RF. The size-invariant feature map reduces the loss of small blood vessels, while the dense connectivity reduces the computational cost. The SID²Net is a bottleneck that has the green channel as the inputs. Two bottleneck modules and three dense blocks were used to extract features, these features are finally merged by two convolutional layers and a sigmoid layer for prediction. The network is divided into two bottlenecks (bottlenecks 1 and 2) with each bottleneck having 36 and 48 output feature maps respectively. The network has 3 dense connectivity blocks. To generate probability maps, the output feature maps of the third dense connectivity are integrated. An ablation experiment was carried out dividing the network into the dense network (DNet), and DNet with size-invariant feature maps (SIDNet). Reference [60] used the multi-instance heating regression to predict RF image segmentation. This method predicts binary maps with the pixels corresponding to the location and labeling of the positive class of the ground truth. The RF images are passed to the UNET framework which extracts features creating multi-instance heatmaps and the local maximal. Finally, the results were interpolated back into the original RF images. The UNet architecture used in this research has 19 layers consisting of 1 input, 9 convolution+ReLU, 4 Transpose convolution, and 1 convolution (output Network).

Reference [61] used the vessel-specific skip chain convolutional network (VSSC Net) for blood vessel segmentation. The VSSC Net involves two stages: preprocessing and segmentation. The preprocessing stage converts RF images to grayscale then the adaptive fractional difference approach [62] is applied to the grayscale image to form the first plane of interest. The CLAHE is applied to the grayscale image to form the second plane of interest. The CLAHE image is applied to the Gaussian filter to form the third plane of

interest. The intensity of the images is reduced by a factor of 2, before concatenating them to give the final preprocessed image. The segmentation stage (VSSC Net) is an end-to-end framework that takes input images of arbitrary size producing a probability map. VSSC Net has two components: base network architecture and novel architecture. The base network consists of different convolutional layers splitted into 4 pairs. The VGG-16 [63] was used as the base network. The proposed novel network has two blocks placed on top of the base network. The first block VE_1 consists of 4 vessel-specific convolutions (VSC) and 4 skip chain convolution (SC) layers. The second block VE_2 consists of 3 VSC layers and 3 SC layers. A skip connection is applied to connect the VE_1 to the VE_2.

The attention-based before-activation residual UNet (BSEResU-Net) proposed by [64] was inspired by the modified UNet architecture. BSEResU-Net exploits the attention mechanism and the dropblock regularization method to reduce overfitting. The images were preprocessed by transforming RGB images to grayscale, then the grayscale images are normalized. The CLAHE algorithm was applied to the grayscale image, then a gamma adjustment was applied to the CLAHE image. The preprocessed image was fed into the network. The BSEResU-Net consists of two parts: BSE residual block and the ResU-Net. The BSE residual block consists of a residual layer, pooling layer, ReLU, sigmoid, and 2 convolutional layers. The ResU-Net has 33 convolutional layers with 16 residual operations, 2 transpose convolutional layers, 2 downsampling layers, and 1 output map. Reference [65] proposed the multipath scale network (MPS-Net) for retinal vessel segmentation. The MPS-Net is an end-to-end network that uses one high-resolution RF input and produces a probability map with two low resolutions as output. The image was first converted to grayscale, then passed on to the MPS-Net. This network has 16 layers and three branches. The first branch has 8 H layers, while the second and third have 4 M and 3 L respectively. The network has 13 multi-path scale modules, 3 convolution +ReLU, 3 Normalization +ReLU, and 1 cropping layer. The multi-path scale module has 3 regional paths concatenated together and arranged horizontally to produce the output. The range entropy [67] definition was introduced to describe vessel information of the feature maps.

Reference [68] proposed the multipath CNN for RF segmentation. This network converts the original image to low-frequency and high-frequency images with the low-pass Gaussian filter and the high-pass Gaussian filter. The low-frequency image is sent to the CNN for segmentation. The CNN consists of a convolution downsampling and convolution upsampling. This CNN has 32 convolutional layers with four blocks of 64, 128, 64, and 32. The downsampling part performs max-pooling while the upsampling employs bilinear interpolation. Meanwhile, the high-frequency image is sent to another CNN with the encoding and decoding regions. This CNN consists of max-pooling + Convolution, Max-pooling +convolution+upsampling, and upsampling + Convolution. Finally, the output of the first and second CNN are concatenated (fusion) to produce the final segmentation. To find out if there is a difference between the preprocessed image and the images without preprocessing, reference [69] used the sin-Net for segmentation of vessels in RF images. The authors segment the images with and without the preprocessing stage. To preprocess, the CLAHE and the multi-scale top hat transform (MTHT) [70] are used to enhance image contrast. The Sin-Net architecture consists of 17 layers comprising of 11 convolution operations, 2 upsampling layers, 2 down-sampling layers, 1 output and input layer each. The upsampling and down-sampling layers are sandwiched in between the convolution operations. Results indicate that the preprocessed images when fed to the network performed slightly better than the images without preprocessing.

The usage of reinforcement learning in RF images is gaining prominence. Reference [71] used CNN with reinforcement learning to segment vessels in RF images. The images are divided into smaller patches and sent to the CNN for training. The CNN has five components: convolution, pooling, dropout, fully connected, and loss function. The dropout increases the generalization ability of the network while the fully connected layer acted as the classifier that connects the CNN to the reinforcement method. The reinforcement sample learning component reinforces the samples with poor performance in training. Overall, the network has 2 convolution layers, 2 max-pooling layers, and 1 layer each for dropout, fully connected and loss function layers. Deep CNN [73] has received tremendous recognition in medical image processing. As an example, wu et al. [72] proposed the network followed network (NFN⁺). The NFN⁺ was preprocessed with the CLAHE algorithm and patched. The enhanced patched image was fed into the network for training. The NFN⁺ consists of four modules: 1) encoder and decoder of the front network, 2) encoder and decoder of the followed network, 3) front group of intra-network skip connection 4) Second group of inter-network skip connection. The intra-network skip connection connects the first and second modules, while the inter-network skip connection bridges the second and third modules. Information gathered at each skip module is incorporated sequentially into the next module. Overall, the NFN⁺ has two connections (front and followed network) with 10 combined parts of convolution, batch normalization, and dropout. At every interval, the network is concatenated.

Fully convolutional networks (FCN) have gained relevance in tasks related to nonmedical imaging. However, such tides are changing, reference [74] used the FCN for segmenting retinal vessels in RF images. They used the method adopted by [75] to pad the region of interest to avoid excessive contrast enhancement at the border of the image. Gaussian filter, gamma correction, and CLAHE were applied to the image. The preprocessed image was passed to the UNet architecture for segmentation. Finally, the output from the network was

subtracted from the original image. Reference [76] proposed RV-Net for vessel segmentation. This method preprocesses the RF images by replacing the black area with an average color (see [78]), then the image is converted to LAB. The CLAHE algorithm is applied to the image and the channels are merged and converted back to RGB. The preprocessed image is augmented by performing image transformation, cropping, and patch extraction [79]. The images are fed into the network for segmentation by the RV-Net. This network (RV-Net) is a U-form network that consists of upsampling and downsampling frameworks. The downsampling has 6 blocks consisting of convolution +ReLU, LCM, and max-pooling. Each layer of the downsampling has a max-pooling at the end. Meanwhile, the upsampling layer has 6 blocks consisting of upsampling, convolution+softmax, and LCM. Reference [80] proposed the hybrid CNN and ensemble random forest (RFs) [81] method. The CNN was used for segmentation while the RF was the trainable traditional method used for classification. The CNN has 5 layers consisting of the convolution, subsampling, and fully connected layer. The subsampling layer is the local averaging method that reduces the spatial resolutions of feature maps.

Since the inception of CNN, several versions and modifications have been proposed in the literature. The research by Hu et al. [82] proposed the multiscale CNN with cross-entropy loss function. The original RF image was augmented, then fed into the network. The multiscale CNN has 4 stages, the first and second stages consist of 4 convolutions and one max-pooling layer each. Two convolution layers are concatenated and the max-pooling transfers to the next stage. The 3rd and 4th stages consist of 6 convolutional layers with a single max-pooling layer. At the starting point of the network, there are 20 convolutions, and 3 max-pooling layers. At last, each map in every stage is upsampled to the original size to either connect to the corresponding side-output or fuse to the feature map. The improved entropy loss function considers the sample balance and inclined the leading process to segment vascular parts (see Fig. 8).

Reference [84] proposed the symmetric equilibrium generative adversarial network (SEGAN) for vessel segmentation. The SEGAN is an end-to-end synthetic neural network that utilizes the adversarial principle. Three principles are used in this research: SEGAN, multiscale feature refine block (MSFRB), and attention mechanism (AM) [86]. The MSFRB is used to extract the shallow-layer features which are high in resolution but have low semantics. The AM on the other hand is used to allocate weights to channel in the MSFRB. Both MSFRB and AM are part of the SEGAN framework. The SEGAN is a combined U-shaped network that has the G and D procedure. The D distinguishes details while the G fakes the details and enhances recognition. The network consists of 20 layers of two end-to-end networks. At the G network, the MSFRB and AM are utilized, concatenated, and passed on to the next layer. Overall, there are 13 traditional layers, 5 MSFRB, and 5 Am layers. A downsampling, upsampling, and skip layer were used in the network. An ablation experiment using the UNet was adopted in this research.

Multitask segmentation is becoming popular in deep learning architecture. A multitask segmentation creates a procedure to segment images in the task of different positions. The research by [87] proposed a hybrid multitask deep learning for segmenting vessels. The original image was annotated before being fed into the deep learning algorithm. The network has two modules: 1) multitask segmentation, 2) fusion network module. For both modules, the improved UNet framework was adopted for segmentation and fusion. The network is an encoder-decoder segmentation consisting of 20 layers. The network consists of 11 convolution +Batch normalization +ReLU, 4 max-pooling, 4 upsampling, 1 sigmoid, input and output layer. The output of the segmentation is passed to the fusion layer for final output.

The deformable U-Net(DUNet) proposed by Jin et al. [88] is a U-shaped architecture with an encoder and decoder framework. Some of the convolutional layers in the traditional UNet are replaced with deformable convolutional blocks. The DUNet integrates the low-level features with high-level features and receptive fields are trained. The design is constructed with 4 convolution layers, 4 batch normalizations, and 4 ReLU layers. In addition, the model consists of 4 convoffset, a global average pooling, 1 dense layer, and 1 softmax. Three deformable convolutional blocks are mapped in the middle of the network. These deformable convolutional blocks have 12 layers with a final output generated by softmax. Reference [90] proposed the strided FCNN for the segmentation of vessels in RF images. The images were preprocessed with the morphological tactics and the principal component analysis (PCA) [91]. The morphological tactics are used to remove the uneven illumination issue achieving the uniform contrast. The PCA transforms the image to grayscale. The network has 5 fully consecutive convolutional blocks with sizes ranging from 16,32,64,128, and 256 features. Apart from the two layers, all the other blocks have three convolutional layers (in the encoder). The encoder has 18 layers with 13 convolution + leaky ReLU [93] and 5 strided convolutional layers. Meanwhile, the encoder has 20 layers with 10 convolution+LReLU, 5 upsampling +ReLU, 4 concatenation + Convolution +LReLU, and 1 Convolution +Sigmoid layer. There is no ablation experiment in this research. Reference [94] proposed the end-to-end improved CNN for vessel segmentation. This network used the multi encoder-decoder principle and a new progressive reduction model that was integrated into the network. The network has 4 interconnected components (multi-encoder and parallel components, RGB-encoder and green channel encoder, decoder component, progressive reduction components). The RGB consists of six levels (convolution, spatial dropout, batch normalization, and max-pooling layers), while the decoder consists of

five levels (deconvolution, concatenation, and batch normalization). The last module consists of convolution, batch normalization, and concatenation layer. Data augmentation was performed to generate more data for the network.

Reference [95] proposed the contextual information enhanced UNet (CIEU-Net) with dilated convolutional module for vessel segmentation. The cascaded dilated module and the pyramid module are integrated to form the segmentation network. The proposed network is a UNet with a modification of cascaded residual dilated module and pyramid module. There are 13 blocks with 47 layers used as the baseline, 5 Residual blocks, 2 convolutional layers, and 3 dilated convolutions. Reference [96] proposed the scale and context-sensitive network for the segmentation of vessels. The model consists of three modules: scale-aware feature aggregation (SFA), adaptive feature fusion (AFF), and multi-level semantic supervision (MSS). The SFA adjusts the receptive field dynamically to extract features. The AFF guides the fusion between features efficiently. While the MSS is used to learn distinctive semantic representative. The SFA consists of multiscale features extraction (MFE) and dynamic features selections (DFE). The SFA has 6 convolutions, 2 convolutions + ReLU, and softmax. The AFF module used the squeeze and extraction operation to model correction among features channels. Finally, the MSS fuses the channel masked to produce the final prediction. The SFA, AFF, and MSS are arranged on the SCS-Net. Overall, the SCS-Net consists of 17 layers for the segmentation of vessels. An ablation experiment was carried out on this research and augmented before training. Reference [97] proposed the enhanced encoder atrous UNet (EEA-UNet) for retinal vessel segmentation. The images were preprocessed with the CLAHE and resized to 512 x 512. Post-processing was done by morphological operations to remove isolated false positives. The EEA-UNet is an asymmetric contraction and expansion path that replaces all the convolutions as the atrous convolution to increase the receptive field. The contracting part has 5 blocks containing 2 atrous convolutions, batch normalization, pooling, and ReLU layer. The atrous convolution reduces the image size without losing the significant features in it [98]. Overall, the EEA-UNet consists of dilated convolution + batch normalization+ ReLU, max-pooling, Depth concatenation, and transpose convolution.

Reference [99] proposed a U-shaped deep learning fusion network (DF-Net) for vessel image segmentation. The method involves 4 stages: multiscale fusion, U-shaped network, feature fusion, and classifier fusion. The original image was multiscaled by the image pyramid [100] and constructed on a multiscale input integrated into the encoder path for information fusion. The U-shaped network collects the pyramid images and processes them for transmission to the next block. The network is an encoder-decoder network that consists of 2 convolution layers, max-pooling with ReLU. The down-sampled feature map is concatenated and doubled to give proper learning. Similarly, the decoder has 2 convolutions, followed by up-sampling with the number of features map halved. The vessel fusion module is attached to the decoder and enhanced with the corresponding output features. This network is combined with the Frangi filter and a deep neural network is trained. Finally, the vessel fusion module integrates masked images to produce the final segmentation. Data augmentation was conducted in this research.

Recently, multi-scale or multitask methods have been used in CNN. The research by Tang et al. [101] adopts this multiscale method. The authors proposed multi-scale channel importance sorting (MSCS) for vessel segmentation. First, the CLAHE algorithm was used to enhance the image before it was fed to the network. The MSCS is an encoder-decoder that consists of 3 encoder and 2 decoder blocks. Each encoder block consists of multi-scale, channel importance, and a convolution layer. At the end of each encoder block, a max-pooling operation is conducted. The multi-scale optimize local topology, while the channel compresses, and regularizes the network to prevent overfitting. Between the encoder and decoder, the spatial attention mechanism was used instead of the traditional skip connection to readjust the output and characterize the encoder generating the attention coefficients. The research by [102] proposed the cascaded attention residual network (AReN-UNet), that integrates the attention and residual modules. The encoder and decoder of the networks are connected to produce the final output. The upper block consists of 16 channels and uses lesser computational memory when compared to the lower part of the network. The downsampling layer is used in the encoder while the up-sampling is used in the decoder. The aggregated residual module consists of concatenated max and average pooling and a shared MLP [105] with the sigmoid layer. Meanwhile, the spatial attention block concatenates the max and average pooling with a convolution layer and sigmoid. A skip connection is used to pair the encoder and the decoder. Reference [103] proposed the multiscale dense network (MD-Net) that makes good use of the multi-scale features and the encoder features. This network is preprocessed with the CLAHE algorithm. The data augmentation and patch segmentation were applied to the CLAHE image to avoid overfitting. A residual atrous spatial pyramid pooling (Res-ASPP) was blended into the error framework and the dense multi-level fusion merges the features in the encoder and decoder. A squeeze and extraction (SE) block is applied to the concatenated layer for effective feature channels. The Res-ASPP has 12 layers all of which are convolution layers with varying dimensions and sizes. The multi-level fusion mechanism and SE block perform the fusion procedure in the network. Overall, the MD-Net has 3 Res-ASPP layers sandwiched in the encoder framework while there are 3 SE blocks in the decoder framework. The skip connections connect the encoder to the decoder. Reference [106] used the combination of edge detection and neural network to segment vessels in RF images. The image was preprocessed by an iterative algorithm that removes

strong contrast between the fundus of the retina and the outer region (see [107]). Then the median and Roberts filters were used to remove noise. The method used feature vectors with eight characteristic pixels. The feature vectors include 1) image gradient obtained with edge detection (Prewitt, Sobel, Canny and Gaussian [108, and 109]), 2) the Laplacian of Gaussian filter, 3) morphological transformation (erosion, dilation, and top hat filtering [110]). The cascaded feed-forward network was used for segmentation. The network has 1 input and output each, and 4 hidden layers. The hidden layer has different neurons with a hyperbolic tangent sigmoid as a transfer function.

Reference [111] proposed the simplified UNet for the segmentation of RF images. A combination of the residual block and batch normalization in the upsampling and downsampling layers produces the required segmentation results. Different patches are extracted from the original images as inputs and trained with a novel loss function to generate the possibility of each input pixel. Then the probability map is binarized with the thresholding algorithm to generate the vessel segmentation. The simplified UNet has 10 blocks consisting of 1 CONV_ReLU1 layer, 1 convolution layer, 3 Block2 layers, 2 Block I1, and 3 Block I2 layers. Skip connections are used to link Block I1 and Block I2 together. The Block2 layer consists of transpose convolution, concatenation layer, and 2 CONV + ReLU layers. Meanwhile, Block I1 and I2 consists of 1 CONV +ReLU, and 4 Batch Normalization + ReLU+ Convolution. Reference [113] combined the attention-based neural network with transfer learning. This research uses the optimized learning method to classify and grade RF images. The attention mechanism adds attention over pixels near the vessel. The Gaussian filter was used to normalize color balance and illumination, then the data was augmented. Finally, the attention network learns and a traditional CNN performs feature extraction. The fully connected layer and the softmax were used for classification. The network comprises of pertaining inception V3, a combination of batch normalization and dropout, attention layer (has 4 convolutions), and the classification layer (convolution, pooling, 2 fully connected layers, and softmax). The softmax graded the health risk with 0 as bad and 2 as good. Reference [114] used morphological process, thresholding, edge detection, and adaptive histogram for segmentation, while CNN was used for classification. After preprocessing and segmentation, the image was fed into the trained CNN for classification (either normal or diseased). The trained CNN has individual neurons tiled to the system such that they respond to overlapping regions. The network has 3 convolutional layers, and 1 layer each for flattened, fully connected, and softmax. The images are classified into 4 namely: Normal, mild, moderate, and severe. Reference [123] used the CNN-RRN to segment retinal images. The image was first preprocessed with the median filter for denoising and smoothing. Then, the image was resized with dual-tree complex wavelet transform and then decomposed into sub-bands, this sub-band was further fed as input into the network. The classification is done using renewal networks (CNN with recurrent neural network concept [124]). The RNN captures information from sequence and time-series data. The concept of CNN was also incorporated by adding recurrent connections to each convolutional network layer. **The Comparison of different methods in this section is available in Tables 5 and 6.**

Table 5: CNN used for Retinal vessel

Author	Ref	Year	CNN Name	Inspiration for research	procedure	Additional comments
Budak et al	[36]	2020	Densely connected/concatenated multi encoder-decoder CNN	Feedforward CNN	Segmentation	Three encoder-decoder blocks with a final softmax layer.
Tang et al.	[38]	2019	Multi-propotion channel ensemble model	Ensemble model	Segmentation	Sub-channel, submodel CNN segmentation
Zhao et al	[40]	2020	RCNN-based junction refinement network	Masked RCNN-model [41]	Segmentation	Multi-task combined RCNN method for segmentation and classification
Yuan et al	[44]	2021	Shallow U-Net	UNet	Segmentation	Image registration combined with a shallow UNET for segmentation
Sun et al	[47]	2021	Multi-path cascaded UNet	UNet	Segmentation,	Refinement and FFA image fusion block for segmentation of arteries
Zhao et al	[50]	2021	Nested U-shaped attention network	UNet	Segmentation	Nested U-shaped network to segment and classify RF images.
Guo et al.	[52]	2019	Bottom-top and top-bottom short connection deep supervised network	Deep supervised network	Segmentation	The deeply supervised network embeds the upsampling and max-pooling with the weight fusion.
Guo et al.	[53]	2018	Multiple deep CNN	Deep CNN [54]	Segmentation	Multiple DCNNs cascaded for segmentation.
Noh et al.	[55]	2019	Scale-space approximated CNN	DRIU [56]	Segmentation	Generation, aggregation, and inference blocks for segmentation.
Zhuo et al.	[57]	2020	Size-invariant and dense connectivity network	DenseNet Network [59]	Segmentation	Synchronization of dense connectivity and size-invariant feature maps.
Hervella et al.	[60]	2020	Multi-instance heat map regression	DNN	Segmentation	Combination of UNET and instance heat map for detection.
P. M Samuel & T. Veeramalai	[61]	2021	Vessel Specific Skip chain CNN	Fully convolutional networks	Segmentation	Preprocessing and VSSC Net segmentation architecture

D. Li & S. Rahardja	[64]	2021	Attention-based before-activation residual U-Net	Modified UNet	Segmentation	Preprocessing BSE residual layer and ResU-Net residual layer.
Lin et al.	[65]	2021	Multi-path scale network	HR-Net[66]	Segmentation	Multi-path scale module combined with several other modules.
Tian et al.	[68]	2020	Multi-path CNN	UNet	Segmentation	Two CNN frameworks for low and high-frequency images.
I. Atli & O. S. Gedik	[69]	2021	Sine-Net CNN	Fully CNN	Segmentation	Uses 17 layers for segmentation of RF images
Guo et al.	[71]	2018	CNN with reinforcement sample learning.	Reinforcement learning	Segmentation	Uses 17 layers for segmentation of RF images
Wu et al.	[72]	2020	A network followed network.	Deep CNN [73]	Segmentation	Front and followed network sandwiched with components.
Hemelings et al.	[74]	2019	Fully convolutional network.	UNet	Segmentation	Preprocessing with UNET framework for segmentation.
Boudegga et al.	[76]	2021	RV-Net	UNet, AlexNet[77], VGG	Segmentation	Preprocessing and RV-Net for vessel segmentation
Wang et al.	[80]	2015	Features and ensemble learning	CNN, and RFs	Segmentation and classification	CNN and RFs for segmentation and classification of vessels.
Hu et al.	[82]	2018	Multiscale CNN	Richer convolutional features [83]	Segmentation	Multiscale CNN with an improved cross-entropy loss function
Zhou et al.	[84]	2021	Equilibrium GAN	UNet, GAN [85]	Segmentation	Uses the MSFRB and AM for segmentation of vessel
Yang et al.	[87]	2021	Improved UNet	UNet	Segmentation	Multitask and fusion block for vessel segmentation
Jin et al.	[88]	2019	DUNet	UNet, Deformable convNet [89]	Segmentation	Replacement of Convolution block in the UNETnetwork with the convoffset

							block for vessel segmentation
Soomro et al.	[90]	2019	Strided FCNN	SegNet [92]	Segmentation		Morphological tactics and PCA with the Encoder-decoder method for vessel segmentation.
Chala et al.	[94]	2021	Improved deep CNN	DCNN	Segmentation		Morphological tactics and PCA with the Encoder-decoder method for vessel segmentation.
Sun et al.	[95]	2021	CIEU-Net	UNet	Segmentation		UNET with integrated residual dilated module and the pyramid module
Wu et al.	[96]	2021	SCS-Net	UNet	Segmentation		Combination of SFA, AFF, and MSS for vessel segmentation
Sathananthavathi & Indumathi	[97]	2021	EEA UNet	UNet	Segmentation		Modified UNET architecture with atrous convolution for segmentation of vessels.
Yin et al.	[99]	2021	DF-Net	UNet	Segmentation		Pyramid U-shaped fusion network for segmentation of vessel.
Tang et al.	[101]	2020	MSCS	UNet	Segmentation		Multi-scale, channel importance, and spatial attention for segmentation of vessels.
Rahman et al.	[102]	2021	Cascaded AReN-UNet	UNet	Segmentation		Concatenated attention and residual modules for vessel segmentation.
Shi et al.	[103]	2021	MD-Net	SegNet, PSPNet [104], UNet	Segmentation		Res-ASPP and SE blocks sandwiched to segment RF vessels.
Tchinda et al.	[106]	2021	Classical edge detection and neural network	Artificial neural network	Segmentation		Feature vectors combined with a cascaded feed-forward network.
Gegundez-Arias et al.	[111]	2021	Simplified UNet	UNet	Segmentation		Combination of residual and batch normalization

						for vessel segmentation.
Maji & Sekh	[113]	2020	Tradition method with CNN	CNN	Classification	Combine the attention-based network with the transfer learning method for vessel segmentation.
Sangeethaa & Maheswari	[114]	2018	Trained CNN	CNN	Segmentation and Classification	Combined thresholding, morphological operation, edge detection, and adaptive histogram for segmentation then CNN for classification.
Muthusamy & Tholkapiyan	[123]	2019	CNN-RNN	CNN	Segmentation and Classification	Feature extraction and CNN-RNN classification.

Table 6: Pros and cons of CNN used for Retinal vessel

No of the Databases used	Ref	Advantages	Disadvantages	Accuracies Accuracy	Accuracies (AUC)
2	[36]	Data handling is performed excellently, this may improve accuracies. The network is very deep, hence providing the ability to handle complex representation.	Produce loss of information with more parameters, hence the network consumes time. Can not perform the translational invariant procedure.	0.97	0.98
4	[38]	Effectively build with residual blocks to avoid vanishing gradient. Create lower gradients, bias, and understand the data effectively	Easy to produce misleading output Computationally expensive	0.98	-
2	[40]	Fast training time when compared to most deep networks. Produce good and improved mean average precision which may result in good accuracy.	When testing the process becomes slow. Training is expensive for space and time.	0.78	0.70
1	[44]	Provides better representation of vessel than some other CNN. Global and local contexts are allowed at the same time, this may help increase accuracy.	Learning becomes slower at the middle layer which results in slow process time.	-	-
3	[47]	Uses the hyperdense connection to avoid gradient descent vanishing problem Uses the attention module to select important features to enhance segmentation accuracy.	More weights are added which can reduce training time. May become cumbersome and confusing if not properly managed.	0.98	0.98
5	[50]	Provide selected focus on the segment of sequence which may improve accuracy. Since feature dimensions are changed deeply, they can give insight into features.	Provides more regularization to the network and interference with the parameter.	0.96	0.97
3	[52]	Deep supervision in the BTS is used to avoid the gradient vanishing problem. Learns easily without human supervision.	Requires large data to perform optimally. Very expensive to train due to the complex data model.	0.82	0.97
2	[53]	Irrespective of data, the network will still produce results. A deeper network may produce better accuracy (The deeper the better)	May produce overfitting which reduces accuracy. Consumes huge processing time.	0.97	0.97
4	[55]	Produce an approximation of the sequence that would have been obtained, hence creating a faster training procedure	Cumbersome and consumes huge processing time.	0.99	0.99
2	[57]	Reduces the loss of small vessels using the size-invariant feature maps. The dense connectivity reduces the computational cost.	The large gaps in the loss and accuracies may reduce accuracies.	0.96	0.98
2	[60]	Uses heuristic strategy to increase information, hence improving the feedback to learn a task. Transform the hard binary labels into soft labels which give the likelihood of locations targeted by pixels.	If not well planned, the middle of the training procedure could be problematic.	0.74	-

6	[61]	Prevents over-amplification of noise and gives good contrast enhancement.	Cumbersome, and involve many feature extraction procedures that good produce error.	0.97	0.99
3	[64]	Reduce overfitting by using the attention mechanism and regularization method.	Large data is needed for optimal performance. The framework may become slow in the middle of the training which may affect the overall training.	0.95	0.98
4	[65]	Use the multi-path scale module to avoid the problem of determination of components to perform a certain task, hence producing better accuracies. Easy to start a new task.	Training and testing time is very slow	0.98	0.98
2	[68]	With the usage of two CNN, it is easy to produce good accuracies. Fast training and quick implementation time	May suffer from gradient vanishing problem.	0.96	0.96
3	[69]	Avoiding dense layers may make the network faster. Can be deployed on a system without larger memory	Maybe prone to overfitting because of the complexity and model parameter	0.97	0.98
2	[71]	Errors that occur during the training process are corrected, hence accuracies may be improved. May be used to solve other complex problems.	If not properly constructed may lead to state overload that can diminish results. May not be suitable for simple problems.	0.92	0.96
3	[72]	Uses augmentation and patch techniques to create more sample data, hence increasing the training data. Reduces noise in the local areas.	Can not perform optimally without lots of data. Cumbersome and may be difficult to understand.	0.96	0.98
2	[74]	Provide good image quality that may improve accuracy	The network may become slow at the middle layer, this will affect overall training.	0.94	-
2	[76]	Produce a fast timing for training and testing	Prone to vanishing gradient that affects accuracy.	0.98	-
2	[80]	Simple and can work efficiently on a large dataset. Robust and can perform effectively even in the presence of noise	When compared to a more sophisticated network, it may perform poorly.	0.98	0.97
2	[82]	The deeper the better, this network may produce accurate segmentation. Because of the augmentation method, that becomes sufficient, hence producing better training.	The network is too cumbersome and computationally expensive.	0.96	0.97
4	[84]	Does not require labeled data, hence training becomes faster.	The training process is hard and can become complex if not properly arranged.	0.96	0.98

		May not require preprocessing because the data are shaper. Data interpretation becomes easy, hence accuracy is improved.			
3	[87]	Does not need multiple runs of training to produce good accuracies. Can learn and perform with a few images.	Overall slow training in the middle of the system.	0.96	-
6	[88]	Perform well in a large and unstructured dataset. Can be computationally efficient.	Very slow for both testing and training. Require the use of a large dataset to function optimally.	0.96	0.98
4	[90]	Improves boundary delineation which in turn can produce good accuracies. Reduce the number of parameters fed into the network	Requires large computational cost and large data.	0.98	0.98
2	[94]	Produce very good accuracies and can reduce vanishing gradient problem	Cumbersome and heavily dependent on large datasets. Prone to noise, since there was no preprocessing in the network	0.98	-
3	[95]	Can easily localize with other networks and still produce a good result. Training patches are larger than the training images.	Very slow in the middle of the network, this can affect the overall network.	0.97	0.98
5	[96]	Capable of dealing with large-scale variation and extracting representative features. Extract varying sizes and shapes of the vessel.	Very slow in the middle of the network, this can affect the overall network.	0.97	0.98
4	[97]	Is not prone to the choice of network depth but modifies the UNET architecture for optimal results.	Very slow in the middle of the network, this can affect the overall network.	0.95	-
3	[99]	Easy to train and gives good inference time for images. Can detect vessels of different sizes.	Heavily dependent on the very large quality of data	0.98	0.98
2	[101]	The channel and multi-scale are used to prevent the network from overfitting. The spatial attention mechanism characterizes the encoder to generate coefficients and improve accuracies.	The network can not retrain a longer sequence of data. More weights are added to the data which can increase the training time.	0.96	0.98
2	[102]	Concatenated networks help to improve the model spatial and structural representation and generalization.	This network is computationally expensive and depends on large data.	0.97	0.98
3	[103]	The data augmentation and patch segmentation were used to avoid overfitting	computationally expensive and depends on large data.	0.97	-
3	[106]	The method is robust to noise and can	Does not have specific rules for	0.95	0.96

		perform optimally despite errors. It uses fast evaluation to learn and produce accuracies	the structure which may make rule assignment difficult and time-consuming.		
3	[111]	The network reduces the covariant shift and the dependence of gradient on the scale of the parameters. The model is regularized and reduces the need for dropout and other regularization techniques.	Computationally expensive and require excessive care for segmentation tasks. Regular changing of the construct can result in a poor generalized scale and shift of input data.	0.97	0.99
2	[113]	Has a good computational efficiency reduction mechanism. May produce faster training time due to convolution replacement.	The system may be hard to parallelize due to the usage of the attention mechanism.	-	-
3	[114]	Automatically detect important features without human supervision.	Requires large dataset for effective result.	0.96	-
3	[123]	The model can collect records so that each pattern can assume dependent status. Extend the pixel neighbor with CNN	Prone to exploding and vanishing gradient problem. Training is very difficult.	-	-

3 Performance Measures

The following are evaluation metrics for the performance of different CNN methods used for the segmentation and classification of retina fundus images.

Tp = true positive, Fp= false positive, Fn= false negative, Tn=true negative.

(1) Intersection over union (IOU): this method quantifies the overlap between the ground truth (GT) and the segmented prediction. The range of prediction is from 0 to 1. The IOU is also known as the Jaccard index.

$$IOU = \frac{GT \cap Prediction}{GT \cup Prediction} (1)$$

(2) Recall: this describes the appropriate part of the positive segmentation prediction relative to the ground truth. This method is the same as sensitivity. It tells the proportion of positive cases that were classified.

$$Recall = \frac{Tp}{Tp + Fn} (2)$$

(3) Precision: this metric describe the positive predictions relative to the ground truth.

$$Precision = \frac{Tp}{Tp + Fp} (3)$$

(4)Accuracy (ACC): measures the addition of true positive and true negative divided by all the positives and negatives

$$ACC = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} (4)$$

(5) Specificity: describe the proportion of the negative cases that were correctly classified.

$$Specificity = \frac{Tn}{Tn + Fp} (5)$$

(6) Area under the curve (AUC): describe the ability of the classifier to differentiate classes. A higher value of AUC means better performance.

$$AUC = \int_a^b F(x). dx(6)$$

Where a and b are upper and lower limits on the curve, and F(x) is the equation of the curve.

(7) Dice measure (DM): measures the overlap between sets of predicted segmentations with the ground truth. The higher values represent good predictions.

$$DM = 2 * \frac{|prediction \cap GT|}{|Prediction| + |GT|}(7)$$

(8) F1 scores: measures the function of precision and recall.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}(8)$$

(9) Mathew correction coefficient (MCC): measures the number of true and false positives and negatives.

$$MCC = \frac{Tp * Tn - Fp * Fn}{\sqrt{(Tp + Fp)(Tp - Fn)(Tn + Fp)(Tn - Fn)}}(9)$$

4 Brief Explanation Of Different Layers For Classification And Segmentation Of Retinal Fundus Images Using Cnn

(1) Activation function

(a) Rectified linear activation function (ReLU): is an activation function that returns 0 when it receives a negative input and returns positive values when it receives a positive number.

$$F(m) = \max(0, m)(9)$$

In most CNN applications it is used as a result.

(b) Softmax: is a mathematical function that converts vectors from numbers to probabilities. The converted probabilities are proportional to the related scale of each value in the vector. The softmax is configured to output values and normalize output.

$$\sigma(w \rightarrow) = \frac{e^{w_n}}{\sum_{i=1}^p e^{w_i}}(10)$$

(c) Hyperbolic tangent activation (Tanh): this activates the hyperbolic tangent in the neurons and ranges from -1 to 1 with an S-shaped configuration. This network is used mainly for classification between two classes in feed-forward networks.

$$F(m) = \tanh(m) = 2\sigma(2m) - 1(11)$$

(d) Sigmoid: a sigmoid function also referred to as squashing function is a special set of all real numbers in the range of 0 and 1 and used to add non-linearity in a network.

$$\sigma(m) = \frac{1}{(1 + e^{-(m)})}(12)$$

(2) Convolution layer: uses a filter to multiply sets of weights with the input. When repeated it produces a map activation which gives a feature map. This layer is the building block in neural networks and CNN.

(3) Dropout layer: used to avoid overfitting in a network by dropping out units. It performs random shifting of inputs with frequency rate at each procedure during training time. The dropout work by removing nodes with a probability of 1 - m and keeping nodes with

probability on m . The reason why dropout is adopted in any network is to reduce the incoming and outgoing edges reducing the network size.

(4) Batch normalization: is a technique that trains a deep neural network and standardizes the inputs to a layer for each mini-batches.

(5) Upsampling: this is a layer that does not have weights but it doubles the dimensions of input and generates a model when it is followed by a convolution layer. The two common types of upsampling are UpSampling 2D, which doubles dimensions, and Conv2DTranspose, an inverse convolution operation.

(6) Downsampling: this layer performs a max operation on inputs or groups of output depending on the position of the layer. Downsampling is used to reduce the feature dimensions resulting in less usage of computational space. This layer is sometimes referred to as the pooling layer.

5 Datasets

In this section, the lists of datasets used for segmenting and classifying retina fundus images are discussed. This datasets are listed in order of usage, they include: DRIVE [125], STARE [126], CHASE-DB1 [127], HRF [128], MESSIDOR [129], IOSTAR [130], ORIGA [131], REFUGE [131], DB1 [132], DB0 [133], Kaggle dataset, DRISHTI-GS [134], NIVE [135], RIM-ONEr3 [136], DRIONS-DB [137], RITE [138], WIDE [139], SYNTHETIC [140], LES-AV [141], RIGA [142], DUKE, DCA, EIARG1, and AV-INSPIRE. The list of datasets that have the highest usage is shown in Fig. 14. Datasets with usage less than 2 were not recorded on the graph. The datasets with the highest usage are DRIVE, STARE, CHASE-DB1, and HRF. Findings reveal that a good dataset should be relevant, usable, and have good quality [143]. Hence, we believe that datasets with higher usage have these aforementioned points.

6 Analyses

Authors searched all online repositories to get papers relating to the classification and segmentation of retina fundus images using CNN. Initially, 200 papers were collated, thereafter the papers were screened, to 100. Finally, the authors decided to consider papers relating to arteries, vessels, and optic disc/ cup (see Fig. 10). A total of 52 research papers from the classification and segmentation of CNN for veins, optic disc, and blood vessels were included in this review (see Fig. 11). The authors made effort to include all research articles relating to the topic scope, however, some interesting studies might have been skipped. Analysis relating to data sources and year of publication is available in Figs. 11 and 12.

The studies included in this review comprised of 2 papers from arteries and veins, 9 papers from optic disc and cup, and 41 papers from retina vessels. A total of 4 studies combined the CNN method with traditional methods (see Tables 6, and 2). Reference [113] was the only study that focused on classification alone, while references [114], [123], and [80] proposed segmentation and classification for RF images. Figure 13 is the analysis of the accuracy metric for optic disc and cup, however, other accuracies can be viewed from the tables.

7 Conclusions And Future Directions

The objective of this study is to review methods relating to CNN segmentation and classification for retina fundus images. The performance of different CNN algorithms for fundus images has been impressive with some CNN methods producing approximately 99% accuracies in some cases. In terms of medical practice, some CNN algorithms have provided helpful clues that clinicians have used to make conclusions on a particular pathology. Several CNN methods have helped to diagnose disease and to predict the existence of the ophthalmic disease in patients. The overall performance of CNN algorithms has made it a promising area and a better replacement for so many traditional methods. For example, the intact dilated and extraction (IDX) has already been approved in the US for practical use in medium and large hospitals, physician groups, and medical management firms [145, 146]. In the future, improvements in CNN algorithms will propel its usage in other critical sectors of health care delivery. We believe that once most CNN algorithms become stable and more efficient their relevance will become widely accepted.

7.1 Observations

1) From the reviewed papers, we observed that most CNN methods for segmenting RF images were built on the UNet, attention, and adversarial networks. A major reason for this could be the performance of these networks (especially UNet) for medical image tasks. Again since segmenting RF images is more than just using a backbone network and getting results, but building a network that solves the task in an accurate and enabling manner, the idea to build a CNN network for segmenting RF on UNet is the right decision.

2) The use of preprocessing methods has helped to improve accuracies. Most researchers have used CLAHE for enhancement and different filters for image denoising. Although we believe that the use of different image denoising methods has not improved results significantly, however, the enhancement with CLAHE has improved the performance of the images.

3) Images used in this review were converted from one color space to another. Most research paper converts the color RF images from RGB format to LAB, then after preprocessing these images were converted back to RGB images. The reason for such conversion was to provide the enabling environment for preprocessing.

4) It is common knowledge that medical images are not available for analysis, hence for effective accuracies, most CNN methods require a large dataset. So far, most researchers used the data augmentation technique as a preprocessing step to acquire more data for the network.

7.2 Limitations and future directions

Despite the enormous contributions of CNN to fundus images segmentation and classification there are still some challenges that confront the application of CNN to RF images.

1. **Unavailability of labeled quality data:** the better the image the better the results, this is a popular slogan commonly used in CNN applications. Although there are several datasets available for fundus images (see section 5), the labeling of enough data in the dataset remains much to desire. This is because experts have a busy schedule and do not have time to accomplish data labeling. For example: to accomplish good data labeling for breast ultrasound images, it is expected that 3 experts do the labeling and come up with a majority voting to arrive at the final image. This task is quite difficult and demanding, requiring great effort and time. To resolve this problem, researchers can use the methods adopted by [147] where a combination of different algorithms with human supervision was used to create labeled data. Another solution will be to use weakly supervised methods to mop-up labeled data for use in the unlabeled data.
2. **Performance of CNN methods in a real-life situation:** most CNN methods reviewed have not been tested in real-life situations. It remains uncertain if some of the algorithms will perform optimally when applied to real-life applications. To solve this problem analysts and computer vision engineers are advised to engage closely with clinicians for a handshake on the gradual deployment of the CNN algorithms on real-life applications.
3. **Unavailability of standard measuring metrics:** It is obvious that research used different evaluation metrics for quantitative results reporting. These metrics measure different evaluations, hence authors could use any metric from the list of available metrics to quantify their research. Although this review solves this problem by collating two common accuracies from the studies, however, more still can be done to reduce the proliferation of metrics. We advocate a procedure to select 3 or 4 metrics as a standard to be used in the segmentation and classification of RF images. This procedure will give uniformity in the metric evaluation procedure.
4. **CNN high computation space:** It is common knowledge that CNN algorithms require a large computational space. This space if not available could jeopardize the chances of attaining higher accuracies in the segmentation task. To solve this problem, we request the creation of more cloud computing platforms (such as Collab) with enough computational space to help researcher run their applications with little or no cost.

Declarations

Acknowledgments

Special thanks to the management of Medical Image Processing Group, Department of Radiology, University of Pennsylvania, and the anonymous referees of the review for valuable remarks.

Author contributions

All authors contributed to the study conception and design. Material preparation and data collection were performed by AEI, TI, and OPI. Analysis was done by AEI. The first draft of the manuscript was written and polished by AEI.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

1. Whitcher JP, Srinivasan M, Upadhyay MP. Corneal blindness: a global perspective. *Bull World Health Organ* 2001;79:214–21.
2. S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, V. Lakshminarayanan, Ophthalmic diagnosis using deep learning with fundus images – A critical review, *102(2020)*, 101758.
3. Nicolela MT, Vianna JR. Optic nerve: clinical examination. *Pearls of Glaucoma Management*. Berlin, Heidelberg: Springer, 2016. p. 17–26.
4. Aziz, T., Ilesanmi, A.E., Charoenlarnopparut, C. Efficient and Accurate Hemorrhages Detection in Retinal Fundus Images Using Smart Window Features. *Appl. Sci.* 2021, 11, 6391. <https://doi.org/10.3390/app11146391>
5. Krolewski AS, Warram JH, Rand LI, Christlieb AR, Busick EJ, Kahn CR. Risk of proliferative diabetic retinopathy in juvenile-onset type 1 diabetes: a 40-yr followup study. *Diabetes Care* 1986;9(5):443–52.
6. M.J. Fowler, Microvascular and macrovascular complications of diabetes, *Clin. Diabetes* 26 (2) (2008) 77–82 .
7. Kauppi, T., Kalesnykiene, V., Kamarainen, J.K., Lensu, L., Sorri, I., Uusitalo, H., Kälviäinen, H., Pietilä, J. DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms, Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology: Lappeenranta, Finland, 2006, Volume 73, pp. 1–17.
8. Age-Related Eye Disease Study Research Group. The relationship of dietary carotenoid with vitamin A, E, and C intake with age-related macular degeneration in a case-control study. *Archives of Ophthalmology* 2007, 125(9): 1225–1232.
9. Age-Related Eye Disease Study Research Group. Risk factors associated with age-related nuclear and cortical cataracts. *Ophthalmology* 2001, 108(8): 1400–1408.
10. U.S. Department of Health and Human Services, Office of the Surgeon General. *The Health Consequences of Smoking: A Report of the Surgeon General* (Washington, D.C., 2004).
11. E .T .D .R .S .R. Group , Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10, *Ophthalmology* 98 (5) (1991) 786–806 .
12. C.J. Rudnisky , B.J. Hinz , M.T. Tennant , A.R. de Leon , M.D. Greve, High-resolution stereoscopic digital fundus photography versus contact lens biomicroscopy for the detection of clinically significant macular edema, *Ophthalmology* 109 (2) (2002) 267–274.
13. Md Mohaimenul Islam, Hsuan-Chia Yang, Tahmina Nasrin Poly, Wen-Shan Jian, Yu-Chuan (Jack) Li, Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis, *Computer Methods and Programs in Biomedicine*, 191(2020), 105320.
14. <https://www.invision2020.com/retinal-imaging-history/> (Accessed: October 10, 2021)
15. Ademola Enitan Ilesanmi, Oluwagbenga Paul Idowu, Stanislav S. Makhanov, Multiscale superpixel method for segmentation of breast ultrasound, *Computers in Biology and Medicine*, 125(2020), 103879.
16. Yang Chen, Guirong Weng, An active contour model based on local pre-piecewise fitting image, *Optik*, 2021, 168130.
17. Daniel Lacker, Kavita Ramanan, Ruoyu Wu, Locally interacting diffusions as Markov random fields on path space, *Stochastic Processes and their Applications*, 140(2021), Pages 81-114,
18. Lei Wang, Juan Gu , Yize Chen, Yuanbo Liang, Weijie Zhang, Jiantao Pu, Hao Chen, Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network, *Pattern Recognition* 112 (2021) 107810.
19. H. Fu , J. Cheng , Y. Xu , D. Wong , J. Liu , X. Cao , Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation, *IEEE Trans Med Imaging* 37 (7) (2018) 1597–1605.
20. O. Ronneberger , P. Fischer, T. Brox, U-Net: convolutional Networks for Biomedical Image Segmentation, *Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015.
21. Yinghua Fu, Jie Chen, Jiang Li, Dongyan Pan, Xuezheng Yue, Yiming Zhu, Optic disc segmentation by U-net and probability bubble in abnormal fundus images, *Pattern Recognition* 117 (2021) 107971.

22. H. Narasimhaiyer, A. Can, B. Roysam , V. Stewart , H.L. Tanenbaum, A. Ma- jerovics, H. Singh, Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy, *IEEE Trans. Biomed. Eng.* 53 (6) (2006) 1084–1098.
23. Xianjing Meng, Xiaoming Xi, Lu Yang, Guang Zhang, Yilong Yin, Xinjian Chen, Fast and effective optic disk localization based on convolutional neural network, *Neurocomputing* 312 (2018) 285–295.
24. Y. LeCun , L. Bottou , Y. Bengio , P. Haffner , Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
25. Xin Yuan, Lingxiao Zhou, Shuyang Yu, Miao Li, Xiang Wang, Xiujuan Zheng, A multi-scale convolutional neural network with context for joint segmentation of optic disc and cup. *Artificial Intelligence In Medicine* 113 (2021) 102035.
26. Ademola Enitan Ilesanmi, Utairat Chaumrattanakul, Stanislav S. Makhanov, A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning, *Biocybernetics and Biomedical Engineering*, 41(2),(2021),802-818.
27. Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans Med Imaging* 2018,37(7):1597–605.
28. Lei Wang, Han Liu, Yaling Lu, Hang Chen, Jian Zhang, Jiantao Pu, A coarse-to-fine deep learning framework for optic disc segmentation in fundus images, *Biomedical Signal Processing and Control* 51 (2019) 82–89.
29. Jen Hong Tan, U. Rajendra Acharya, Sulatha V. Bhandary, Kuang Chua Chua, Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science* 20 (2017) 70–79.
30. H.N. Veena, A. Muruganandham, T. Senthil Kumaran. A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images, *Journal of King Saud University – Computer and Information Sciences* xxx (xxxx) xxx.
31. Lei Zhang, Lang Zou, Chuanyu Wu, Jiangming Jia, Jianneng Chen, Method of famous tea sprout identification and segmentation based on improved watershed algorithm, *Computers and Electronics in Agriculture*, 184(2021), 106108.
32. Neha Mathur, Shruti Mathur, Divya Mathur, A Novel Approach to Improve Sobel Edge Detector, *Procedia Computer Science*, 93(2016), Pages 431-438.
33. Rakhshanda Imtiaz, Tariq M. Khan, Syed Saud Naqvi, Muhammad Arsalan, Syed Junaid Nawaz, Screening of Glaucoma disease from retinal vessel images using semantic segmentation, *Computers and Electrical Engineering* 91 (2021) 107036.
34. Fantin Girarda, Conrad Kavalec, Farida Cheriet, Joint segmentation and classification of retinal arteries/veins from fundus images, *Artificial Intelligence In Medicine* 94 (2019) 96–109.
35. Jos´e Morano, ´Alvaro S. Hervella, Jorge Novo, Jos´e Rouco, Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images, *Artificial Intelligence In Medicine* 118 (2021) 102116.
36. Girard F, Kavalec C, Cheriet F. Joint segmentation and classification of retinal arteries/veins from fundus images. *Artif Intell Med* 2019,94:96–109. <https://doi.org/10.1016/j.artmed.2019.02.004>.
37. Ümit Budak, Zafer Cömert, Musa Çıbuk, Abdulkadir Şengür, DCCMED-Net: Densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images, *Medical Hypotheses* 134 (2020) 109426.
38. Peng Tanga, Qiaokang Lianga, Xintong Yana, Dan Zhang, Gianmarc Coppola, Wei Sun, Multi-proportion channel ensemble model for retinal vessel segmentation, *Computers in Biology and Medicine* 111 (2019) 103352.
39. J. Odstrcilik, et al., Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database, *IET Image Process.* 7 (4) (2013) 373–383.
40. He Zhao, Yun Sun, Huiqi Li, Retinal vascular junction detection and classification via deep neural networks, *Computer Methods and Programs in Biomedicine* 183 (2020) 105096.
41. K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: *Proceedings of the IEEE International Conference on Computer Vision, 2017*, pp. 2961–2969.
42. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016*, pp. 770–778.
43. T.-Y. Lin , P. Dollar , R. Girshick , K. He , B. Hariharan , S. Belongie , Feature pyramid networks for object detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017*, pp. 2117–2125.
44. Maciej Czepita, Anna Fabija ´nska, Image processing pipeline for the detection of blood flow through retinal vessels with subpixel accuracy in fundus images, *Computer Methods and Programs in Biomedicine* 208 (2021) 106240.

45. S. Klein, Staring, M.K. Murphy, M.A. Viergever, J.P.W. Pluim, Elastix: a toolbox for intensity based medical image registration, *IEEE Trans. Med. Imaging* 29 (1) (2010) 196–205, doi: 10.1109/TMI.2009.2035616.
46. L.D. Hubbard, R.J. Brothers, W.N. King, et al., Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study, *Ophthalmology* 106 (12) (1999) 2269–2280 Dec, doi: 10.1016/s0161-6420(99)90525-0.
47. Gang Sun, Xiaoyan Liu, Xuefei Yu, Multi-path cascaded U-net for vessel segmentation from fundus fluorescein angiography sequential images, *Computer Methods and Programs in Biomedicine* 211 (2021) 106422.
48. Michael Yeung, Evis Sala, Carola-Bibiane Schonlieb, Leonardo Rundo, Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy, *Computers in Biology and Medicine* 137 (2021) 104815.
49. Lei Sang, Min Xu, Shengsheng Qian, Xindong Wu, Knowledge Graph enhanced Neural Collaborative Filtering with Residual Recurrent Network, *Neurocomputing* 454 (2021) 417–429.
50. Ruohan Zhao, Qin Li, Jianrong Wu, Jane You, A nested U-shape network with multi-scale upsample attention for robust retinal vascular segmentation, *Pattern Recognition* 120 (2021) 107998.
51. J.V. Soares, J.J. Leandro, R.M. Cesar, H.F. Jelinek, M.J. Cree, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification, *IEEE Trans. Med. Imaging* 25 (9) (2006) 1214–1222.
52. Song Guo, Kai Wang, Hong Kang, Yujun Zhang, Yingqi Gao, Tao Li, BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation, *International Journal of Medical Informatics* 126 (2019) 105–113.
53. Yanhui Guo, Ümit Budak, Abdulkadir S engür, A novel retinal vessel detection approach based on multiple deep convolution neural networks, *Computer Methods and Programs in Biomedicine* 167 (2018) 43–48.
54. P. Liskowski, K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Trans. Med. Imag.* 35 (2016) 2369–2380, doi: 10.1109/TMI.2016.2546227.
55. Kyoung Jin Noh, Sang Jun Park, Soochahn Lee, Scale-space approximated convolutional neural networks for retinal vessel segmentation, *Computer Methods and Programs in Biomedicine* 178 (2019) 237–246.
56. K. Maninis, J. Pont-Tuset, P.A. Arbeláez, L.V. Gool, Deep retinal image understanding, in: *Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2016, pp. 140–148, doi: 10.1007/978-3-319-46723-8_17.
57. Zhongshuo Zhuo, Jianping Huang, Ke Lu, Daru Pan, A size-invariant convolutional network with dense connectivity applied to retinal vessel segmentation measured by a unique index, *Computer Methods and Programs in Biomedicine* 196 (2020) 105508.
58. Y. Luo, H. Cheng, L. Yang, Size-invariant fully convolutional neural network for vessel segmentation of digital retinal images, in: *Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016 Asia-Pacific, IEEE, 2016, pp. 1–7.
59. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *CVPR*, vol. 1, 2017, p. 3.
60. Álvaro S. Hervella, JoséRouco, Jorge Novo, ManuelG. Penedo, Marcos Ortega, Deep multi-instance heatmap regression for the detection of retinal vessel crossings and bifurcations in eye fundus images, *Computer Methods and Programs in Biomedicine* 186 (2020) 105201.
61. Pearl Mary Samuel, Thanikaiselvan Veeramalai, VSSC Net: Vessel Specific Skip chain Convolutional Network for blood vessel segmentation, *Computer Methods and Programs in Biomedicine* 198 (2021) 105769.
62. S. Pearl Mary, V. Thanikaiselvan, Unified adaptive framework for contrast enhancement of blood vessels, *Int. J. Electr. Comput. Eng.* 10 (2020) 767–777, doi: 10.11591/ijece.v10i1.pp767-777.
63. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.
64. Di Li, Susanto Rahardja, BSEResU-Net: An attention-based before-activation residual U-Net for retinal vessel segmentation, *Computer Methods and Programs in Biomedicine* 205 (2021) 106070.
65. Zefang Lin, Jianping Huang, Yingyin Chen, Xiao Zhang, Wei Zhao, Yong Li, Ligong Lu, Meixiao Zhan, Xiaofei Jiang, Xiong Liang, A high resolution representation network with multi-path scale for retinal vessel segmentation, *Computer Methods and Programs in Biomedicine* 208 (2021) 106206.
66. K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation, *arXiv preprint arXiv:1902.09212* (2019).
67. C. Meng, K. Sun, S. Guan, Q. Wang, R. Zong, L. Liu, Multiscale dense convolutional neural network for DSA cerebrovascular segmentation, *Neurocomputing* 373 (2020) 123–134.

68. Chun Tian, Tao Fang, Yingle Fan, Wei Wu, Multi-path convolutional neural network in fundus segmentation of blood vessels, *biocybernetics and biomedical engineering* 40 (2020) 583 – 595.
69. Ibrahim Atli, Osman Serdar Gedik, Sine-Net: A fully convolutional deep learning architecture for retinal blood vessel segmentation, *Engineering Science and Technology, an International Journal* 24 (2021) 271–283.
70. X. Bai, F. Zhou, B. Xue, Image enhancement using multi-scale image features extracted by top-hat transform, *Opt. Laser Technol.* 44 (2) (2012) 328–336.
71. Yanhui Guo, Ümit Budak, Lucas J. Vespa, Elham Khorasani, Abdulkadir Şengür, A retinal vessel detection approach using convolution neural network with reinforcement sample learning strategy, *Measurement* 125 (2018) 586–591.
72. Yicheng Wu, Yong Xia, Yang Song, Yanning Zhang, Weidong Cai, NFN+: A novel network followed network for retinal vessel segmentation, *Neural Networks* 126 (2020) 153–162.
73. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. <http://dx.doi.org/10.1109/tpami.2017.2699184>.
74. Ruben Hemelings, Bart Elend, Ingeborg Stalmans, Karel Van Keer, Patrick De Boever, Matthew B. Blaschko, Artery–vein segmentation in fundus images using a fully convolutional network, *Computerized Medical Imaging and Graphics* 76 (2019) 101636.
75. Soares, J.V.B., Leandro, J.J.G., Cesar, R.M., Jelinek, H.F., Cree, M.J., 2006. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* 25, 1214–1222.
76. Henda Boudegga, Yaroub Elloumi, Mohamed Akil, Mohamed Hedi Bedoui, Rostom Kachouri, Asma Ben Abdallah, Fast and efficient retinal blood vessel segmentation method based on deep learning network, *Computerized Medical Imaging and Graphics* 90 (2021) 101902.
77. Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60 (May (6)), 84–90. <https://doi.org/10.1145/3065386>.
78. Jiang, Z., Zhang, H., Wang, Y., Ko, S.-B., 2018. Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Comput. Med. Imaging Graph.* 68 (September), 1–15. <https://doi.org/10.1016/j.compmedimag.2018.04.005>.
79. Liskowski, P., Krawiec, K., 2016. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* 35 (November (11)), 2369–2380. <https://doi.org/10.1109/TMI.2016.2546227>.
80. Shuangling Wang, Yilong Yin, Guibao Cao, Benzhen Wei, Yuanjie Zheng, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning, *Neurocomputing* 149 (2015) 708–717.
81. L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
82. Kai Hu, Zhenzhen Zhang, Xiaorui Niu, Yuan Zhang, Chunhong Cao, Fen Xiao, Xieping Gao, Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function, *Neurocomputing* 309 (2018) 179–191.
83. Y. Liu, M.-M. Cheng, X. Hu, K. Wang, X. Bai, Richer convolutional features for edge detection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5872–5881.
84. Yukun Zhou, Zailiang Chen, Hailan Shen, Xianxian Zheng, Rongchang Zhao, Xuanchu Duan. A refined equilibrium generative adversarial network for retinal vessel segmentation, *Neurocomputing* 437 (2021) 118–130.
85. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., “Generative adversarial nets,” in *Neural Information Processing Systems Conference*, Montreal, Quebec, Canada, 2014.
86. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al, “Attention is all you need,” in *International Conference on Learning Representations*, Long Beach, California, USA. 2017.
87. Lei Yang, Huaixin Wang, Qingshan Zeng, Yanhong Liu, Guibin Bian, A hybrid deep segmentation network for fundus vessels via deep-learning framework, *Neurocomputing* 448 (2021) 168–178.
88. Qiangguo Jin, Zhaopeng Meng, Tuan D. Pham, Qi Chen, Leyi Wei, Ran Su, DUNet: A deformable network for retinal vessel segmentation, *Knowledge-Based Systems* 178 (2019) 149–162.
89. J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: *International Conference on Computer Vision*, 2017, pp. 764–773.
90. Toufique Ahmed Soomro, Ahmed J. Afifib, Junbin Gao, Olaf Hellwich, Lihong Zheng, Manoranjan Paul, Strided fully convolutional neural network for boosting the sensitivity of retinal blood vessels segmentation. *Expert Systems With Applications* 134 (2019) 36–

91. Soomro, T. A. , Afifi, A. J. , Gao, J. , Hellwich, O. , Khan, M. U. , Paul, M. , & Zheng, L. (2017). Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network. In *International conference on digital image computing: Techniques and applications (DICTA)*.
92. Badrinarayanan , Vijay , Kendall , Alex , Cipolla , & Roberto (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99. 1–1.
93. Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. (pp. 1–5). arXiv: 1505.00853v2.
94. Mohamed Chala, Benayad Nsiri, My Hachem El yousfi Alaoui, Abdelmajid Soulaymani, Abdelrhani Mokhtari, Brahim Benaji. Chandler, An automatic retinal vessel segmentation approach based on Convolutional Neural Networks, *Expert Systems With Applications* 184 (2021) 115459.
95. Muyi Sun, Kaiqi Li, Xingqun Qi, Hao Dang, Guanhong Zhang, Contextual information enhanced convolutional neural networks for retinal vessel segmentation in color fundus images, *J. Vis. Commun. Image R.* 77 (2021) 103134.
96. Huisi Wu, Wei Wang, Jiafu Zhong, Baiying Lei, Zhenkun Wen, Jing Qin, SCS-Net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation, *Medical Image Analysis* 70 (2021) 102025.
97. Sathananthavathi V., Indumathi G., Encoder Enhanced Atrous (EEA) Unet architecture for Retinal Blood vessel segmentation, *Cognitive Systems Research* 67 (2021) 84–95.
98. Fisher, Y., & Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. In *Proceeding of Computer Vision and Pattern Recognition ICLR 2016*. arXiv:1511.07122v3 [cs.CV] 30 Apr 2016..
99. Pengshuai Yin, Hongmin Cai, Qingyao Wu, DF-Net: Deep fusion network for multi-source vessel segmentation, *Information Fusion*, 78(2022), 199-208.
100. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2881–2890.
101. Xianlun Tang, Bing Zhong, Jiangping Peng, Bohui Hao, Jie Li, Multi-scale channel importance sorting and spatial attention mechanism for retinal vessels segmentation, *Applied Soft Computing Journal* 93 (2020) 106353
102. Aamer Abdul Rahman, Birendra Biswal, Geetha Pavani P, Shazia Hasan, M.V.S. Sairam, Robust segmentation of vascular network using deeply cascaded AReN-UNet, *Biomedical Signal Processing and Control* 69 (2021) 102953.
103. Zhengjin Shi, Tianyu Wang, Zheng Huang, Feng Xie, Zihong Liu, Bolun Wang, Jing Xu, MD-Net: A multi-scale dense network for retinal vessel segmentation, *Biomedical Signal Processing and Control* 70 (2021) 102977.
104. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in *Proc IEEE Comput Soc Conf Comput Vision Pattern Recognit (2017)*, Venice Italy, pp. 2881- 2890. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/papers/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.pdf.
105. R. Collobert and S. Bengio (2004). Links between Perceptrons, MLPs and SVMs. *Proc. Int'l Conf. on Machine Learning (ICML)*.
106. Beaudelaire Saha Tchinda, Daniel Tchiotsop, Michel Noubom, Valerie Louis-Dorr, Didier Wolf, Retinal blood vessels segmentation using classical edge detection filters and the neural network, *Informatics in Medicine Unlocked* 23 (2021) 100521.
107. Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Trans Med Imag* 2006,25(9):1214–22.
108. Ziou D, Tabbone S. Edge detection techniques- an overview. *International Journal of Pattern Recognition and Image Analysis* 1998,8(4):537–59.
109. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 1986,8:679–714.
110. Gonzalez RC, Woods RE, Eddins SL. *Digital image processing using MATLAB*. Upper Saddle River, New Jersey: Pearson/Prentice Hall, 2004.
111. Manuel E. Gegundez-Arias, Diego Marin-Santos, Isaac Perez-Borrero, Manuel J. Vasallo-Vazquez, new deep learning method for blood vessel segmentation in retinal images based on convolutional kernels and modified U-Net model, *Computer Methods and Programs in Biomedicine* 205 (2021) 106081.
112. L Srinidhi, C., Aparna, P. & Rajan, J. Recent Advancements in Retinal Vessel Segmentation. *J Med Syst* 41, 70 (2017). <https://doi.org/10.1007/s10916-017-0719-2>.

113. Debasis Maji, Arif Ahmed Sekh, Automatic Grading of Retinal Blood Vessel in Deep Retinal Image Diagnosis, *Journal of Medical Systems* (2020) 44: 180.
114. S. N. Sangeetha & P. Uma Maheswari, An Intelligent Model for Blood Vessel Segmentation in Diagnosing DR Using CNN, *Journal of Medical Systems* (2018) 42: 175.
115. Zhe Xie, Tonghui Ling, Yuanyuan Yang, Rong Shu, Brent J. Liu, Optic Disc and Cup Image Segmentation Utilizing Contour-Based Transformation and Sequence Labeling Networks, *Journal of Medical Systems* (2020) 44: 96.
116. Yu F, Koltun V (2016) Multi-scale context aggregation by dilated convolutions. In: 4th international conference on learning representations, ICLR 2016 - conference track proceedings,
117. A. Viterbi, Optimum detection and signal selection for partially coherent binary communication *IEEE Trans. Inf. Theor.*, 11 (2) (Apr. 1965), pp. 239-246.
118. A.J. Viterbi, A personal history of the Viterbi algorithm, *IEEE Signal Process. Mag.*, 23 (4) (Jul. 2006), pp. 120-142.
119. Sandip Sadhukhan, Arpita Sarkar, Debprasad Sinha, Goutam Kumar Ghorai, Gautam Sarkar, Ashis K. Dhara, Attention Based Fully Convolutional Neural Network for Simultaneous Detection and Segmentation of Optic Disc in Retinal Fundus Images, *World Academy of Science, Engineering and Technology International Journal of Medical and Health Sciences*, Vol:14, No:8, 2020.
120. R. Priyanka, S. J. Grace Shoba, A. Brintha Therese, Segmentation of Optic Disc in Fundus Images using Convolutional Neural Networks for Detection of Glaucoma. *International Journal of Advanced Engineering Research and Science (IJAERS)*, Vol 4, Issue 5, 2017
121. C. Raja, N. Vinodhkumar, An Efficient Segmentation Of Optic Disc Using Convolution Neural Network For Glaucoma Detection In Retinal Images, *European Journal of Molecular & Clinical Medicine*, Volume 07, Issue 03, 2020.
122. F Hoppner, F. Klawonn, A contribution to convergence theory of fuzzy c-means and derivatives, *IEEE Trans. Fuzzy Syst.* 11 (5) (2003) 682–694.
123. Revathi Priya Muthusamy, S. Vinod, M. Tholkapiyan, Automatic Detection of Abnormalities in Retinal Blood Vessels using DTCWT, GLCM Feature Extractor and CNN-RNN Classifier, *International Journal of Recent Technology and Engineering (IJRTE)*, Volume 8, Issue 1S4, 2019.
124. Retinal Vessel Segmentation via A Coarse-to-fine Convolutional Neural Network, 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
125. J. Staal , M.D. Abràmoff, M. Niemeijer , M.A. Viergever , B.V. Ginneken , Ridge-based vessel segmentation in color images of the retina, *IEEE Trans. Med. Imaging* 23 (4) (2004) 501–509.
126. A. Hoover, M. Goldbaum, Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels, *IEEE Trans. Med. Imaging* 22 (8) (2003) 951–958.
127. Q. Li, B. Feng, L.P. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, *IEEE Trans. Med. Imaging* 35 (2016) 109–118.
128. J. Orlando, E. Prokofyeva, M. Blaschko, A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images, *IEEE Trans. Biomed. Eng.* 64 (2017) 16–27.
129. L. Wang , H. Liu , Y. Lu , H. Chen , J. Zhang , J. Pu , A coarse-to-fine deep learning framework for optic disc segmentation in fundus images, *Biomed Signal Process Control* 51 (2019) 82–89.
130. S. Abbasi-Sureshjani, I. Smit-Ockeloen, E. Bekkers, B. Dashtbozorg, B. ter Haar Romeny, Automatic detection of vascular bifurcations and crossings in retinal images using orientation scores, in: *International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016, pp. 189–192.
131. J. Orlando , H. Fu , J. Breda , K. Keer , D. Bathula , A. Pinto , R. Fang , P. Heng , J. Kim , J. Lee , J. Lee , X. Li , P. Liu , S. Lu , B. Murugesan , V. Naranjo , S. Phaye , S. Shankara-narayana , A. Sikka , J. Son , A. Hengel , S. Wang , J. Wu , Z. Wu , G. Xu , Y. Xu , P. Yin , F. Li , X. Zhang , Y. Xu , H. Bogunovic , REFUGE Challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs, *Med Image Anal* 59 (2020) 101570.
132. R.V.J.P.H. Kälviäinen, H. Uusitalo, Diaretdb1 diabetic retinopathy database and evaluation protocol, *Med. Image Underst. Anal.* 20 07 (20 07) 61.
133. T. Kauppi, V. Kalesnykiene, J.K. Kamarainen, L. Lensu, I. Sorri, H. Uusitalo, H. Kälviäinen, J. Pietilä, Diaretdb0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms, *Machine Vision and Pattern Recognition Research Group*, 73, Lappeenranta University of Technology, Finland, 2006.

134. Sivaswamy J, Krishnadas S, Chakravarty A, Joshi G, Tabish AS. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomed Imaging Data Papers* 2015,2(1):1004.
135. R. Wang , L. Zheng , C. Xiong , C. Qiu , H. Li , X. Hou , B. Sheng , P. Li , Q. Wu , Retinal optic disc localization using convergence tracking of blood vessels, *Multimedia Tools & Applications* 76 (22) (2016) 1–23.
136. Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. Rim-one: An open retinal image database for optic nerve evaluation. 2011 24th international symposium on computer-based medical systems (CBMS) 2011:1–6.
137. M. Abdullah, M. Fraz, S. Barman, Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm, *PeerJ* (2016).
138. Kang H, Gao Y, Guo S, Xu X, Li T, Wang K. AVNet: A retinal artery/vein classification network with category-attention weighted fusion. *Comput Methods Prog Biomed* 2020,195:105629. <https://doi.org/10.1016/j.cmpb.2020.105629>.
139. R. Estrada, C. Tomasi, S.C. Schmidler, S. Farsiu, Tree topology estimation, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (8) (2015) 1688–1701.
140. H. Zhao, H. Li, S. Maurerstroh, L. Cheng, Synthesizing retinal and neuronal images with generative adversarial nets, *Med. Image Anal.* 49 (2018) 14–26.
141. Orlando, J.I. , Breda, J.B. , Van Keer, K. , Blaschko, M.B. , Blanco, P.J. , Bulant, C.A. ,2018. Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 65–73.
142. Almazroa A. A, Alodhayb S, Osman E, Ramadan E, Hummadi M, Dlaim M, Alkatee M, Raahemifar K, Lakshminarayanan V (2018) Retinal fundus images for glaucoma analysis: The RIGA dataset.
143. L.M. Koesten, E. Kacprzak, J.F.A. Tennison, E. Simperl, The trials and tribulations of working with structured data: -a study on information seeking behavior, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA (2017), pp. 1277-1289, 10.1145/3025453.3025838
144. Tao Li, Wang Bo, Chunyu Hu, Hong Kang, Hanruo Liu, Kai Wang, Huazhu Fu, Applications of deep learning in fundus images: A review, *Medical Image Analysis* 69 (2021) 101971.
145. Sourya Sengupta, Amitojdeep Singh, Henry A. Leopold, Tanmay Gulati, Vasudevan Lakshminarayanan, Ophthalmic diagnosis using deep learning with fundus images – A critical review, *Artificial Intelligence In Medicine* 102 (2020) 101758.
146. David, A.M. , Lou, Y. , Ali, E. , Warren, C. , Ryan, A. , Folk, J.C. , Meindert, N. , 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology & Visual Science* 57 (13), 5200–
147. Jayaram K. Udupa, Dewey Odhner, Liming Zhao, Yubing Tong, Monica M.S. Matsumoto, Krzysztof C. Ciesielski, Alexandre X. Falcao, Pavithra Vaideeswaran, Victoria Ciesielski, Babak Saboury, Syedmehrdad Mohammadianrasanani, Sanghun Sin, Raanan Arens, Drew A. Torigian, Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images, *Medical Image Analysis* 18 (2014) 752–771.
148. Wang, C., Zhao, Z. & Yu, Y. Fine retinal vessel segmentation by combining Nest U-net and patch-learning. *Soft Comput* **25**, 5519–5532 (2021). <https://doi.org/10.1007/s00500-020-05552-w>
149. Kaur, M., Kamra, A. Detection of retinal abnormalities in fundus image using transfer learning networks. *Soft Comput* (2021). <https://doi.org/10.1007/s00500-021-06088-3>

Figures

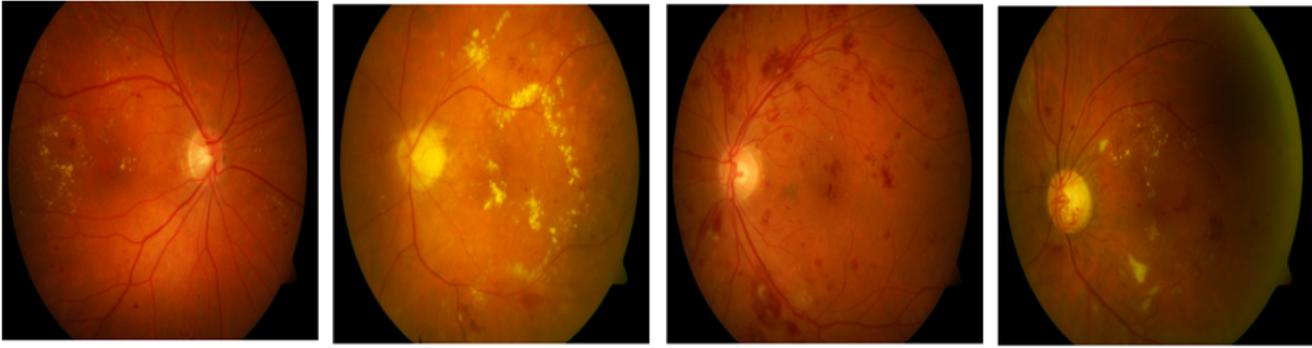


Figure 1

Retinal fundus images showing different signs of the DR [7]

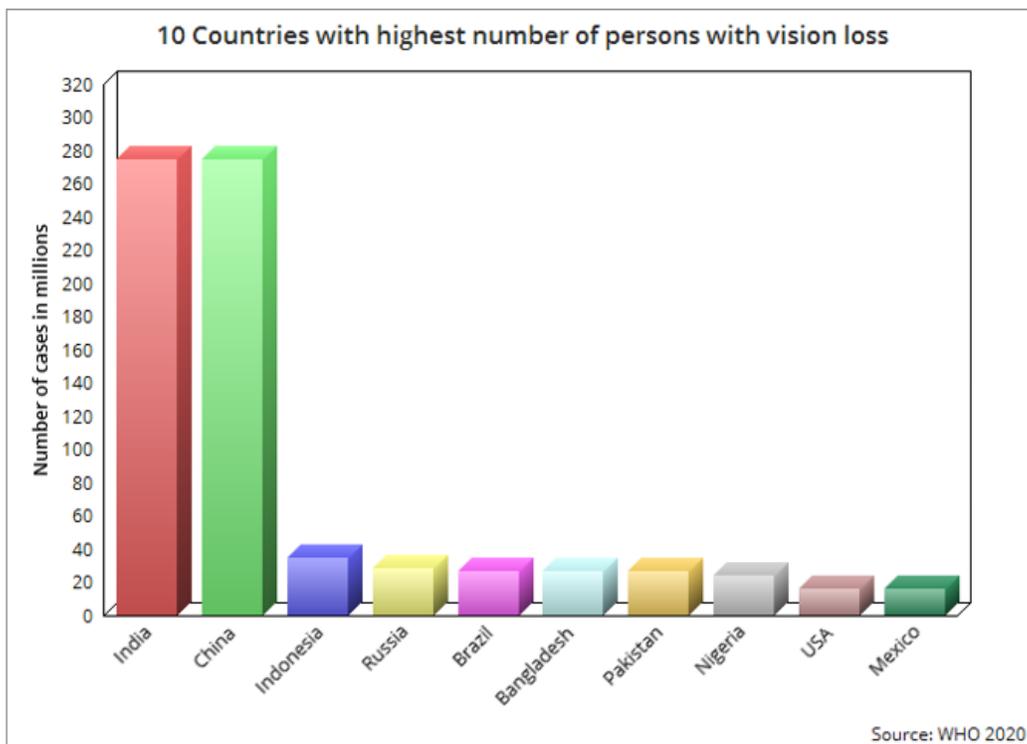


Figure 2

Cases of vision loss by Countries

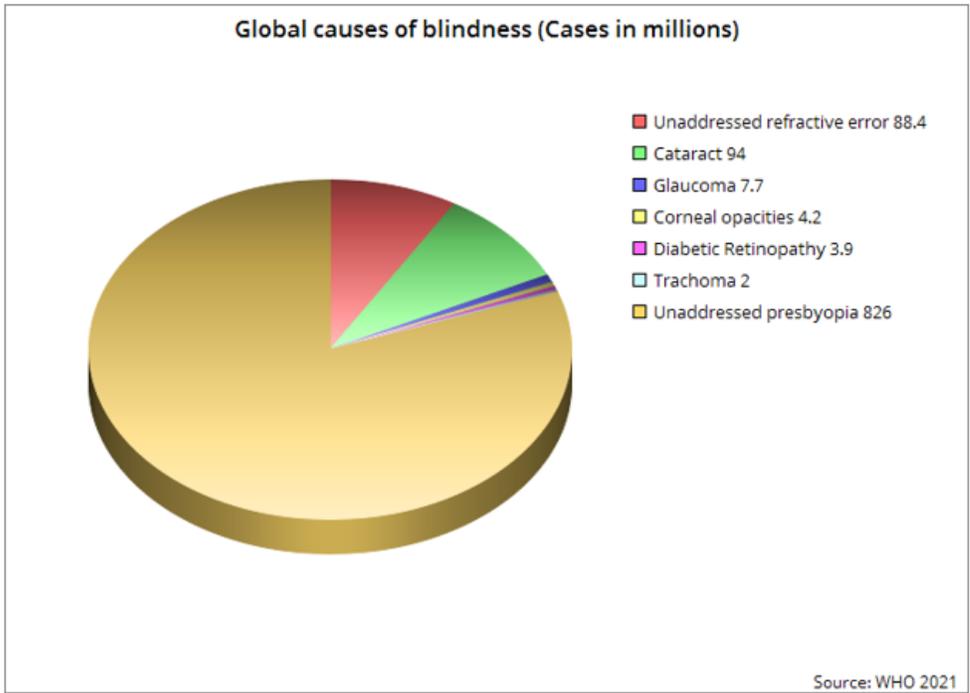


Figure 3

Global blindness figure

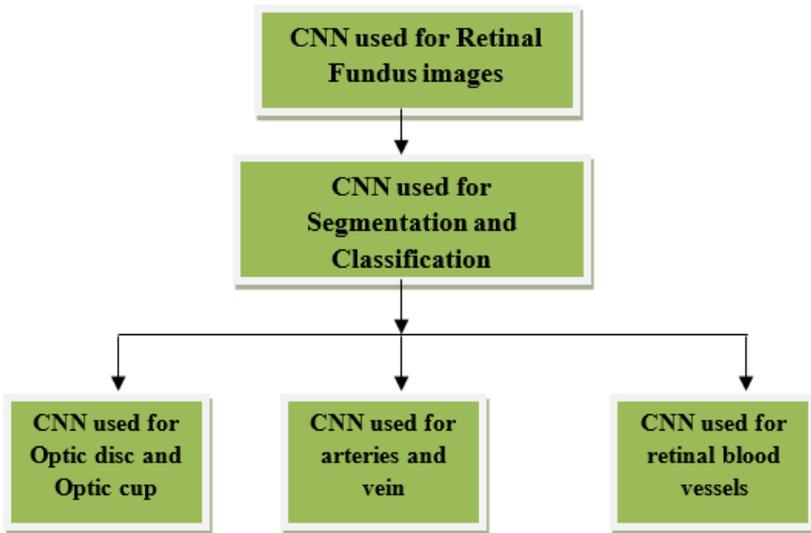


Figure 4

Categories of CNN methods for retinal fundus images

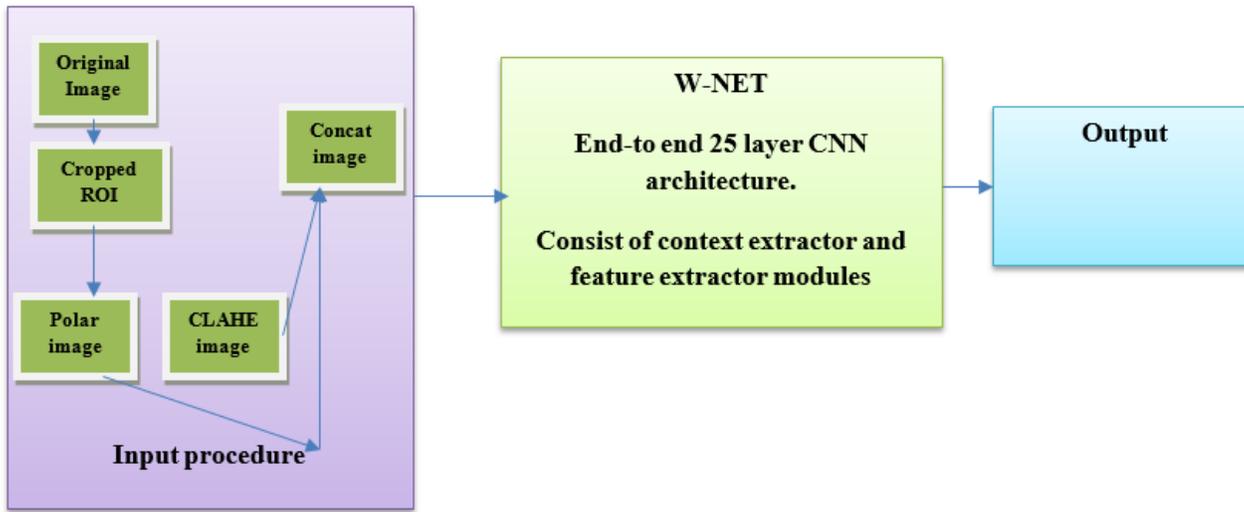


Figure 5

Multi-scale CNN [25]

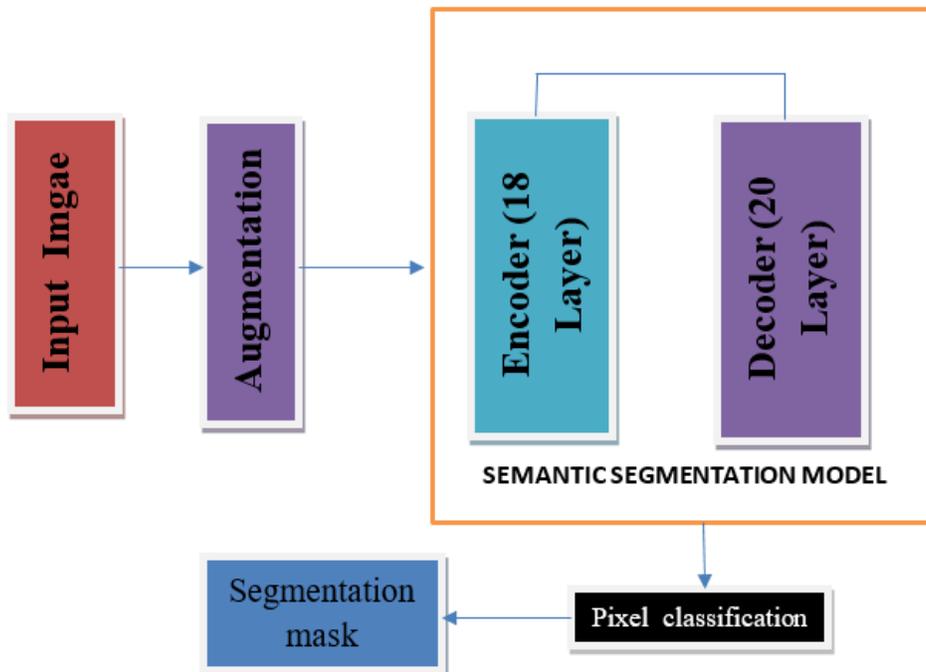


Figure 6

Label based semantic segmentation [33]

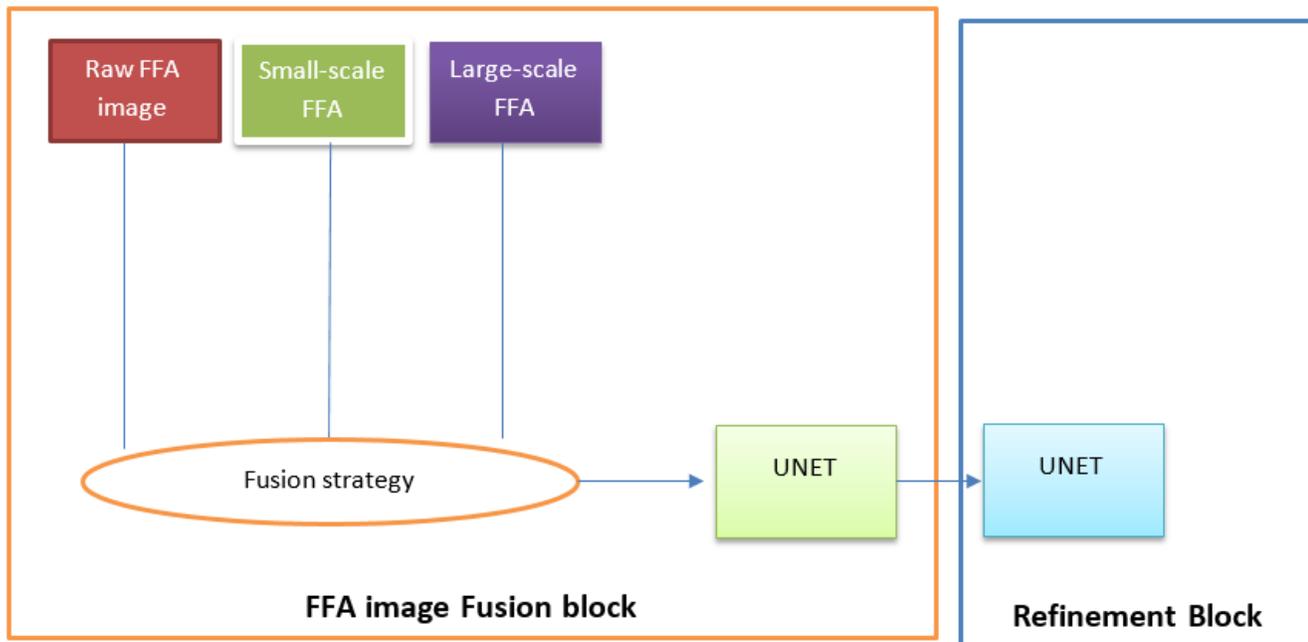


Figure 7

MCU-Net [47]

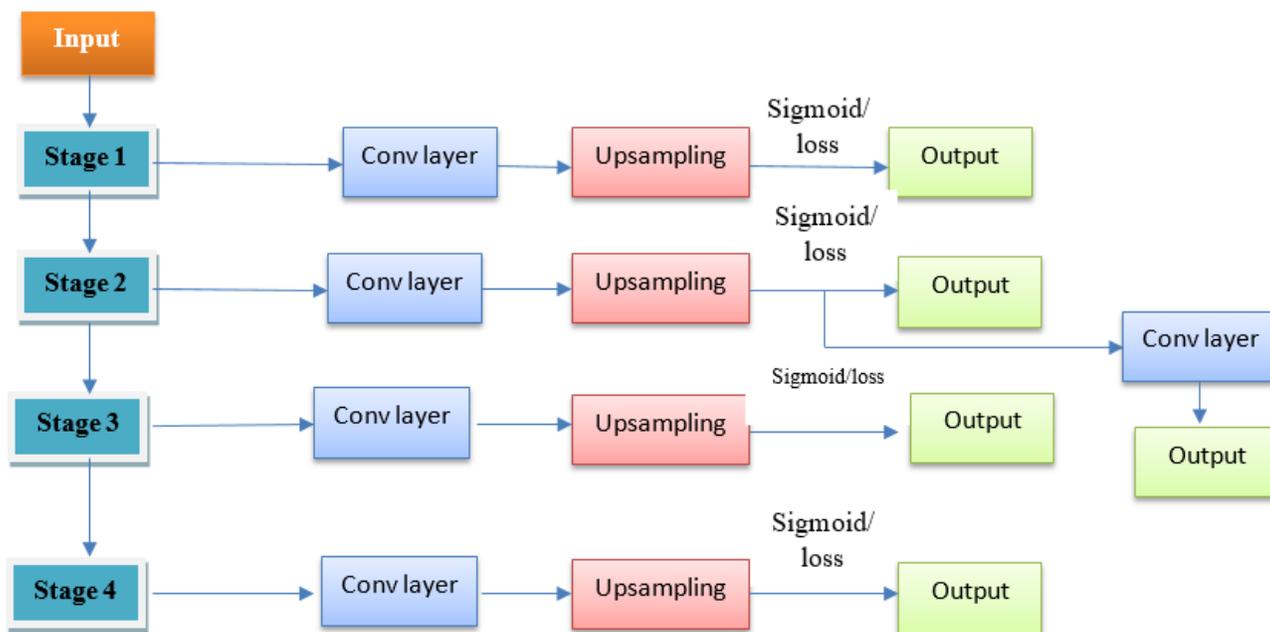


Figure 8

Multiscale CNN [82]

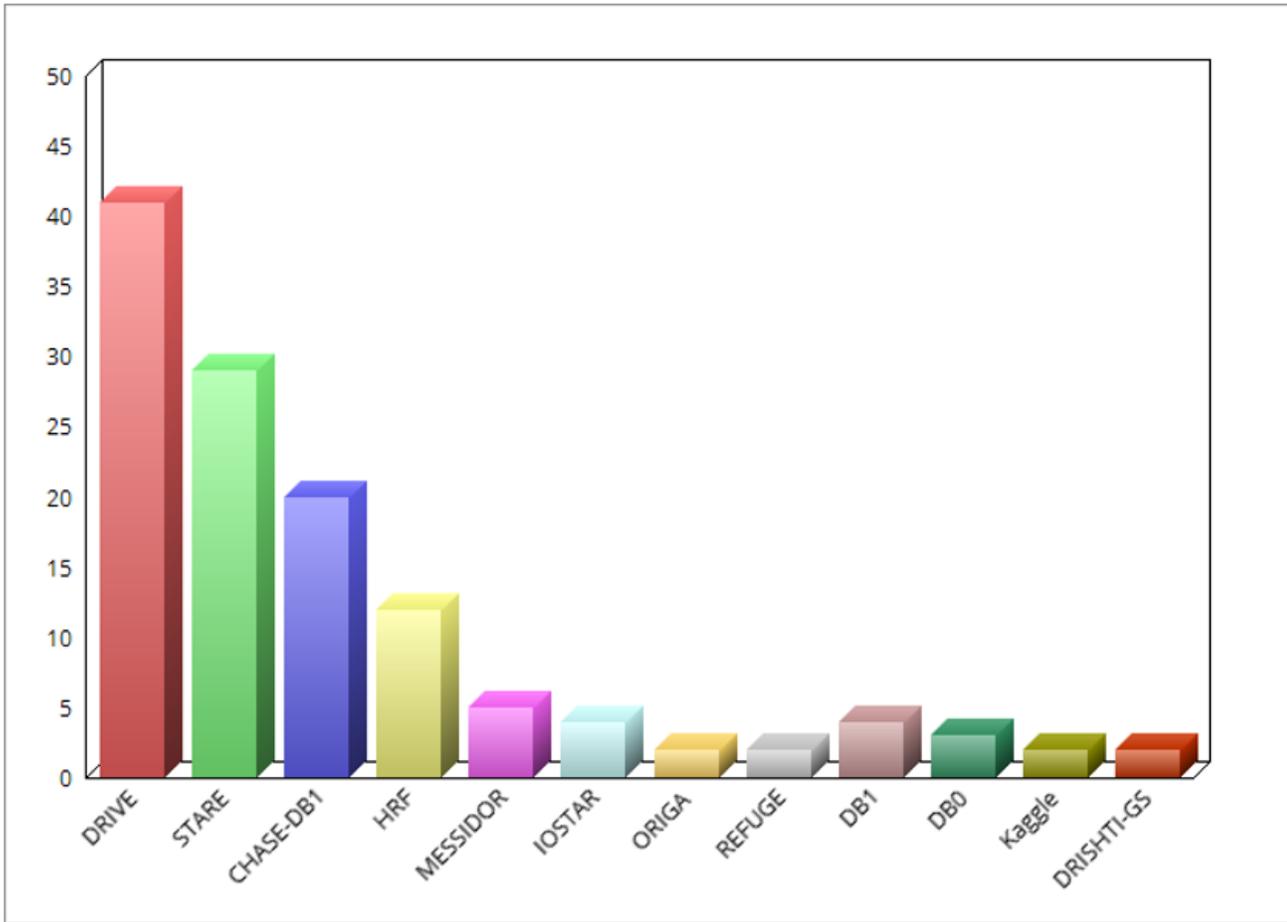


Figure 9

Datasets used for classification and segmentation of retina fundus images

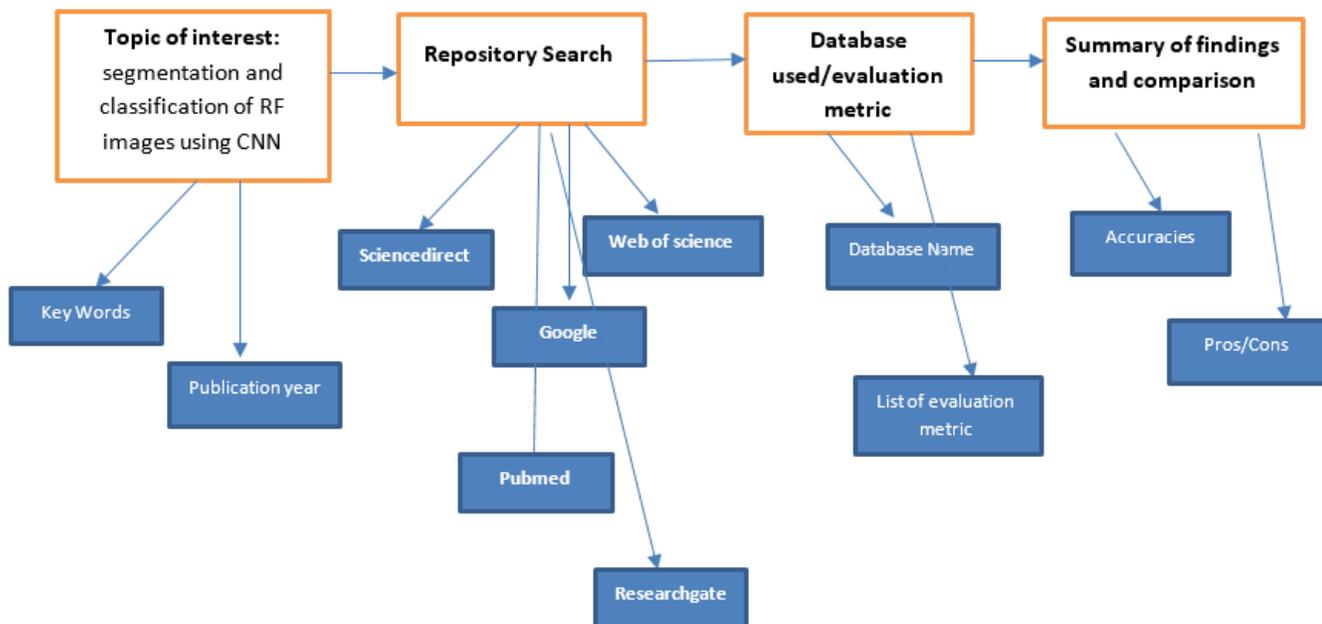


Figure 10

Approach for the review process

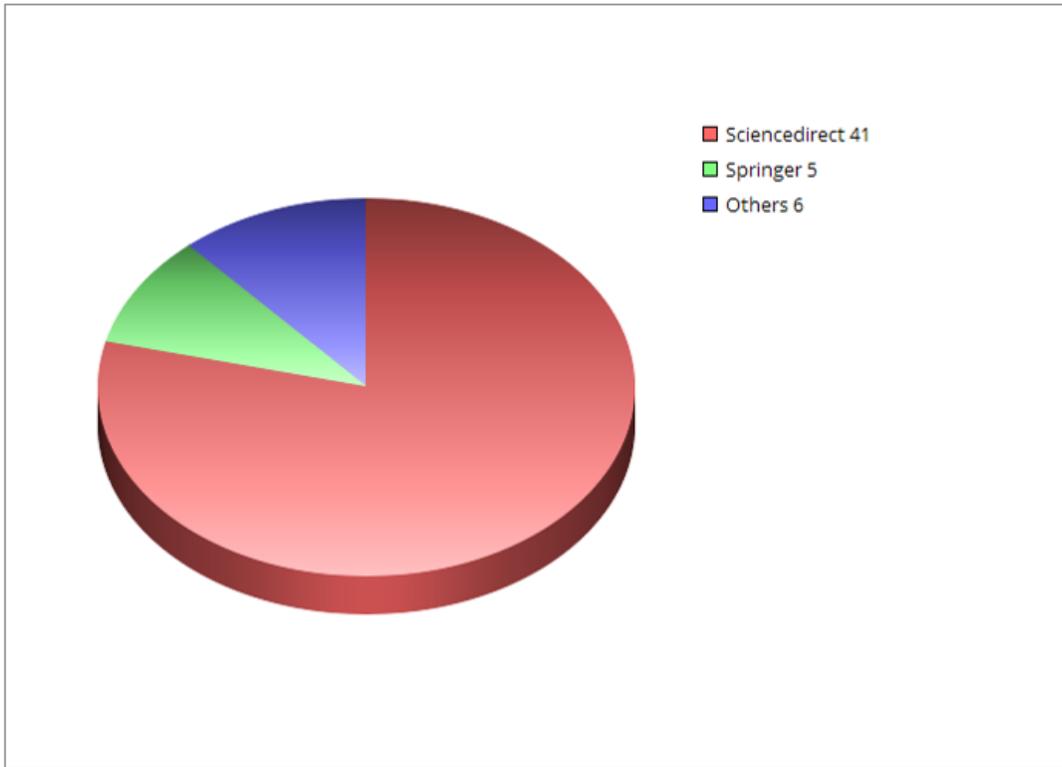


Figure 11

Analysis of data sources

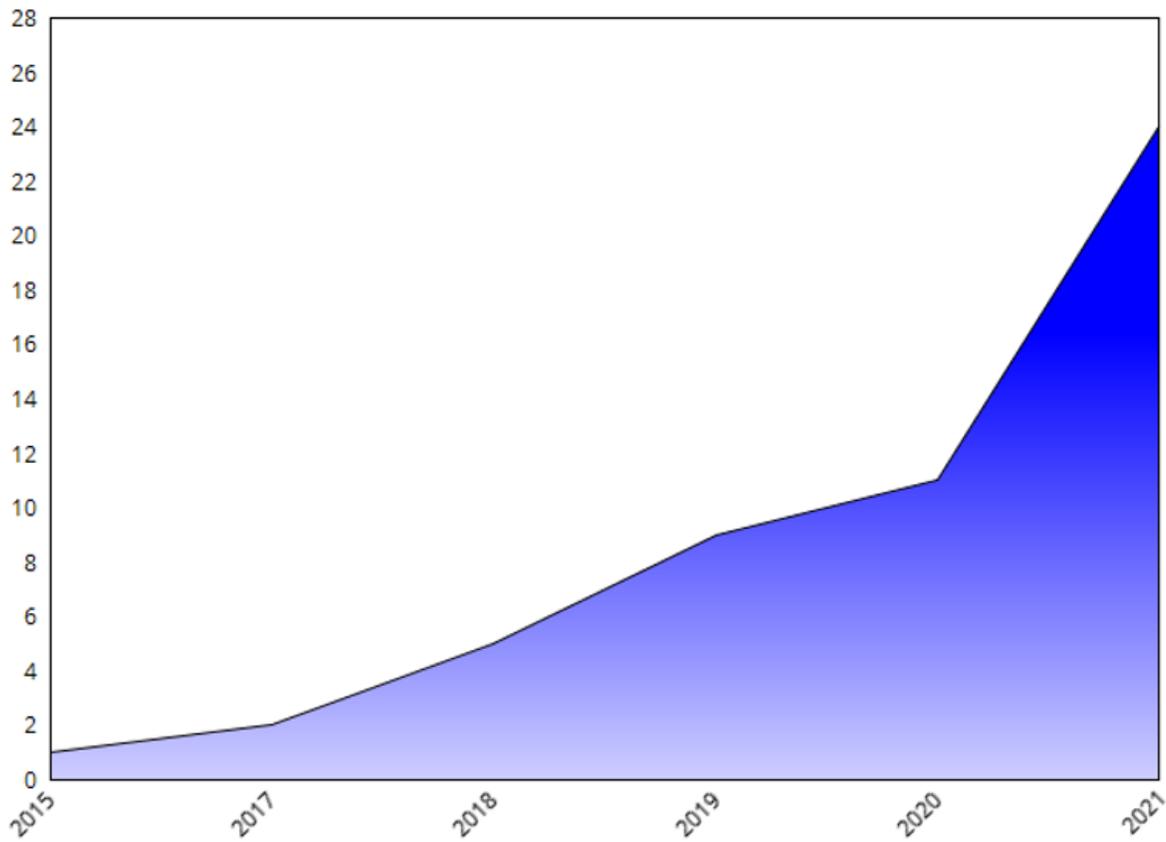


Figure 12

Number of papers published yearly

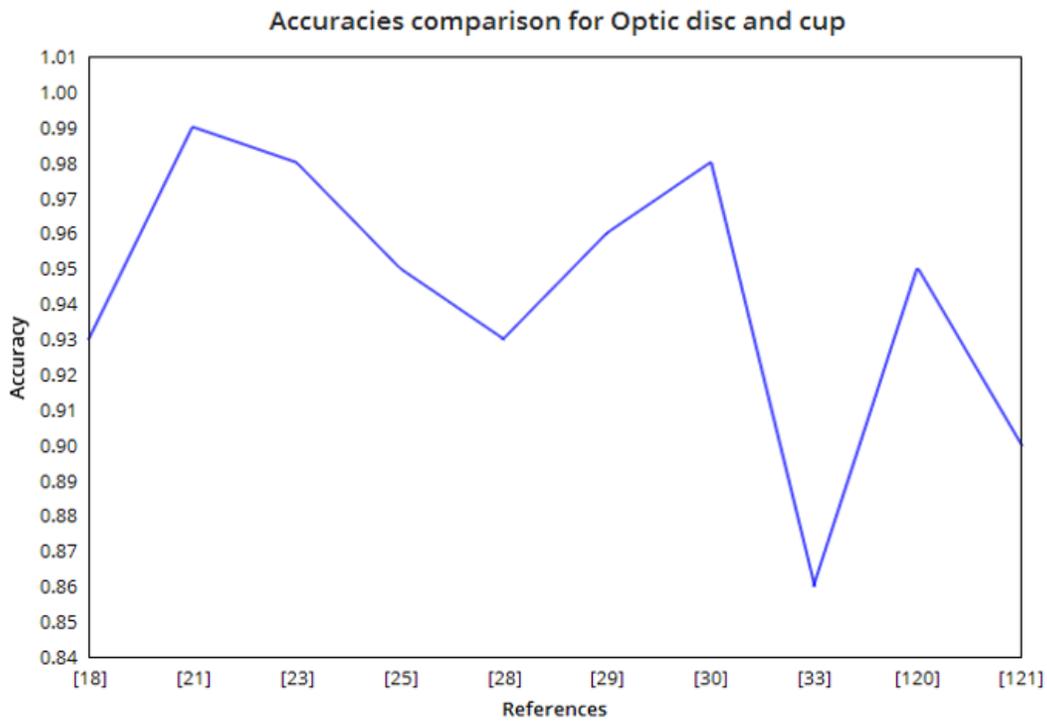


Figure 13

Accuracy comparison for optic disc and cup