A SYSTEMATIC STUDY OF DEEP LEARNING ARCHITECTURES FOR ANALYSIS OF GLAUCOMA AND HYPERTENSIVE RETINOPATHY

Madhura Prakash M, Deepthi K Prasad,
Meghna S Kulkarni, Spoorthi K, Venkatakrishnan S

Forus Health Private Limited, Bangalore, India

ABSTRACT

Deep learning models are applied seamlessly across various computer vision tasks like object detection, object tracking, scene understanding and further. The application of cutting-edge deep learning (DL) models like U-Net in the classification and segmentation of medical images on different modalities has established significant results in the past few years. Ocular diseases like Diabetic Retinopathy (DR), Glaucoma, Age-Related Macular Degeneration (AMD / ARMD), Hypertensive Retina (HR), Cataract, and dry eyes can be detected at the early stages of disease onset by capturing the fundus image or the anterior image of the subject's eye. Early detection is key to seeking early treatment and thereby preventing the disease progression, which in some cases may lead to blindness. There is a plethora of deep learning models available which have established significant results in medical image processing and specifically in ocular disease detection. A given task can be solved by using a variety of models and or a combination of them. Deep learning models can be computationally expensive and deploying them on an edge device may be a challenge. This paper provides a comprehensive report and critical evaluation of the various deep learning architectures that can be used to segment and classify ocular diseases namely Glaucoma and Hypertensive Retina on the posterior images of the eye. This review also compares the models based on complexity and edge deployability.

KEYWORDS

Deep learning architectures, U-Net, Ocular diseases, Retina, Glaucoma, Hypertensive retinopathy

1. INTRODUCTION

The quick and widespread application of deep learning in the field of medical image analysis has been greatly aided by developments in image processing, artificial intelligence, and computer vision-based techniques. This is also aiding the advancement of ophthalmology in practice. Detection of ocular diseases at their onset is extremely crucial to prevent the disease progression to advance stages or blindness in many cases. The latest computer vision techniques powered with the algorithms of artificial intelligence can ease the load of relying entirely on ophthalmologists for detection of the onset of disease or any other abnormalities in the fundus images, which might otherwise prove to be labour intensive for the experts.[1] With performance better than manual diagnosis, this can also aid in remote medical facilities for early diagnosis. Ocular diseases like glaucoma and hypertensive retina are one of the main reasons for the global blindness burden. Early diagnosis is a vital step in avoiding disease progression by availing of timely treatment. Due to the complexity and heterogeneity of the retinal image data, adopting a reliable automated approach for medical image analysis is still a difficult undertaking.

DOI: 10.5121/ijaia.2022.13603

Traditional image processing techniques extract low-level, hand-crafted features like color, texture, sharpness, brightness, morphological features, or a combination of these like in saliency maps for the application of classification algorithms. Machine learning algorithms like Support Vector Machine (SVM), Gaussian Mixture Models (GMM), K-Nearest Neighbour (KNN), random forest, AdaBoost, and naïve Bayes can be applied to the extracted feature set for the binary or multi-class classification on the fundus images. Traditional image processing techniques can also be applied to extract the region of interest in each fundus image. As compared to learned features in deep neural networks, good hand-crafted features are challenging to design manually and frequently incapable of achieving good classification performance.

Recently deep learning techniques are being used in the classification of the fundus images and in the segmentation of the image to point out the relevant areas of interest. Deep learning techniques can accommodate vast amounts and heterogenous data and they outperform traditional machine learning algorithms [2]. The ImageNet Challenge [3], 2012 to classify the 1000 classes of objects has created a revolution in the deep learning arena giving rise to several standard deep learning architectures like AlexNet, VGGNet, GoogLeNet, ResNet, DenseNet, and so on each with an established breakthrough performance. These architectures and their variants have been used in literature by several researchers for the classification and segmentation of fundus images.

This research work has focused on the critical review of the existing deep learning solutions. The performances of these solutions have been analyzed and the work also focuses on throwing light on those solutions which aid in developing models that are lightweight and can be deployed easily on any device. A discussion on the future directions in the application of deep learning models in medical image classification and segmentation is also a significant contribution of this work.

The rest of the paper is organized as follows. Section 2 provides an overview of deep learning methodology with a note on the basic architectures used in image segmentation and classification. Section 3 provides a review of the recent works and solutions for glaucoma and hypertensive retina disease detection. Section 4 provides an overview of recent deep-learning architectures used in ocular disease classification. Section 5 discusses the necessities and possible solutions for generating lightweight deep learning models for edge deployability. Section 6 gives a list of publicly available datasets for ocular disease images with their corresponding links. Section 7 discusses some of the common performance metrics used in the analysis of classification and segmentation solutions in deep learning. The review concludes with a discussion on the future direction in the application of DL models for medical image segmentation and classification, in section 8.

2. AN ENCAPSULATION OF DEEP LEARNING TECHNIQUE

In essence, a deep learning architecture is a neural network with several layers. A neural network is a network of several neurons or perceptrons. A neuron is an elementary unit in an artificial neural network. It is a mathematical entity that takes multiple inputs and produces an output. The input signals to a neuron are linearly combined, an activation function is used to bring these linearly combined values within the range of interest and this output signal is passed to the subsequent neurons in the network [4].

A neural network depicted in figure 1 has several layers of artificial neurons. Neurons in each layer receive input data from the neurons in the previous layers. The input signal passes through the layers of the network. The layers are broadly categorized as the input layer, hidden layer, and output layer. Each layer processes the input signal resulting in the desired outcome.

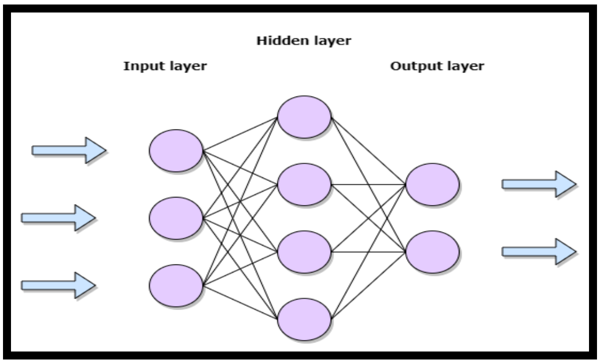

Figure 1. An illustrative diagram of a Neural Network

The three kinds of learning that take place within neural networks are

- Supervised Learning - Based on a set of labelled data
- Unsupervised Learning - There is no labelled data, and the model generates the output based on patterns found in the output data.
- Reinforcement Learning - network learns based on the continuous feedback it receives from the environment

A Neural Network has several advantages

- Resilience -Even if a few neurons in a neural network are malfunctioning, the network will still be able to produce outputs.
- Synchronously learn and adapt: Capable of synchronized learning and quick contextual adaptation. The networks also have the ability to adapt and learn how to do various jobs.
- Perform multiple jobs simultaneously: capable of handling numerous tasks at once based on the input data to produce the desired output. [4]

Some of the pitfalls of neural architectures are:

- Reasoning- Neural networks offer a remedy for a problem. The network's complexity prevents it from explaining why it made certain conclusions. Hence, explaining ability is a major area of concern in deep neural networks.
- Determination of appropriate network architecture for a particular problem is challenging. There is no specified rule of thumb for a neural network procedure. A proper network structure is achieved by trying the best network, in a trial-and-error approach. It is a process that involves refinement.
- Dependency: Neural networks are highly dependent on high end systems with parallel processing capabilities [4].

2.1. Categories of the Deep Architectures

2.1.1. Convolutional Neural Networks (CNN)

The block diagram of CNN models is depicted in figure 2. CNN consists of the convolutional layer, the Activation layer, the Pooling layer, and the Dense layer. Activation functions mimic the firing of a neuron and provide the model with non-linearity. The model is trained using the backpropagation algorithm, which is used to adjust weights. Dropout and batch normalization are two frequent techniques for accelerating convergence and preventing overfitting.

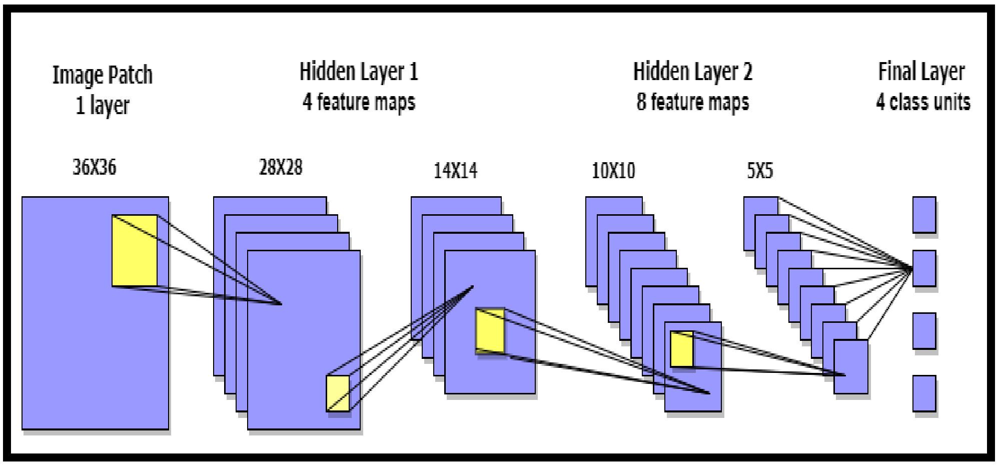

Figure 2. A schematic diagram of a Convolution Neural Network

Convolutional layer generates feature maps from the input image by making use of kernel filters to convolve around the image. The activation layer includes the application of activation functions like rectified linear activation (ReLu) on the output of the convolutional layer so as to bring the values of the feature maps within a standard range. Pooling is a subsampling process used to squeeze the feature size. A fully connected layer is used as the final layer for classification. A fully Convolutional Network, often known as CNN without a fully connected layer, is utilized for applications like image segmentation [5].

2.1.2. Fully Convolutional Networks

Fully convolutional networks as depicted in figure 3 consist of downsampling (convolution), upsampling (deconvolution), and pooling layers. Faster calculation and fewer parameters result from the absence of a dense layer. The down-sampling path is generally made of convolutional layers and an up-sampling path is made of deconvolutional layers. These can incorporate information transfer through layers with multiple configurations and skip connections that skip more than one layer for improved learning [5].

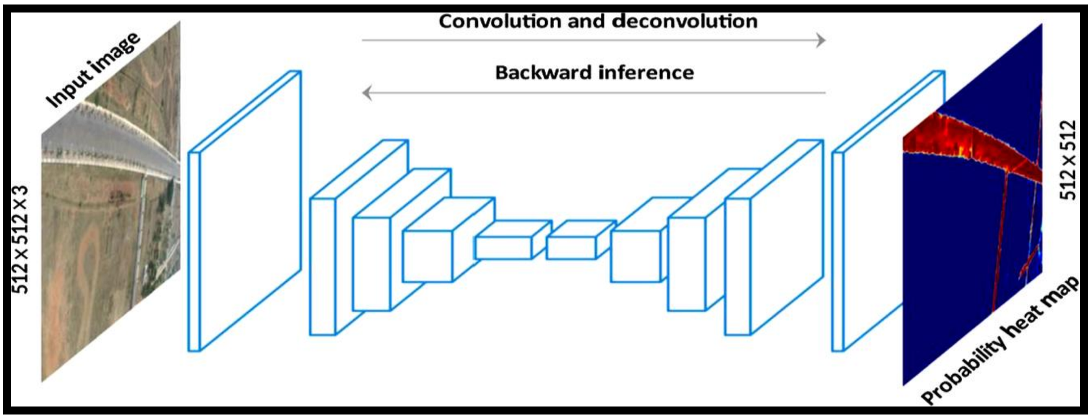

Figure 3. A schematic diagram of a Fully Convolution Neural Network

2.1.3. Autoencoders

In autoencoders data is effectively compressed and encoded using an unsupervised neural network, which is then utilized to reconstitute data from the smaller encoded form. As shown in figure 4, it is comprised of an associate encoder that reduces the dimensions and a bottleneck that has all-time low dimensions of input data, a decoder that learns to reconstruct the information from the coding and reconstruction loss that is employed to compute the performance of the decoder's output with relevancy the first data. The network is trained using backpropagation to minimize the reconstruction loss in order to get the reconstructed data to be as similar to the original data as possible [5].

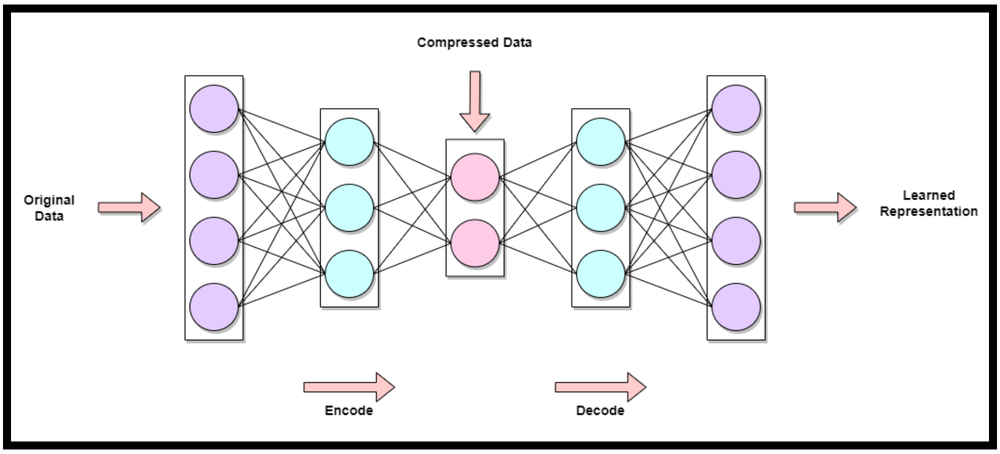

Figure 4. A schematic diagram of an Auto Encoder architecture

2.2. Model Training Procedures

The weights associated with the edges of the neural network are initialized at random. Training of the network includes tuning the weights associated using stochastic gradient descent-based backpropagation algorithm. Some of the techniques to train a neural network are discussed below.

2.2.1. Transfer Learning process

Requirement of the large dataset is usually a critical constraint for the efficient training of Deep learning algorithms. The datasets available for ophthalmic disease diagnosis may not have sufficient varied images which might lead to overfitting of the model. Transfer learning, illustrated in figure 5 presents a possible solution to this problem. In transfer learning, a pre-trained network on another but the similar domain is used and the weights are fine-tned from here. This proves to be advantageous because the weights of the network can be initialized to a meaningful starting value instead of initializing at some random values [5].

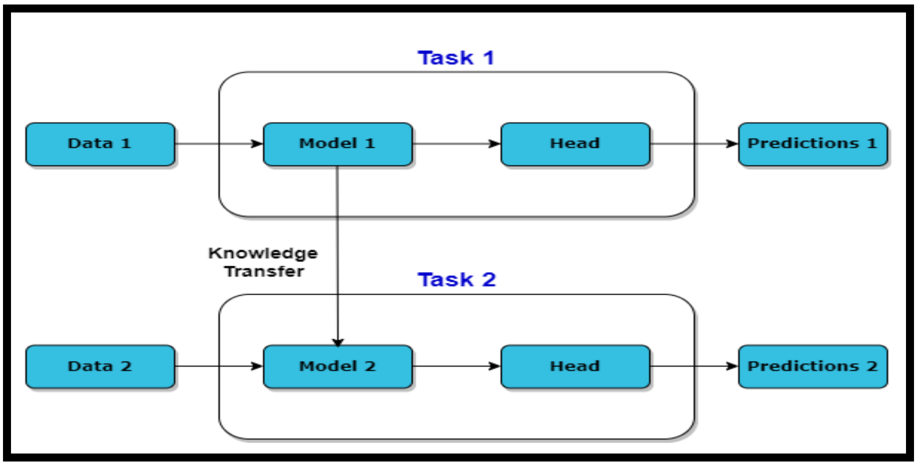

Figure 5. Overview of Transfer Learning procedure

2.2.2. Ensemble Learning

In ensemble learning based several deep learning models are trained independently. The results are obtained by polling. The majority voting method, which is frequently used, includes choosing the most frequent outcome as the final forecast. It is predicated on the idea that errors from models that have undergone independent training are unlikely to coincide [5].

3. Review of Works in Glaucoma and Hypertensive Retina Classification and Segmentation using DL Approaches

3.1. Glaucoma

The ocular disease glaucoma harms the optic nerve and impairs vision. Glaucoma causes persistent vision loss that is incurable and cannot be reversed with current medical practices. However, glaucoma progresses slowly and without any earlier warning signs, making the diagnosis difficult.

The Online Retinal Fundus Image Database for Glaucoma Analysis and Research (ORIGA) dataset images were used in [6] to perform binary classification of the presence or absence of Glaucoma. A U-Net model was used to segment the optic cup followed by feature extraction of the segmented image using pretrained DenseNet-201 model and classification using Deep Convolutional Neural Network (DCNN). A pre-processing pipeline sig Otsu thresholding and morphological operation was defined in [7] to generate ground truth for optic disc localization. A Region based Convolution Neural Network (RCNN) was used for extracting the relevant region in the fundus images and a DCNN was used to classify the images. The proposed methodology was subjected to exhaustive testing on various publicly available datasets like ORIGA, Digital Retinal Images for Vessel Extraction (DRIVE) (HRF), and MESSIDOR.

Pre-processing techniques like Gaussian Filtering and morphological operations were proposed in [8]. The processed image is further subjected to processing with Contrast limited adaptive histogram equalization (CLAHE) and Sobel edge detection algorithms. The region of interest was extracted based on the watershed algorithm. The optic disc and optic cup segmentation were performed based on two different CNN encoder-decoder-based segmentation algorithms. The authors in [9] combined accurate vertical cup-to-disc ratio (VCDR) estimation and glaucoma detection in one study. They analysed the importance of optic nerve hypoplasia (ONH) and peripapillary regions in deep learning experiments. The work was based on cropped fundus images. The presence of significant pixel information on glaucoma and VCDR outside the ONH was revealed by this work. They have reported the area under the curve (AUC) values up to 0.88. A DL-based approach namely EfficientDet-D0 with EfficientNet-B0 was used in [10] for Glaucoma classification and localization. The features from the samples were extracted by using the EfficientNet-B0 architecture. Then, the Bi-directional Feature Pyramid Network (BiFPN) module of EfficientDet-D0, repeatedly executes top-down and bottom-up key point fusion on the features extracted. In the final stage, a localized area comprising glaucoma lesions and the associated class was output. Optic Disk localization and Glaucoma Diagnosis Network (ODGNet) based on two phases was proposed by the authors in [11]. For efficient OD localization from the fundus images, a visual saliency map combined with shallow CNN was used in the first phase. Pre-trained models based on transfer learning were employed in the second phase to diagnose glaucoma. On five publicly available retinal datasets (ORIGA, HRF, DRIONS-DB, DR-HAGIS, and RIM-ONE), transfer learning-based models like AlexNet, ResNet, and VGGNet combined with saliency maps were tested for their ability to distinguish between normal and glaucomatous pictures.

The focus of the work in [12] was to consider the cup and disk segmentation problem jointly instead of individually segmenting the optic disc and cup. The minimum bounding boxes of the optic disc and cup are initially detected. The combined task of the optic disc and cup segmentation was based on a region-based convolutional neural network named Joint R-CNN. The feature extraction was based on the Atrous-VGG16 framework. Here, the last three convolution layers of VGG16 are replaced with 9 atrous convolution layers. The atrous convolutions enlarge the receptive field while keeping the size of the feature map. Without changing the parameters, it may be used to extract higher-resolution, deeper feature maps. The authors in [13] suggested JointRCNN, disc proposal network (DPN), and cup proposal network (CPN) to construct bounding box proposals for optical disc and cup, respectively. Since it is already known that the optic cup lies within the optic disc, a disc attention unit connecting DPN and CPN is recommended. In this module, an appropriate bounding box of the optic disc is initially chosen and is then carried ahead as the foundation for optic cup identification. The vertical cup-to-disc ratio (CDR), which is computed and utilised as a criterion for disease prediction, is obtained after acquiring the disc and cup regions, which are the adorned ellipses of the respective detected bounding boxes.

3.2. Hypertensive Retinopathy

The thickening of the retina’s blood vessel walls may occur as a result of high blood pressure. This results in the narrowing down of the blood vessels, restricting the blood flow into the retina. The retina might become swollen because of this restricted blood flow. The retinal blood vessels gradually get damaged. This restricts the retina's capacity to operate and puts too much pressure on the optic nerve, impairing eyesight. The name of this condition is hypertensive retinopathy (HR). A hypertensive retinopathy (HYPER-RETINO) framework has been developed by authors in [14] to classify HR based on five grades. Based on previously trained HR-related lesions, a HYPER-RETINO system was implemented. Several procedures were used to create this HYPER-RETINO system, including pre-processing, semantic and instantiation categorization for the discovery of lesions related to HR, as well as a DenseNet structure to categorize the stages of HR.

The researchers in [15] focused on developing an evaluation system for retinal vessel alterations caused by hypertension using a deep learning algorithm. Both the retinal venules and arterioles' combined area were measured. Each image's retinal vessels were automatically identified as either venules or arterioles. The complete venular area (VA) and complete arteriolar area (AA) were then calculated. The relationships between AA, VA, age, systolic blood pressure (SBP), and diastolic blood pressure were carefully observed. Authors in [16] have suggested a Dense Accumulation Vessel Segmentation Network (DAVS-Net), which can provide good segmentation accuracy with only a few trainable parameters. This network employs an encoder-decoder framework, where edge information is passed from the initial layers of the encoder to the final layer of the decoder, in order to achieve faster convergence. Dense connections are taken into account by DAVS-Net as a way to improve the precision of semantic segmentation.

The authors of [17] offer the recent approaches utilized by scientists to forecast hypertensive retinopathy. The use of features found in retinal pictures for the early identification of HR has also been discussed. The authors of [18] have suggested a deep learning-based automated system that uses U-Net and Dense-Net architectures to detect and score papilledema. There are two primary steps to the suggested strategy. First, in the fundus retinal image, the optic disc and its surroundings were located and clipped for input to Dense-Net, which categorizes the optic disc as normal or papilledematous. The second stage entails applying the Gabor filter to the Dense-Net categorized papilledema fundus image. The segmented vascular network was created from the pre-processed papilledema image using U-Net, and the vessel discontinuity index is then obtained (VDI).

The authors in [19] offer a method for the simultaneous segmentation and categorization of the retinal arteries and veins from eye fundus pictures. This joint task is broken down into three segmentation problems using the approach: the segmentation of arteries, veins, and the entire vascular tree. A unique loss called BCE3 was proposed. This combines the independent segmentation losses of the three classes of interest, to train the networks using this methodology. To review the classification of hypertensive retinopathy, the authors in [20] have proposed a method to calculate the multiple uses of artery and vein diameter ratio (AVR) in addition to changes in position with the optic disc in retinal images utilizing Deep Neural Networks (DNN) and Boltzmann Machines approach.

The authors in [21] have proposed a method to determine the combined features of artery and vein diameter ratio (AVR). The changes in position with the optic disc in retinal images were considered to compute the classification of hypertensive retinopathy using the Deep Neural Networks (DNN) and Boltzmann Machines approach. The transfer learning-based convolutional neural network (E-TLCNN) model was proposed for diagnosing HR using high-quality images from fundus images in [22]. Transfer learning was used to identify the results' stages and to appropriately evaluate the findings. In order to classify the features and emphasize the severity, such as reading the diabetic retinopathy and AVR, a new model using CNN architecture based on DenseNet was proposed. A comparative study of the performance of deep learning algorithms versus traditional approaches in limited data sources for detecting Hypertensive Angiopathic Retinopathy was conducted by researchers in [23], The study was carried out in a clinical environment based on a multi-racial population.

4. CUTTING-EDGE DL MODELS IN CLASSIFICATION AND SEGMENTATION

Several deep learning architectures have been used in literature for image segmentation and classification purposes. Few of these follow transfer learning procedures to transfer the model training weights from a standard architecture while the rest follow a hybrid approach combining the parts of various architectures to accomplish multitasking of classification followed by segmentation. Deep learning architectures do not need hand-crafted features and there is no strict requirement of pre-processing the image before passing through the network. However, a pre-processed image reduces the complexity and overload on the architecture thereby improving performance efficiency. Table 1 overviews the architectures that have achieved significant results in ocular disease classification and segmentation.

Table 1. A review of significant architectures in image classification and segmentation

Refer ence	Focus Area	DL Architecture Used	Pre-processing applied	Performance
[24]	Retinal Image Segmentation for blood vessel segmentation on fundus images	Faster R-CNN	Features from CNN	92.81% sensitivity
[25]	Classification of RD (Retinal Detachment) versus Non-RD fundus images.	ResNet50	Features from CNN	Sensitivity, Specificity, Precision of 99.00%, 99.99%, 99.99%, respectively
[26]	Medical Image Segmentation	CNN with Residual Block	Features from CNN	Precision - 0.9173 Recall - 0.9139
[27]	Synthesis of fundus images for classification of Age-Related Macular Degeneration (AMD).	GANs with ResNet-18 architecture	Hough Circle Transform, resizing, and central cropping to remove the background.	Accuracy - 77.5%
[28]	Assessing the quality of Fundus images	Inception-V3	Features from CNN	Mean absolute error - 0.61 (0.54-0.68)
[29]	Classification of AMD and non-AMD fundus images	Basic CNN	Features from CNN	Accuracy - 77.5%
[30]	Classification and segmentation of Drusens in AMD	DeepLabV3+ and UNet	Features from CNN	Accuracy - 82%
[31]	Registration technique based on deep learning to align multi-modal retinal images gathered from clinical studies	Dual VGG16 extractors used in a Siamese Network architecture	Correlation matrix for feature matching	Sensitivity- 0.997 Specificity- 0.662
[32]	Classification of AMD and drusen identification on OCT images	AlexNet, VGG, GoogLeNet	SIFT features, correlation matrix using L2 norm metric.	Mean errors ranging from 54-69 µm
[33]	Classification of fundus image into AMD and Non-AMD	VGG16	Cropped and resized to 224*224 pixels	Accuracy- 91.2%
[34]	Multiclass classification of retinal fundus images	Generative Adversarial Networks (GANs)	Features from CNN	Accuracy 87%
[35]	Hard exudates segmentation method to diagnose Diabetic Retinopathy (DR) in the early stage	U-Net based architecture	Super-pixel algorithm	Accuracy- 97.95%
[36]	Medical image classification and segmentation	U-Net based architecture	Features from CNN	Accuracy - 92%

All the proposed solutions mentioned in Table 4.1 pre-process the image by using standard image processing algorithms. These techniques help in gaining better efficiency when the pre-processed image is passed through the deep learning architecture. U-Net is a standard model that is used for segmentation purposes. Generative Adversarial Networks are gaining momentum, especially in situations where there is data sparsity as these architectures help in generating sample images for the model to learn.

5. TOWARDS LIGHTWEIGHT MODELS FOR EDGE DEPLOYABILITY

Deep learning models require high computational power and resources because of which it is a challenge to customize these models to suit lightweight devices such as mobiles and the Internet of Things (IoT). Some of the challenges in optimizing the performance and efficiency of deep learning architecture solutions are: -

- Limited availability of relevant data
- Determining the number of algorithm runs (epochs)
- Model variability and reproducibility
- Developing a model that can rapidly learn, segment accurately, and automatically suppress outputs that were misclassified
- Enabling on-device deployment of the models
- Explosion of models in a problem solution leading to exhaustion of available resources

Efficient deep learning must focus on Inference Efficiency (Number of Model parameters and memory requirement) and Training Efficiency (Time and number of devices required for training). Some of the techniques that can be used for generating lightweight models are [37]: -

5.1. Compression Techniques

Compressing the layers and edges of the deep neural networks using pruning. Further, with little loss in quality, quantization can be used to compress the weight matrices of a layer by lowering its precision (for example, from 32-bit floating point values to 8-bit unsigned integers). Binarization is an extreme case of quantization where the weight representation is reduced to only two bits.

5.2. Learning Techniques

Distillation techniques that enable the improvement in the accuracy of a smaller model by learning to mimic a larger model can be used. This can be used in the Student-Teacher architecture model. With no loss of validity, knowledge distillation moves information from a big model to a small one. Smaller models can be used on less potent hardware because their evaluation costs are lower.

5.3. Regularization

In regularization, the following can be incorporated to improvise on the model training efficiency

- The larger weights can be penalized and allowed to decay.
- Large activations can be penalized by introducing an activation constraint

5.4. Automation

Automated hyper-parameter optimization (HPO), which increases accuracy by adjusting the hyper-parameters Then, a model with fewer parameters could replace this. locating a model that maximizes loss/accuracy as well as other factors like model latency and model size.

5.5. Systematized Architectures

Parameter sharing between Convolution Neural Networks used for image classification. By doing so, it is not necessary to learn unique weights for each input pixel, and they become more resistant to overfitting.

5.6. Infrastructure

Along with the tools necessary specifically for deploying effective models, such as Tensorflow Lite (TFLite) and PyTorch Mobile, this also contains the model training framework such as Tensorflow, PyTorch, etc.

Some of the notable frameworks for cross-platform deployability are as follows [38]:

- TensorFlow Lite is a multi-platform framework for on-device machine learning. It can be used on Android, iOS, Linux, and microcontrollers. TensorFlow Lite provides features like compression during model conversion from TensorFlow-to-TensorFlow Lite format and GPU/NPU acceleration support.
- PyTorch Mobile is another multiplatform framework to work with PyTorch models on Android, iOS, and Linux. This tool is in its beta stage but it already features an 8-bit quantization during conversion from the PyTorch to PyTorch Mobile format and support of the GPU/NPU acceleration.
- CoreML is a machine learning framework specific to the iOS platform. It leverages Apple hardware, including CPU, GPU, and NPU to maximize the model performance while minimizing memory and power consumption. It also supports such key features as quantization from 16 to 1 bits and on-device fine-tuning with local data, which can help to personalize application behavior for each user.
- A new format for exchanging deep learning models is called ONNYX, or Open Neural Network Exchange Format. Deep learning models will supposedly become portable, preventing vendor lock-in.

6. PUBLICLY AVAILABLE DATASET FOR OCULAR DISEASE

The model generated by the deep learning architecture is only as good as the data. For the model to achieve better accuracy, sensitivity, and specificity of the data provided to the model must be of good and relevant quality. It is also imperative to generate models that generalize beyond the given data for training. Hence an important challenge in deploying deep learning models is the availability of data. Table 2 lists some of the standard datasets available for ocular disease identification.

Table 2. A list of the publicly available standard dataset for ocular disease identification

Sl no	Fundus Image Dataset	Details	Link
1	e-optha	The two sub-databases are e-ophtha-MA (MicroAneurysms) and e-ophtha-EX (EXudates). 47 photos with exudates and 35 without lesions are included in the database of images with exudates. Database of images with microaneurysms: It has 233 images without a lesion and 148 images with microaneurysms or minor haemorrhages.	https://www.adcis.net/en/third-party/e-ophtha/
2	Ocular Disease Intelligent Recognition (ODIR)	ODIR contains information on 5,000 individuals, including age, color images of the fundus in both eyes, and diagnostic terms used by doctors. This dataset is a collection of patient data that Shanggong Medical Technology Co., Ltd. has gathered from various hospitals and medical facilities in China. These institutes use a variety of cameras, including Canon, Zeiss, and Kowa, to acquire fundus images, which produce photos with different image resolutions.	https://odir2019.grand-challenge.org/dataset/
3	Automated Retinal Image Analysis (ARIA) Data Set	143 photos of either diabetics, healthy individuals, or those with age-related macular degeneration (AMD). obtained using a 50-degree FOV Zeiss FF450+ fundus camera. The OD and fovea are annotated in addition to the vessel segmentation.	http://www.damianjjfarnell.com/?page_id=276
4	FIRE (Fundus Image Registration Dataset)	A dataset for registering retinal images that have been annotated with real-world data is available. 129 retinal pictures make up 134 image pairs in the collection. Depending on their qualities, these image pairs are divided into three groups. The photos were taken using a Nidek AFC-210 fundus camera, which has a FOV of 45 degrees in both the x and y directions with a resolution of 2912x2912 pixels.	https://projects.ics.forth.gr/cvrl/fire/
5	HRF- (High-Resolution Fundus)	Currently, there are 15 photos of healthy individuals, 15 photographs of patients with diabetic retinopathy, and 15 images of patients with glaucoma in the public database. Additionally, field of view (FOV) masks are offered for specific datasets.	https://www5.cs.fau.de/research/data/fundus-images/
6	DR HAGIS:	Diabetic Retinopathy, Hypertension, Age-related macular degeneration, and Glaucoma Images The following four co-morbidity subgroups are included in the DR HAGIS database. Images Glaucoma subgroup images, 1–10 Hypertension subgroup images, 11-20 21–30: Images of diabetic retinopathy 31–40: Subgroup for age-related macular degeneration	https://personalpages.manchester.ac.uk/staff/niall.p.mcloughlin/
7	AV-WIDE 12	30 ultra-wide FOV images, including both	https://people.duke.edu

		normal and AMD-affected eyes. Each image is approximately 900*1400 pixels in size and was captured with an Optos 200Tx UWFI camera.	/~sf59/Estrada_TMI_2015_dataset.htm
8	STARE (Structured Analysis of the Retina)	97 images (59 AMD and 38 normal), each with a resolution of 605700 pixels and obtained with a fundus camera (TOPCON TRV-50; Topcon Corp., Tokyo, Japan) with a 35° field of view. - There are also depictions of Drusen kinds.	http://www.ces.clemson.edu/~ahoover/stare
9	ADAM Challenge	-(Automatic Detection challenge on Age-related Macular degeneration) Baidu Research (Sunnyvale, CA, United States) gathered a sizable dataset of 1200 retinal fundus images from people with and without AMD.	https://ichallenges.grand-challenge.org/iChallenge-AMD/
10	RIADD (Retinal Image Analysis for multi-Disease Detection Challenge)	Contains 3,200 fundus images recorded using 3 different cameras and multiple conditions. The dataset has been classified into six categories based on six disorders or diseases: branch retinal vein occlusion, media haze, drusen, diabetic retinopathy, and age-related macular degeneration.	https://riadd.grand-challenge.org/download-all-classes/
11	JSIEC	(Joint Shantou International Eye Centre)-1000 fundus images which belong to 39 classes	https://www.kaggle.com/datasets/linchundan/fundusimage1000
12	DRIVE: Digital Retinal Images for Vessel Extraction	Designed to facilitate comparative research on the segmentation of blood vessels in retinal pictures. For the diagnosis, screening, treatment, and evaluation of various ophthalmologic diseases like diabetes, hypertension, arteriosclerosis, and choroidal neovascularization, retinal vessel segmentation, and delineation of morphological attributes of retinal blood vessels, such as length, width, tortuosity, branching patterns, and angles are used.	https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction

7. Performance Metrics for Classification and Segmentation techniques using DL

The performance of a model during training and testing is assessed using several metrics. Loss functions are separate from metrics. Loss functions show how well a model is performing. A deep learning model is trained using loss functions employing optimization methods like gradient descent. Classification models have discrete outputs, consequently, metrics that contrast discrete classes are necessary. Classification Metrics assess a model's performance and indicate whether the classification is accurate or not, but they each assess in a different manner [39]. Classification metrics are given below:

- Accuracy equals the proportion of correct forecasts to all predictions, multiplied by 100.
- Confusion Matrix: a tabular comparison of model predictions and ground-truth labels
- Precision is measured by the ratio of true positives to all anticipated positives.
- Recall is the proportion of true positives to all the ground truth's positives.
- The harmonic mean of precision and recall is the F1-score.

- AU-ROC (Area under Receiver operating characteristics curve) - AUC stands for the level or measurement of separability, and ROC is a probability curve. It reveals how well the model can differ across classes. The model is more accurate at classifying 0 classes as 0, and classifying 1 class as 1, the higher the AUC. TPR (True Positive Rate) and FPR (False Positive Rate) are represented on the y-axis and x-axis, respectively, of the ROC curve.

The common segmentation metrics are Pixel Accuracy and Intersection over Union (or IoU) [40].

Pixel Accuracy: Pixel accuracy is computed as given in equation 7.1

$$Pixel\,Accuracy = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN} \quad \text{-------------->7.1}$$

True Positive (TP): pixel classified appropriately as y
False Positive (FP): pixel classified inappropriately as y
True Negative (TN): pixel classified appropriately as not y
False Negative (FN): pixel classified appropriately as not y

Intersection over Union (IoU)

The formula in equation 7.2 is used to calculate the Intersection over Union (IoU) metric. It is also known as the Jaccard index, and it is basically a way to measure how much of the target mask and the forecast output overlap. The IoU metric calculates the fraction of pixels shared by the targeted and prediction masks that are also present in both masks.

IoU= (target ∩ prediction) / (target ∪ prediction) ----------> 7.2

Mean IoU (mIoU)

This measure depicts the average IoU across all courses. It is a reliable predictor of how well a segmentation model for images performs across all categories that the model might hope to identify.

8. CONCLUSION

This research work was carried out with the goal of extensive study of available deep learning architecture for fundus image classification and segmentation for identifying ocular diseases namely Glaucoma and Hypertensive Retinopathy. This work provides a summary of the existing solutions and compares them based on their performance. The work also provides a comprehensive list of publicly available image datasets for ocular diseases. The work provides significant evidence that the transfer learning-based approach is currently state-of-the-art in generating efficient models. Further, ensemble architectures combining various standard models can aid in generating efficient solutions.

9. FUTURE DIRECTIONS IN APPLICATION OF DL MODELS IN MEDICAL IMAGE CLASSIFICATION AND SEGMENTATION

One of the crucial aspects of applying deep learning models is reducing processing time and efficiency. Selective information processing is one of the options, which focuses on processing part of the scene or image. This can be implemented by simulating human vision systems and human attention principles on image processing algorithms. Explainability is also one of the significant directions in the application of DL models as the deep architectures essentially act as black boxes and it is important for justifying the solutions provided by these architectures for human acceptability of the solution [41]. Other notable directions in the application of DL models in medical image classification and segmentation are given below:

- **Hybrid Model Integration**

Integration of multiple models and creating a hybrid architecture for the purposes of classification followed by segmentation. Hybrid models can improve the speed, accuracy, and completeness of decision-making.

- **The Vision Transformer**

Commonly referred to as ViT, an image classification model developed by researchers at the University of Washington, it is used in sentiment analysis, object recognition, and image captioning. The application of Vision Transformers on Image classification and Segmentation tasks is gaining momentum because of the significant results established results in the literature.

- **Self-Supervised Learning**

Application of deep architectures requires huge amounts of labeled and annotated data. In the case of Medical Images, this costs the expert’s time and availability. Self-supervised learning in Deep architectures tries to avoid this by helping with automation. It learns to categorize the raw data automatically rather than depending on labeled data to train a system. Each input component can predict any other part of the input.

- **Incorporating Edge Intelligence (EI)**

Edge intelligence is transforming how data is acquired and analyzed. It shifts operations away beyond cloud-based systems for data storage and toward the edge. EI has increased decision-independence making from data storage devices by putting them nearer the data source.

REFERENCES

[1] X. He, Y. Deng, L. Fang and Q. Peng, "Multi-Modal Retinal Image Classification with Modality-Specific Attention Network," in IEEE Transactions on Medical Imaging, vol. 40, no. 6, pp. 1591-1602, June 2021, doi: 10.1109/TMI.2021.3059956.

[2] Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. 2, 420 (2021). https://doi.org/10.1007/s42979-021-00815-1

[3] ImageNet Large Scale Visual Recognition Challenge - https://dspace.mit.edu/handle/1721.1/104944#:~:text=The%20ImageNet%20Large%20Scale%20Visual,from%20more%20than%20fifty%20institutions.

[4] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2022). Image Segmentation Using Deep Learning: A Survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

[5] T. A. Soomro et al., "Deep Learning Models for Retinal Blood Vessels Segmentation: A Review," in IEEE Access, vol. 7, pp. 71696-71717, 2019, doi: 10.1109/ACCESS.2019.2920616.

[6] M.B. Sudhan, M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, Yosef Asrat Waji, "Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model", Journal of Healthcare Engineering, vol. 2022, Article ID 1601354, 10 pages, 2022. https://doi.org/10.1155/2022/1601354

[7] Bajwa, M.N., Malik, M.I., Siddiqui, S.A. et al. Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. BMC Med Inform Decis Mak 19, 136 (2019). https://doi.org/10.1186/s12911-019-0842-8

[8] 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, 2021, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.02.003

[9] Hemelings, R., Elen, B., Barbosa-Breda, J. et al. Deep learning on fundus images detects glaucoma beyond the optic disc. Sci Rep 11, 20313 (2021). https://doi.org/10.1038/s41598-021-99605-1

[10] Nawaz M, Nazir T, Javed A, Tariq U, Yong H-S, Khan MA, Cha J. An Efficient Deep Learning Approach to Automatic Glaucoma Detection Using Optic Disc and Optic Cup Localization. Sensors. 2022; 22(2):434. https://doi.org/10.3390/s22020434

[11] Latif, J., Tu, S., Xiao, C. et al. ODGNet: a deep learning model for automated optic disc localization and glaucoma classification using fundus images. SN Appl. Sci. 4, 98 (2022). https://doi.org/10.1007/s42452-022-04984-3

[12] Yuliang Ma, Zhenbin Zhu, Zhekang Dong, Tao Shen, Mingxu Sun, Wanzeng Kong, "Multichannel Retinal Blood Vessel Segmentation Based on the Combination of Matched Filter and U-Net Network", BioMed Research International, vol. 2021, Article ID 5561125, 18 pages, 2021. https://doi.org/10.1155/2021/5561125

[13] Y. Jiang et al., "JointRCNN: A Region-Based Convolutional Neural Network for Optic Disc and Cup Segmentation," in IEEE Transactions on Biomedical Engineering, vol. 67, no. 2, pp. 335-343, Feb. 2020, doi: 10.1109/TBME.2019.2913211.

[14] Abbas, Q.; Qureshi, I.; Ibrahim, M.E.A. An Automatic Detection and Classification System of Five Stages for Hypertensive Retinopathy Using Semantic and Instance Segmentation in DenseNet Architecture. Sensors 2021, 21, 6936. https://doi.org/10.3390/s21206936\

[15] Kanae Fukutsu, Michiyuki Saito, Kousuke Noda, Miyuki Murata, Satoru Kase, Ryosuke Shiba, Naoki Isogai, Yoshikazu Asano, Nagisa Hanawa, Mitsuru Dohke, Manabu Kase, Susumu Ishida, A Deep Learning Architecture for Vascular Area Measurement in Fundus Images, Ophthalmology Science, Volume 1, Issue 1, 2021, 100004, ISSN 2666-9145, https://doi.org/10.1016/j.xops.2021.100004. (https://www.sciencedirect.com/science/article/pii/S2666914521000026)

[16] Raza M, Naveed K, Akram A, Salem N, Afaq A, Madni HA, et al. (2021) DAVS-NET: Dense Aggregation Vessel Segmentation Network for retinal vasculature detection in fundus images. PLoS ONE 16(12): e0261698. https://doi.org/10.1371/journal.pone.0261698

[17] D. Nagpal, S. N. Panda and M. Malarvel, "Hypertensive Retinopathy Screening through Fundus Images-A Review," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 924-929, doi: 10.1109/ICICT50816.2021.9358746.

[18] Saba T, Akbar S, Kolivand H, Ali Bahaj S. Automatic detection of papilledema through fundus retinal images using deep learning. Microsc Res Tech. 2021 Dec;84(12):3066-3077. doi: 10.1002/jemt.23865. Epub 2021 Jul 8. PMID: 34236733.

[19] Morano J, Hervella ÁS, Novo J, Rouco J. Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images. Artif Intell Med. 2021 Aug;118:102116. doi: 10.1016/j.artmed.2021.102116. Epub 2021 May 29. PMID: 34412839.

[20] Abbas, Q.; Qureshi, I.; Ibrahim, M.E.A. An Automatic Detection and Classification System of Five Stages for Hypertensive Retinopathy Using Semantic and Instance Segmentation in DenseNet Architecture. Sensors 2021, 21, 6936. https://doi.org/10.3390/s21206936

[21] Krismono Triwijoyo, Bambang & Pradipto, Yosef. (2017). Detection of Hypertension Retinopathy Using Deep Learning and Boltzmann Machines. Journal of Physics: Conference Series. 801. 012039. 10.1088/1742-6596/801/1/012039.

[22] N. Deepa and D. T, "E-TLCNN Classification using DenseNet on Various Features of Hypertensive Retinopathy (HR) for Predicting the Accuracy," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1648-1652, doi: 10.1109/ICICCS51141.2021.9432255.

[23] Şüyun, Süleyman & Tasdemir, Sakir & Biliş, Serkan & Milea, Alexandru. (2021). Using a Deep Learning System That Classifies Hypertensive Retinopathy Based on the Fundus Images of Patients of Wide Age. Traitement du Signal. 38. 207-213. 10.18280/ts.380122.

[24] Hoque, Mohammed & Kipli, Kuryati & Zulcaffle, Tengku & Al-Hababi, Saleh & D.A.A, Mat & Sapawi, Rohana & Joseph, Annie. (2021). A Deep Learning Approach for Retinal Image Feature Extraction. Pertanika Journal of Science and Technology. 29. 10.47836/pjst.29.4.17.

[25] Yadav, S., Das, S., Murugan, R. et al. Performance analysis of deep neural networks through transfer learning in retinal detachment diagnosis using fundus images. Sādhanā 47, 49 (2022). https://doi.org/10.1007/s12046-022-01822-5

[26] Kaul, Chaitanya et al. "Focusnet++: Attentive Aggregated Transformations for Efficient and Accurate Medical Image Segmentation." 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (2021): 1042-1046.

[27] Oliveira, Guilherme & Rosa, Gustavo & Pedronette, Daniel & Papa, João & Kumar, Himeesh & Passos Júnior, Leandro & Kumar, Dinesh. (2022). Which Generative Adversarial Network Yields High-Quality Synthetic Medical Images: Investigation Using AMD Image Datasets.

[28] Abramovich, Or & Pizem, Hadas & Van Eijgen, Jan & Stalmans, Ingeborg & Blumenthal, Eytan & Behar, Joachim. (2022). FundusQ-Net: a Regression Quality Assessment Deep Learning Algorithm for Fundus Images Quality Grading.

[29] Gong D, Kras A, Miller JB. Application of Deep Learning for Diagnosing, Classifying, and Treating Age-Related Macular Degeneration. Semin Ophthalmol. 2021 May 19;36(4):198-204. doi: 10.1080/08820538.2021.1889617

[30] Pham, Q.T.M.; Ahn, S.; Song, S.J.; Shin, J. Automatic Drusen Segmentation for Age-Related Macular Degeneration in Fundus Images Using Deep Learning. Electronics 2020, 9, 1617. https://doi.org/10.3390/electronics9101617

[31] Tharindu De Silva, Emily Y. Chew, Nathan Hotaling, and Catherine A. Cukras, "Deep-learning based multi-modal retinal image registration for the longitudinal analysis of patients with age-related macular degeneration," Biomed. Opt. Express 12, 619-636 (2021)

[32] Chen, Yao-Mei & Huang, Wei Tai & Ho, Wen-Hsien & Tsai, Jinn-Tsong. (2021). Classification of age-related macular degeneration using convolutional-neural-network-based transfer learning. BMC Bioinformatics. 22. 10.1186/s12859-021-04001-1.

[33] Heo TY, Kim KM, Min HK, Gu SM, Kim JH, Yun J, Min JK. Development of a Deep-Learning-Based Artificial Intelligence Tool for Differential Diagnosis between Dry and Neovascular Age-Related Macular Degeneration. Diagnostics (Basel). 2020 Apr 28;10(5):261. doi: 10.3390/diagnostics10050261

[34] Smitha, A., Jidesh, P. Classification of Multiple Retinal Disorders from Enhanced Fundus Images Using Semi-supervised GAN. SN COMPUT. SCI. 3, 59 (2022). https://doi.org/10.1007/s42979-021-00945-6

[35] Y. Zong et al., "U-net Based Method for Automatic Hard Exudates Segmentation in Fundus Images Using Inception Module and Residual Connection," in IEEE Access, vol. 8, pp. 167225-167235, 2020, doi: 10.1109/ACCESS.2020.3023273.

[36] Md. Badiuzzaman Shuvo, Rifat Ahommed, Sakib Reza, M.M.A. Hashem, CNL-UNet: A novel lightweight deep learning architecture for multimodal biomedical image segmentation with false output suppression,Biomedical Signal Processing and Control,Volume 70,2021,102959, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.102959.

[37] Menghani, G. (2021). Efficient deep learning: A survey on making deep learning models smaller, faster, and better. arXiv preprint arXiv:2106.08962.

[38] Wang, H.; Yang, J.; Wu, Y.; Du, W.; Fong, S.; Duan, Y.; Yao, X.; Zhou, X.; Li, Q.; Lin, C.; Liu, J.; Huang, L.; Wu, F. A Fast Lightweight Based Deep Fusion Learning for Detecting Macula Fovea Using Ultra-Widefield Fundus Images. Preprints 2021, 2021080469 (doi: 10.20944/preprints202108.0469.v2).

[39] Du, Xiaogang & Nie, Yinyin & Wang, Fuhai & Lei, Tao & Wang, Song & Xuejun, Zhang. (2022). AL-Net: Asymmetric Lightweight Network for Medical Image Segmentation. Frontiers in Signal Processing. 2. 842925. 10.3389/frsip.2022.842925.

[40] R. M. T. Yanni, N. E.-K. El-Ghitany, K. Amer, A. Riad, and H. El-Bakry, "A New Model for Image Segmentation Based on Deep Learning", Int. J. Onl. Eng., vol. 17, no. 07, pp. pp. 28–47, Jul. 2021.

[41] Saleem, Muhammad Asif & Senan, Norhalina & Wahid, Fazli & Aamir, Muhammad & Samad, Ali & Khan, Mukhtaj. (2022). Comparative Analysis of Recent Architecture of Convolutional Neural Network. Mathematical Problems in Engineering. 2022. 10.1155/2022/7313612.