

Table of Contents

Introduction	2
Literature Survey	3
Method	7
Vessel Segmentation	7
Datasets	8
Automated Retinal Image Analysis (ARIA) dataset	8
AFIO dataset	8
High Resolution Fundus (HRF) dataset	8
DualModal2019 dataset	8
STructured Analysis of the REtina (STARE) dataset	8
Digital Retinal Images for Vessel Extraction (DRIVE) dataset	9
CHASE-DB1 dataset	9
Preprocessing	10
Model and Training	10
Boundary Enhancement	11
Feature Denoising	12
Training	12
Artery-Vein Classification	13
Datasets	13
DRIVE Artery Vein dataset - Retinal Images vessel Tree Extraction (RITE) dataset	13
HRF Artery Vein dataset	13
Preprocessing	14
Model and Training	15
Vessel Attention Map	15
Training	16
Post Processing	17
Segment Label Unification	17
Evaluation Metrics	18
Results	18
Discussion	21
Conclusion	21
References:	22
Appendix	24

Vessel Segmentation and Artery-Vein Classification in Retinal Fundus Images

Introduction

Retina is a multi-layered tissue of light-sensitive cells that is present in the posterior cavity of the eye and it is responsible for creating neural impulses that are interpreted by the brain for visual perception. Retinal fundus imaging is a non-invasive imaging technique that is used to visualize the retina. A retinal fundus image is usually a RGB image of the backside of the eye that shows the various components of the retina such as the optic disk, fovea and the macula along with the retinal vasculature - an intricate network of arteries and veins that supply blood to the retina. Fundus images are commonly used to scrutinize the retina and check for any sort of anomalies in it or in its vasculature.

The retinal vasculature are generally monitored for abnormalities that can act as early symptoms of several retinal and ocular diseases. For instance, microenvironmental factors simulated by ocular diseases like age-related macular degeneration and diabetic retinopathy can cause atypical neovascular growths in the capillaries of the inner retinal surface [1]. Furthermore, retinal fundus imaging is one of the very few techniques that is both non-invasive and that allows the visualization of blood vessels inside the body. The distribution of blood vessels viewed through the fundus images can reflect the whole body's vasculature and thus the usability of retinal fundus images can be extended well beyond ophthalmology. It can provide valuable insights into the patient's overall vascular health and act as an indicator for multiple systemic, neurological, cerebrovascular and cardiovascular morbidities and risks. Studies have proven that chronic diseases like diabetes and hypertension can cause unusual dilation and tortuousness in retinal venules. Wong et al on their research have concluded that reduced arteriole-to-venule diameter ratios measured from retinal images can be correlated to the occurrence and development of hypertension [2]. It has also been observed that narrowed arterioles, dilated venules or a combination of both can be associated with increased risk for stroke events and coronary heart diseases [3].

Studies have made it fairly evident that through accurate analysis and evaluation, it is possible to obtain vital diagnostic information from the retinal vasculature. However, segmenting vessels from retinal images and classifying them into arteries and veins manually can be a highly time-consuming and laborious task with low efficiency. Additionally, manual segmentation and classification also means subjectivity and thus variability in quality across annotators. The objective of this work is to develop a deep learning based automatic vessel segmentation and artery vein classification model for retinal fundus images. This model is expected to greatly benefit the current clinical setting and also act as an aid for risk factor estimation or detection of mainly ocular diseases and additionally other ailments discussed earlier like hypertension, diabetes and vascular diseases.

Figure 1(a) shows a sample retinal fundus image. In 1(b) is the vessel segmentation that corresponds to the retinal image in 1(a) and in 1(c) is its artery vein classification. The vessel pixels in red represent the arteries and the vessel pixels in blue belong to the veins.

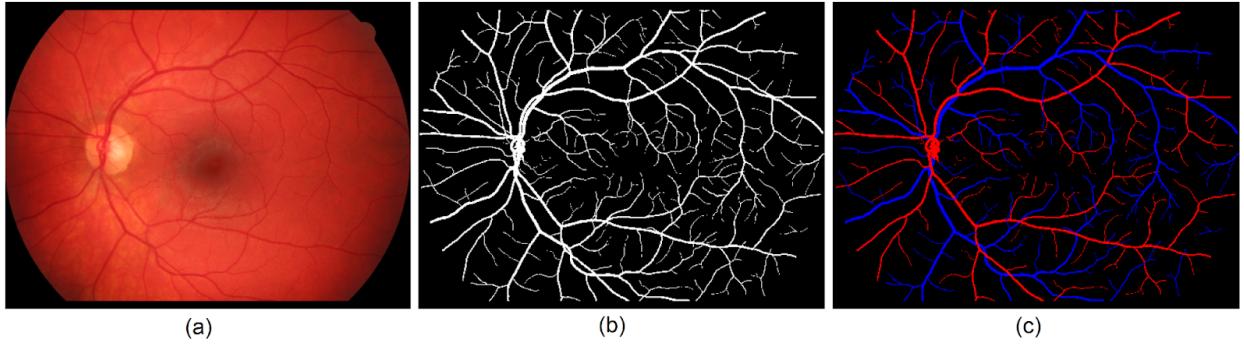


Figure 1: A sample retinal fundus image from the HRF dataset [28].

(a) Raw fundus image. **(b)** Vessel segmentation of the retinal fundus image. **(c)** Vessels classified into arteries (red) and veins (blue).

Literature Survey

Automated vessel segmentation is an active field of research and there are numerous algorithms that have been proposed. In the past years, machine learning techniques are prominently used for the development of automatic vessel segmentation models. These models can be broadly categorized into unsupervised and supervised learning methods.

Unsupervised learning techniques focus on intrinsic properties that can help separate the vessel pixels from the background. Nguyen et al. proposed an unsupervised learning algorithm that used line detectors at different scales to segment the retinal vasculature out from retinal fundus images [4]. The authors claim that their technique is time efficient and can handle high resolution retinal images. They were able to attain an average accuracy of 94.07% and 93.24% respectively on the public datasets DRIVE and STARE. Clustering is another unsupervised classification algorithm and it involves grouping similar kinds of objects or, as in this case, pixels into a ‘cluster’. Yavuz et al. have proposed a vessel segmentation technique based on this algorithm [5]. They use Gabor, Gaussian and Frangi filters for vessel enhancement and top-hat transform on the filter outputs to further highlight the vessel pixels. They then experiment with two clustering algorithms: K-means and Fuzzy C-means clustering on the vessel enhanced images to separate the vessel pixels from the background. They were able to achieve an average accuracy of 95.94% and 95.71% on the datasets DRIVE and STARE respectively.

Supervised learning techniques involve a training phase where the models are made to learn how to differentiate between vessel and non-vessel pixels based on ground truth retinal images that are precisely labelled. The performance of these models are generally better than that of unsupervised models as they can benefit from the manually labelled ground truths used during the training. Dastgheib et al. in their work propose an unsupervised vessel segmentation model that uses multi-scale spatial filters to locate blood vessels in a retinal fundus image [6]. They got an average accuracy of 94.28% on the DRIVE dataset. They then, on top of their unsupervised model, built a support vector machine (SVM) classifier, a supervised classification model. The SVM classifier was trained on the features extracted from the outputs of the multi-scale spatial filters. This improved the average accuracy on the DRIVE dataset to 94.48%. Another supervised retinal vasculature segmentation model was proposed by Zhu et al. and it used an

extreme learning machine (ELM) [7]. For each pixel in the fundus image, a 39-dimensional feature vector was created using local features, morphological features, phase congruency etc. and this acted as input to the ELM classifier. The output of the classifier was then collected to create a binary vessel segmentation image. This model gave an average accuracy of 96.07% on the DRIVE dataset.

The performance results of the supervised and unsupervised models discussed so far for vessel segmentation have been summarized in table 1. There are numerous other retinal vessel segmentation models and for further reference Akbar et al. have compiled a list of major machine learning based automated techniques for blood vessel segmentation in [8].

Table 1: Performance summary of the supervised and unsupervised vessel segmentation models discussed.

Method	Datasets	Feature Extraction Technique	Classifier	Accuracy	Specificity	Sensitivity
<i>Unsupervised Methods:</i>						
Nguyen et al. [4]	DRIVE STARE	Multi-scale line detectors	-	0.9407 0.9324	-	-
Yavuz et al. [5]	DRIVE STARE	Gabor Filter, Gaussian-Matched filter, Frangi filter, top-hat transform	K-Means and Fuzzy C-Means clustering	0.9570 0.9594	0.9905 0.9816	0.6102 0.6869
<i>Supervised Methods:</i>						
Dastgheib et al. [6]	DRIVE	Multi-scale Spatial Filters	Support Vector Machine (SVM) classifier	0.9448	-	-
Zhu et al. [7]	DRIVE	Local features, morphological features, phase congruency, Hessian and divergence of vector fields	Extreme Learning Machine (ELM) classifier	0.9607	0.9868	0.7140

Vessel segmentation using supervised models usually contains two stages: feature extraction and pixel classification. The performance of these models depends a great deal on choosing the appropriate set of features and it demands excellent domain knowledge. Furthermore, even with the right set of features, the amount of information that is fed into the model is restricted which in turn might limit the generalization capability of the model [9]. Deep learning models are adopted in order to overcome these disadvantages. Deep learning methods have the ability to automatically learn features from training data without much human interference. They are also said to have better generalization and recognition abilities as they learn from patterns at different levels and scales of the input and try to utilize most of the useful information available in the input [9].

Initially convolutional neural networks (CNNs) were a popular choice out of the different deep learning models for retinal vessel segmentation. Fan et al. proposed a 5-layered CNN for vessel segmentation and compared the use of red, green and blue channels from the RGB retinal fundus image as input to the network [10]. They concluded that the green channel provided the best vessel-background contrast and were able to get an average accuracy of 96.12%, 96.54% and 95.73% respectively on the public datasets DRIVE, STARE and CHASE-DB1. Later, studies showed that fully convolutional networks (FCNs) and U-Nets [11] were more leveraged to make dense predictions and thus were better suited for segmentation tasks. Numerous U-Net based models with different modifications and improvements to support performance gain on the vessel segmentation task were designed. Alom et al. proposed a deep learning model called the recurrent residual convolutional neural network (R2U-Net) by combining the features of U-Nets, residual networks and recurrent convolutional neural networks for semantic segmentation of medical images [12]. The model was tested on three segmentation tasks and one of it was blood vessel segmentation in retinal images. It achieved an average accuracy of 95.56%, 96.34% and 97.12% respectively on the DRIVE, CHASE-DB1 and STARE datasets. Wu et al. introduce an inception-residual convolution block for improved feature representation that will help in more accurate segmentation of the fine capillaries in fundus images. They embedded these convolution blocks within the encoder-decoder architecture of the U-Net and introduced a model called the Vessel-Net for accurate retinal vessel segmentation [13]. Vessel-Net was able to give an average accuracy of 95.78% and 96.61% on DRIVE and CHASE-DB1 datasets respectively. Recently, the use of generative adversarial networks (GANs) for vessel segmentation have also been proposed as in [14] where Kamran et al. introduced RV-GAN for retinal vessel segmentation. They obtained an average accuracy of 97.90%, 96.97% and 97.54% on the DRIVE, CHASE-DB1 and STARE datasets respectively. Although the results produced by GANs are good, when compared to the training of other deep learning models discussed, there is a need to train the generators and the discriminators alternatively and that can be unstable or troublesome sometimes [9]. Table 2 summarizes the performance of some deep learning models that have been proposed so far for vessel segmentation. For further reference, Chen et al. in [9] have listed and compared different deep learning models available for retinal vessel segmentation.

Table 2: Performance summary of some deep learning models for retinal vessel segmentation.

Model	Datasets	Accuracy	Specificity	Sensitivity
<i>CNN-based Models:</i>				
Fan et al. [10]	DRIVE, STARE, CHASE-DB1	0.9612, 0.9654, 0.9573	0.9788, 0.9799, 0.9702	0.7814, 0.7834, 0.7661
<i>U-Net based Models:</i>				
R2U-Net Alom et al. [12]	DRIVE, STARE, CHASE-DB1	0.9556, 0.9712, 0.9634	0.9813, 0.9862, 0.9820	0.7792, 0.8298, 0.7756
Vessel-Net Wu et al. [13]	DRIVE, CHASE-DB1	0.9578, 0.9661	0.9802, 0.9814	0.8038. 0.8132

DEU-Net Wang et al. [22]	DRIVE, CHASE-DB1	0.9567, 0.9661	0.9816, 0.9821	0.7940, 0.8074
RVSeg-Net Wang et al. [23]	DRIVE, CHASE-DB1	0.9681 ,0.9726	0.9845, 0.9836	0.8107, 0.8069
LadderNet Zhuang et al.[24]	DRIVE, CHASE-DB1	0.9561, 0.9656	0.9810, 0.9818	0.7856, 0.7978
IterNet Li et al. [25]	DRIVE, STARE, CHASE-DB1	0.9573, 0.9701, 0.9655	0.9838, 0.9886, 0.9823	0.7735, 0.7715, 0.7970
BEFD-Net Zhang et al. [21]	DRIVE	0.9701	0.9845	0.8215
<i>GAN-based Models:</i>				
RV-GAN Kamran et al. [14]	DRIVE, STARE, CHASE-DB1	0.9790, 0.9754, 0.9697	0.9969, 0.9864, 0.9806	0.7927, 0.8356, 0.8199

Moving on to automated artery-vein classification, there are graph-based and feature-based algorithms that are available [15]. Joshi et al. proposed a method where from a vessel segmentation image individual vessel trees are identified through graph search. Each individual vessel tree is then labelled as either an artery or a vein based on their color properties [16]. This method yielded an accuracy of 91.44% on a dataset of 50 fundus images. Graph-based methods make use of the natural shape and structure of the retinal vasculature and are considered to be logic-based simpler algorithms for artery-vein classification. However, sometimes during vessel segmentation there are chances for only part of a vessel to be detected, especially during the segmentation of fine capillaries. On that aspect, one of the major downsides of graph-based algorithms is that these vascular regions that are stranded cannot be linked with any of the vessel trees with certainty. Feature-based algorithms for artery-vein classification overcame this drawback. Zamperini et al. in [17] have focused on finding optimal features for artery-vein discrimination in fundus images. From a set of 86 features, they chose the 16 most prominent features by using the greedy backward feature selection algorithm. Their results showed that features that relate to color, contrast inside and outside the vessel region and positional information are most relevant for artery-vein separation. They also prove that with the use of these good descriptors it is possible to achieve good classification results. As for a classification model, Niemeijer et al. proposed a feature-based supervised model for artery-vein classification [18]. They initially extracted 24 features based on intensity values and derivative information for each centerline pixel in the vessel segmentation. They then reduced the dimensionality of the feature space by selecting the 12 most prominent features based on the sequential forward floating selection (SFFS) algorithm. Different classifiers like linear and quadratic discriminant analysis (LDA and QDA), SVM and K-nearest neighbours were trained based on this set of 12 features to assign a label for each of the centerline pixels. These labels are then processed to label all the vessel pixels in the segmentation image. They were able to obtain an AUC (area under the ROC (receiver operating characteristic) curve) of 0.88 on the DRIVE test dataset using their algorithm. Although the results seem satisfactory, the problem with both graph and feature based algorithms for artery-vein classification is that they need a

vessel segmentation to start with and in order to get good results the accuracy of the vessel segmentation algorithm needs to be high.

Deep learning models for artery-vein classification have been proposed to overcome the disadvantages discussed above. Following the successful application of U-Nets for vessel segmentation, Hemelings et al. proposed the use of these models for artery-vein segmentation [19]. A 5-layered U-Net model with 16 channels was designed to take as input the RGB fundus image and give as output a segmentation map with four labels - background, artery, vein and unknown (the authors marked the pixels with unclear categorization as unknown during preprocessing of the datasets). They were able to obtain an accuracy of 93.94% on the DRIVE dataset and 97.28% on the HRF dataset. It is to be noted that this approach removes the previously stated disadvantage of needing an accurate vessel segmentation model. As another approach, Hu et al. proposed a deep learning model called VC-Net that produces both vessel segmentation and artery-vein classification outputs [20]. The vessel detection part of the model is an U-Net embedded with multi-scale feature extraction modules. The output from the vessel segmentation module is used to calculate a vessel map. This vessel map is then passed on to the artery-vein classification module as an aid to enhance the vessel pixels and to improve classification accuracy. The VC-Net when tested on the DRIVE dataset gave an accuracy of 95.54%.

Based on all the previous works that were referred to and their results, in this work, initially a U-Net based model called the boundary enhancement and feature denoising network (BEFD-Net) [21] was used for vessel segmentation. It was trained to give a vessel segmentation output by taking as input the RGB fundus image. For artery-vein classification, another modified U-Net module was trained to classify the input pixels into one of three categories - artery, vein or background by taking as input the RGB fundus image and a vessel map calculated using the corresponding vessel segmentation module output. More detailed descriptions of the modules along with the details of the various retinal fundus datasets that were used for the training and testing of the models are provided in the next section of this report.

Method

This section is divided into two parts. The first part deals with the vessel segmentation task and the second part focuses on the artery-vein classification task.

Vessel Segmentation

As already stated, the aim of the vessel segmentation model is to take a retinal fundus image like the one shown in figure 1(a) as input and produce as output a binary image similar to the one shown in figure 1(b) where the pixels that correspond to blood vessels in the input image are marked true and the other pixels that belong to the background are marked false.

Datasets

Below are the datasets that were used during training, validation and testing of the vessel segmentation model. The datasets contain retinal fundus images and ground truths where the blood vessels were manually labelled by human experts.

Automated Retinal Image Analysis (ARIA) dataset

This dataset was created by Damian Farnell [26] in collaboration with the ophthalmic imagers who worked at St. Paul's Eye Unit, UK. This dataset contains 143 fundus images each of size 768×576 and these were collected from adult test subjects who were either healthy or were affected by age-related macular degeneration (AMD) or diabetes. The dataset comes with two ground truths for each of the 143 fundus images with the vessel traces given by two expert retinal image graders at St. Paul's Eye Unit.

AFIO dataset

This dataset contains 100 retinal fundus images obtained from Armed Forces Institute of Ophthalmology (AFIO), Pakistan. The resolution of all the images is 1504×1000 and out of the 100 fundus images, 86 were collected from subjects with ocular diseases and the remaining 14 were collected from subjects with healthy or normal retinas. The dataset was annotated by four experts from AFIO and these annotations contain ground truths for both vessel segmentation and artery-vein classification. More details on the dataset can be found in [27].

High Resolution Fundus (HRF) dataset

This dataset contains 45 retinal fundus images out of which 15 belong to healthy patients, another 15 belong to patients with diabetic retinopathy and the remaining 15 images belong to glaucomatous patients. Each image is of resolution 3504×2336 and come with a gold standard vessel segmentation annotated by a group of experts working in the field of retinal image analysis. More information can be found in [28, 29].

DualModal2019 dataset

This is a set of 30 retinal images with resolution 1024×1024 . The dataset comes with manually annotated ground truths of arteries and veins making it suitable for both vessel segmentation and artery-vein classification tasks. More details on the dataset can be found in [30].

STructured Analysis of the REtina (STARE) dataset

The STARE is a project initiated by the University of California and funded by the U.S. National Institutes of Health. The dataset contains about 400 retinal images provided by the Shiley Eye Center at the University of California and by the Veterans Administration Medical Center in San Diego [31]. Out of all the images in the dataset, 20 images of resolution 700×605 were annotated by Hoover et al. in order to be used in their work aimed at developing a vessel segmentation model [32] and these 20 images along with their vessel segmentation ground truths are used in this work.

Digital Retinal Images for Vessel Extraction (DRIVE) dataset

This is a dataset with 40 images that is split into 20 training and 20 testing images. The images were acquired during a diabetic retinopathy screening program in the Netherlands. 33 of the 40 images are from healthy patients and the remaining 7 images (3 training and 4 testing) are of retinas that show symptoms of diabetic retinopathy. The resolution of the images is 768×584 and the 20 images in the training set have one manual vessel segmentation provided by an expert whereas the 20 images in the test set have two manual vessel segmentations provided by an expert and a human observer. More information on the dataset can be found in [33].

CHASE-DB1 dataset

This dataset is part of the Child Heart and Health Study in England (CHASE), a cardiovascular health survey spanning across 200 primary schools. There are 28 optic disc centered retinal images of size 999×960 collected from school children and the dataset comes with two vessel segmentations given by two expert annotators for each retinal image.

Table 3: Datasets

Used As	Dataset	Image Size	Granularity	Number of Images	Labels	Number of Annotators
Training/Validation	ARIA	768×576	Medium	143	Vessels	2
Training/Validation	AFIO	1504×1000	Medium	100	Artery/Vein	1
Training/Validation	HRF	3504×2336	High	45	Vessels	1
Training/Validation	DualModal2019	1024×1024	Low	30	Artery/Vein	1
Training/Validation	STARE	700×605	Medium	20	Vessels	1
Test	DRIVE	768×584	High	40	Vessels	1 for 20 images 2 for 20 images
Test	CHASE-DB1	999×960	Low	28	Vessels	2

Table 3 lists all the datasets along with some associated details. The property granularity represents the extent to which the fine capillaries are included by the annotators of the datasets in the ground truth vessel segmentation. Sample fundus images from each dataset and their ground truth vessel segmentations are provided in the appendix and the difference in granularity and annotation styles across the datasets can be observed there.

Usually training and test sets for the models are created using images belonging to one dataset. However, this work is also aimed at testing the cross dataset generalization capability of the vessel segmentation model. Therefore, as indicated in table 3, ARIA, AFIO, HRF, DualModal2019 and STARE datasets were used for model training and validation while DRIVE and CHASE-DB1 datasets were used for testing. This results in 68 test images and 338 images to split between training and validation sets. 65 random images were chosen from the 338 images to form the validation dataset and the remaining 273 images formed the training dataset for the vessel segmentation model.

Preprocessing

As it can be seen from table 3, the fundus images from different datasets are of different resolutions. Thus as a preprocessing step, all the fundus images and their ground truths were resized to a resolution of 512×512 . The ground truths are binary images and thus for the resized ground truths to also be binary images, nearest neighbor interpolation is used for their resizing.

Model and Training

One of the challenges in vessel segmentation is to detect the fine capillaries while maintaining a decent precision score by controlling the number of false positives. This is important because when aiming to detect the fine capillaries, as there are some underlying similarities, there is always a risk of wrongly focusing on the noise in the input image. With input resizing involved, the loss in information adds to the risk of increased focus on the noise. The boundary enhancement and feature denoising network (BEFD-Net) proposed by Zhang et al. in [21] majorly addresses the problem discussed above. Figure 2 shows the architecture of a BEFD-Net. It can be seen that it is an U-Net with added boundary enhancement (BE) and feature denoising (FD) functions. The boundary enhancement uses an edge attention map to increase the model's focus on the vessel edges and in turn increases the model's capability to detect fine blood capillaries. The feature denoising feature is introduced to reduce the focus on the noise in the input image that might have been wrongly highlighted by the boundary enhancement layer. The BE feature increases the recall score of the model by reducing the number of false negatives while the FD feature is expected to keep the precision score from dropping by keeping the number of false positives under check.

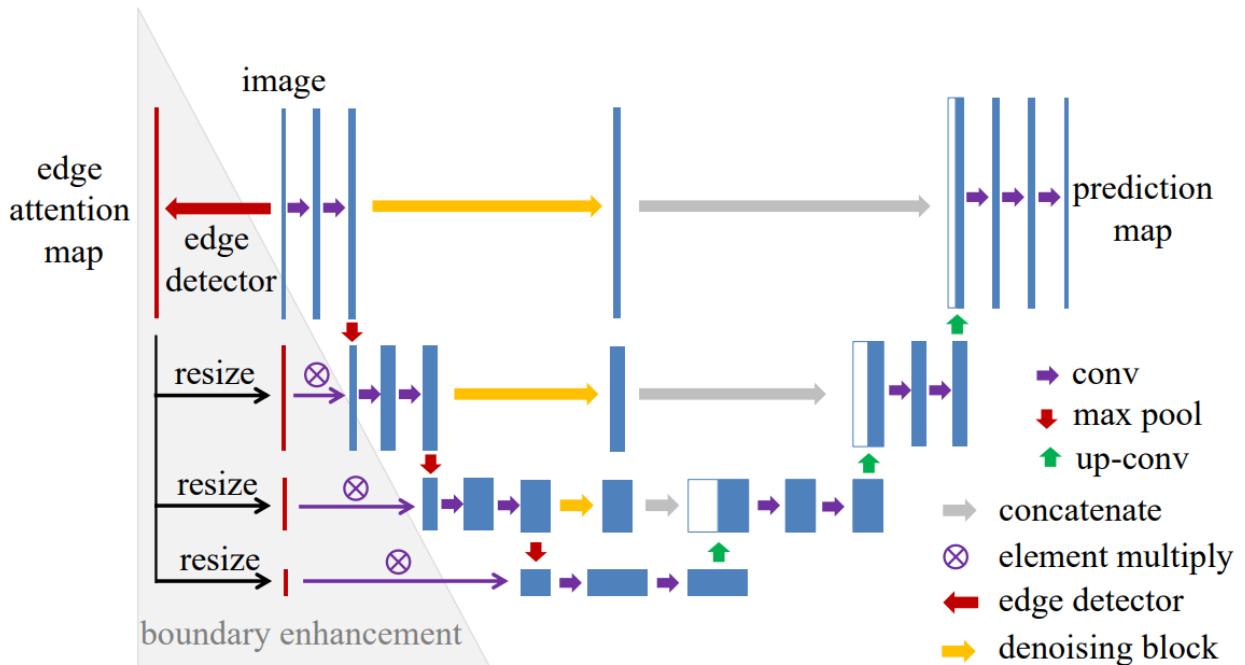


Figure 2: Architecture of BEFD-Net. Image sourced from [21]

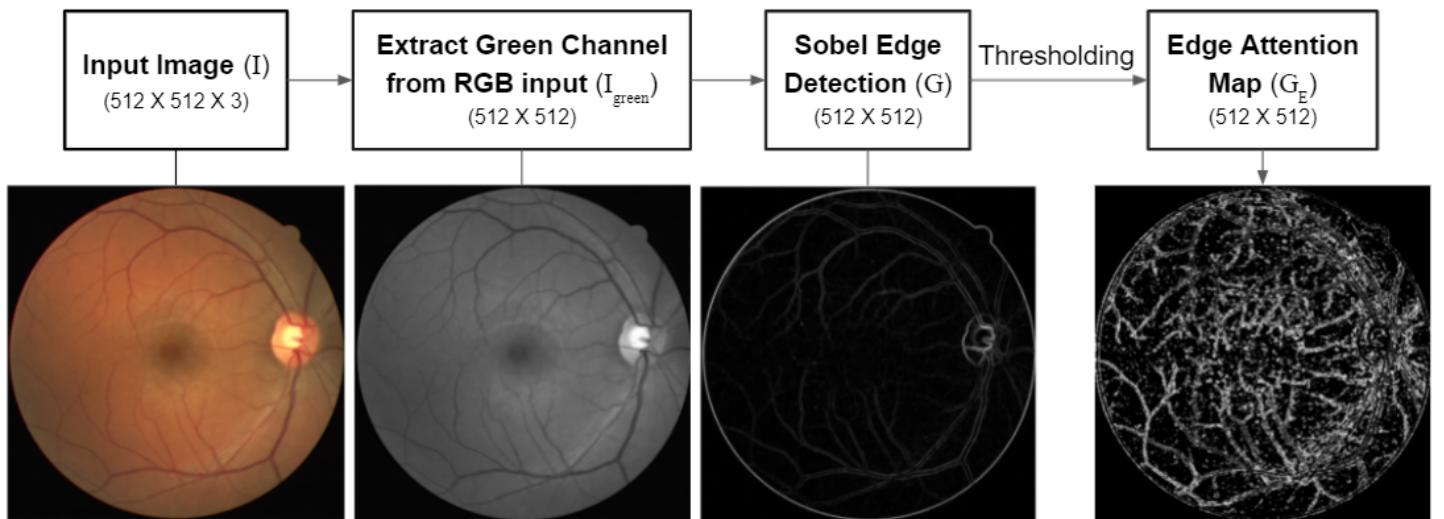
Boundary Enhancement

Figure 3 shows the steps involved in the computation of the edge attention map for the boundary enhancement feature of the BEFD-Net model. The sobel operators need a single channel input. After some experimentation, it was found that using the green channel of the RGB fundus input for edge detection yielded good results. The next step was to use the sobel operators as shown in equation (1) to obtain the edges. I_{green} in equation (1) represents the green channel of the input fundus image (I), $*$ represents the convolution operator, G_x and G_y denote the filters to compute the gradients along the x and y axis respectively and finally G is the Sobel edge map.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I_{green}, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I_{green}, G = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$G_E(x, y) = \begin{cases} 1, & \text{if } G(x, y) > \lambda_{max} \text{ or } < \lambda_{min} \\ (1 - \frac{G(x, y) - \lambda_{min}}{\lambda_{max} - \lambda_{min}}) \cdot \alpha + \beta, & \text{otherwise} \end{cases} \quad (2)$$

The Sobel edge map, G is then thresholded as per equation (2) to get the edge attention map G_E . $G(x, y)$ denotes the pixel value at (x, y) in the Sobel edge map. α , β , λ_{max} and λ_{min} are thresholding parameters. α was set to 2 and β was set to 1 as done in [21]. With some experimentation it was observed that it was better if the values for λ_{max} and λ_{min} were calculated



from the Sobel edge map and thus λ_{max} was set to $[0.9 * Maximum(G)]$ and λ_{min} to $[0.5 * \lambda_{max}]$.

Figure 3: Steps to calculate the edge attention map

The thresholding operation ends up setting all the non-edge pixels and strong-edge pixels to one in the edge attention map. It highlights the weak-edge pixels by assigning higher weights to

them in the edge attention map and this brings the micro vessels with low contrast into focus. However, as it could be gathered from the edge attention map shown in figure 3, the noise in the input image is also emphasized along with the weak-edge pixels. The denoising block is designed to deal with this problem.

Feature Denoising

Xie et al. in [34] proposed a method to improve the adversarial robustness of an image classification model. Based on the hypothesis adversarial perturbation induces noise into the features extracted by the classification model, they proposed a non-local means based feature denoising block and proved that by adding these blocks into the model and training it the model's adversarial robustness can be improved. Figure 4 shows how this feature denoising block works for an input x of size $H \times W$ with C channels. The grey part illustrates the calculation of the non-local means, the denoising operation. The denoising block is then formed by wrapping the denoising operation with a 1×1 convolution and an identity skip connection. This denoising block is included into the BEFD-Net architecture as shown in figure 2 to help deal with the noise in the feature maps.

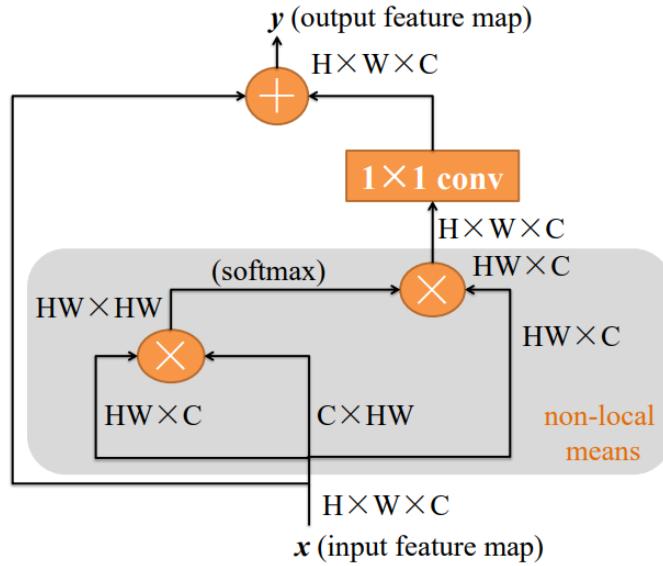


Figure 4: The feature denoising block used in BEFD-Net. Image sourced from [21]

Training

The BEFD-Net trained for vessel segmentation was 4 layers deep with 64 channels and kernel size 3. This implies that all the convolutions in the encoding and decoding path of the U-Net used 3×3 kernels and the number of filters started with 64 for the first layer and doubled for every layer down. A dropout of rate 0.1 was set for regularization and batch normalization after each convolution layer was used for increased stability during training. The output layer at the end of the decoding path of the model was a single filter 1×1 convolution layer with sigmoid activation.

The model was trained using adam optimizer with a learning rate of 0.001 and on a loss that was the sum of binary cross entropy and dice loss. The batch size was set to 5 and the model was trained for 100 epochs on the 273 training images with data augmentations that included

flipping, blurring, scaling, shifting, rotating and randomly changing brightness and contrast of the input images. At the end of each epoch, the loss and accuracy on the validation set was calculated. If there was no decrease in the validation loss for 10 consecutive epochs, the learning rate was decreased by a factor of 10. Finally, the model that gave the best binary accuracy on the validation data was saved as the final model for vessel segmentation. The performance of this model will be summarized in the results section of this report.

Artery-Vein Classification

The aim of the artery-vein classification task is to categorize the pixels in the retinal fundus images into one of the three classes - arteries, veins and background as shown in figure 1(c). A U-Net based deep learning model was trained for this purpose. Below are the details on the datasets used along with specifications on the model and its training.

Datasets

The datasets AFIO and DualModal2019 already come with ground truths for artery vein classification. In addition to these two datasets, we were able to find artery vein classifications for DRIVE and HRF datasets. Below are the details on the datasets that might be considered relevant.

DRIVE Artery Vein dataset - Retinal Images vessel Tree Extraction (RITE) dataset

Hu et al. for their work in [35] created the RITE (retinal images vessel tree extraction) dataset with standard artery vein ground truths for the 40 fundus images in the DRIVE dataset. The ground truths were RGB images where the arteries were marked in red, the veins in blue, the areas where the arteries and veins overlapped were marked in green and the areas of uncertainty where the label for the vessel was entirely clear were marked in white. More details on the dataset can be found in [36].

HRF Artery Vein dataset

Hemelings et al., to use for their FCN based artery vein classification model [19], manually separated the vessel segmentations of the 45 retinal fundus in the HRF to generate artery vein ground truths. These ground truths were again RGB images where the arteries were indicated by red pixels, the veins by blue pixels and the areas of uncertainty where the label for the vessel was entirely clear were indicated by green pixels. More details on the dataset and the labelling procedure can be found in [19, 37].

Splitting available data for model training, validation and testing

As explained earlier for the vessel segmentation task, the aim is to develop a model that can generalize well outside the datasets that were used for its training and validation. Thus, initially, the DRIVE dataset was set aside for testing and the remaining three datasets AFIO, DualModal2019 and HRF were used for model training and validation. That left 40 images in the test set. As for the training and validation sets, from the available 175 images, random 25 were chosen for validation and the remaining 150 became training images. However, the models that were trained based on this data split were not able to perform well when tested on the test set.

The veins and the arteries are usually identified with the help of the difference in their colour. As the veins carry deoxygenated blood, they have a colour that is darker than that of the arteries on the fundus images. But, it is common to be able to see this difference in the colour of the vessels only in the region close to their origin at the optic disc. As we move away from the optic disk, it becomes difficult to differentiate between arteries and veins. Experts generally trace the path of the vessels to identify and categorize the blood capillaries that are away from the optic disk. Therefore, artery vein classification is a complicated task. By using different retinal image datasets, the tasks becomes more challenging because of the differences in the annotation styles, variations in the granularity of the ground truths, differences in the quality of the images and the information losses due to the need to resize the images to bring them to a uniform size. It could be observed from figure 5 that additionally, the arteries and veins can also visually differ greatly across datasets. These can make developing an artery vein classification model that can generalize across different retinal datasets particularly challenging and thus are also the most probable reasons behind the under performances of the models on a test set that was very different from its training data.

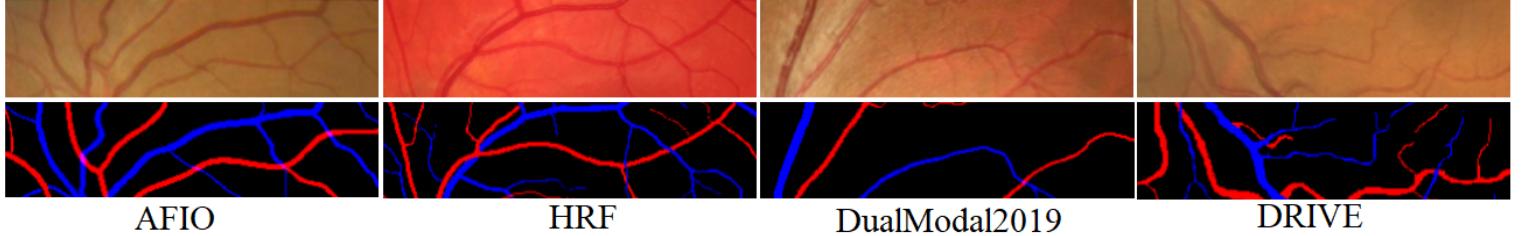


Figure 5: Data samples from each of the four datasets that were used for artery vein classification. The differences in the annotation style, granularity level, image quality and visual appearances of the arteries and veins across the datasets could be observed. Also, the samples were collected from regions close to the optic disc and therefore the color difference between arteries and veins (veins appear darker than arteries) can be noted.

Based on these observations, the 215 samples from all the four datasets were together split into training, validation and test sets. During this final data split, approximately 70 percent of samples from each of the four datasets was set aside for training and the remaining samples were divided equally to form the validation and test sets. This resulted in 155 training images (74 from AFIO, 22 from DualModal2019, 31 from HRF and 28 from DRIVE) and 30 images (13 from AFIO, 4 from DualModal2019, 7 from HRF and 6 from DRIVE) in both validation and test sets.

Preprocessing

Similar to preprocessing for vessel segmentation, all the input fundus images and the ground truths were resized to a uniform size of size 512×512 . In addition to that, there is also a need to deal with the inconsistent labelling methods across the datasets. AFIO and DualModal2019 datasets came with only two labels - artery and vein and thus did not need any processing. As for HRF, it came with three labels - artery, vein and uncertain. All the pixels that were marked uncertain were removed and ground truths were created with only artery and vein pixels. The DRIVE dataset came with four labels - artery, vein, overlapping and uncertain. The pixels that were under the label ‘overlapping’ were re-labelled to ‘artery’ and the pixels marked ‘uncertain’ were removed.

The DL model output will have three channels, one for each class - artery, vein and background. Thus, for the ground truth to be compatible with the model output, it should be a binary image with three channels. The first channel should have the artery pixels marked one and all other pixels zero. Similarly, the second channel should have the vein pixels marked one and all other pixels zero. The third channel should have the artery and vein pixels marked zero and all background pixels marked one. Such three channel binary ground truths were created for all the retinal fundus images.

Model and Training

The model for artery vein classification is mostly similar to the BEFD-Net model that was used for the vessel segmentation. The edge attention map was replaced with a ‘vessel attention map’ inspired from the work of Hu et al. in [20] and since the attention map is no longer noisy the denoising block was removed. Figure 6 shows the architecture of the network modified for artery vein classification.

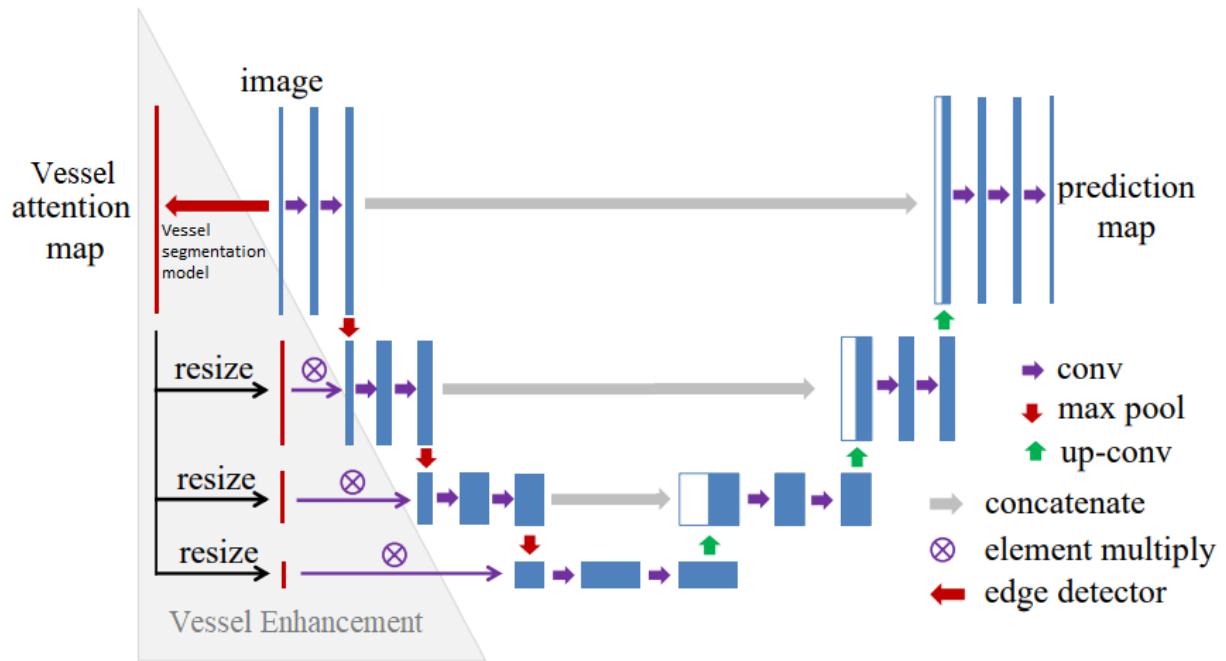


Figure 6: Architecture of artery vein segmentation model

Vessel Attention Map

The vessel attention map is to suppress the background features and help the model pay more attention to the vessel features. Figure 7 shows the steps involved in the computation of the vessel attention map from the output of the vessel segmentation model. The retinal fundus image I is passed as input to the vessel segmentation model to obtain the output I_{vessel} . The areas prone for misclassification are where the vessel features are subtle and difficult to separate from the background like in the microvascular areas or vascular boundaries. Usually the probability of these misclassification prone pixels are concentrated around 0.5 and thus to highlight these pixels in the vessel segmentation output an activation function is defined as per $F(x)$ in equation (3) where x is the vessel probability (pixels of I_{vessel}). The activation weights for

the pixels with vessel probability close to 0.5 is set to values close to $[(1 - e^{-0.5}) + 1]$ and for the pixels with vessel probability close to 0 and 1 (background pixels and main thick vessels pixels) the activation weights are close to 1.

$$F(x) = \left(e^{-|x-0.5|} - e^{-0.5} \right) + 1 \quad (3)$$

Finally the vessel attention map I_{Vmap} is calculated as $I_{Vmap} = [I_{vessel} * F(I_{vessel})] + 1$ where * represents element-wise multiplication. I_{Vmap} is in the range [1,2]. Its value for background pixels is close to 1 and for a pixel that is most probable to belong to a vessels, its value is close to 2.

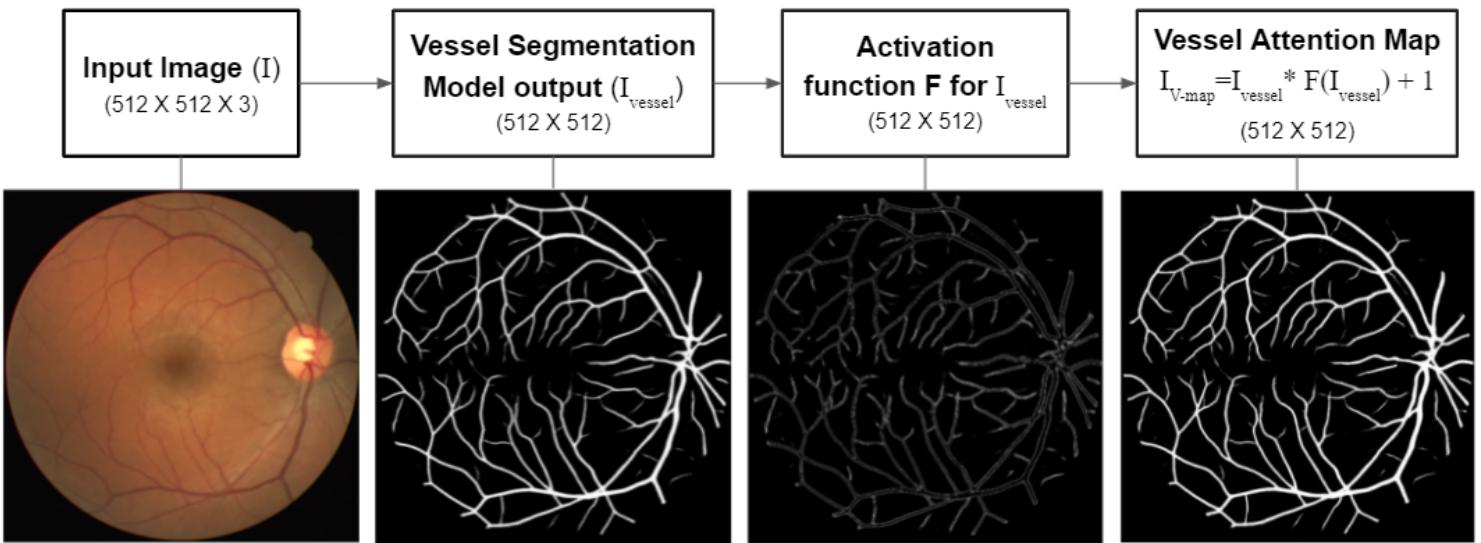


Figure 7: Steps to compute the vessel attention map

Training

Similar to the vessel segmentation model, the artery vein classification model was also 4 layers deep with 64 channels and kernel size 3. The dropout rate was again 0.1 and batch normalization was also used to increase stability during training. The activation function of the output layer in the vessel segmentation model was sigmoid. Sigmoid activation suits binary classification tasks well but since now the classification has three labels, the output layer at the end of the decoding path of the artery vein classification model was a 1×1 convolution layer with three filters and softmax activation.

The model was trained using adam optimizer with a learning rate of 0.001 and on a loss that was the sum of categorical cross entropy and dice loss. Including data augmentation during training either reduced the model performance or did not produce any significant improvements in it. Therefore, the model was trained for 100 epochs on the 155 training images without any data augmentations with a batch size of 5. At the end of each epoch, the model's validation loss and accuracy were calculated. Similar to the vessel segmentation model training, if there was no decrease in the validation loss for 10 consecutive epochs, the learning rate was decreased by a factor of 10. The model that gave the best accuracy on the validation data was saved.

Considering the fact that the output from the vessel segmentation model flows into the artery vein classification model, after the initial training, as a final step, the weights of the vessel segmentation model can be fine tuned by unfreezing its weights and training both the vessel segmentation and artery vein classification model with a small learning rate. Thus the model that was saved at the end of the first training was taken and after unfreezing the weights of the vessel segmentation model it was trained with a learning rate of 10^{-5} for another 25 epochs. The model from this training that gave the best categorical accuracy on the validation data was saved as the final artery vein classification model.

Post Processing

The proposed U-Net based artery vein classification model makes classification decisions mainly based on the local features. Firstly, as already mentioned, away from the optic disk the arteries and veins have similar appearances and usually manual artery vein classifications are based on vessel tracing. Thus considering just the local features around the vessel might not yield accurate results. Secondly, the local features mostly reflect the profiles of image components like color and shape and these vary greatly among fundus images that are acquired using different retinal cameras and are also greatly influenced by factors like illumination, image resolution and quality. Therefore there is always scope for misclassification errors. The post processing steps are aimed at correcting the labelling inconsistencies in the model outputs that could be identified based on the structure of the vessel system.

Segment Label Unification

One labelling inconsistency is the appearance of artery and vein labels in a vessel segment as shown in figure 8(b) and the segment label unification algorithm is used to process these inconsistencies. As the first step, a binary image for the detected vessels is obtained from the output of the artery vein classification model. The binary image is then skeletonized and all the key points - cross points and end points are detected. Vessel end points are pixels that have no more than one neighbor and vessel crossover points include vessel bifurcation points and points of vessel overlap and these are identified by pixels that have three or more neighbors in the skeletonized vessel binary image. Vessel pixels between keypoints are identified as segments. For each segment, artery and vein scores are calculated using the artery vein classification model output. Artery score is the sum of the artery probabilities of all the pixels in the segment and similarly vein score is the sum of the vein probabilities of the segment pixels. Finally, the segment is assigned the label with the highest score.

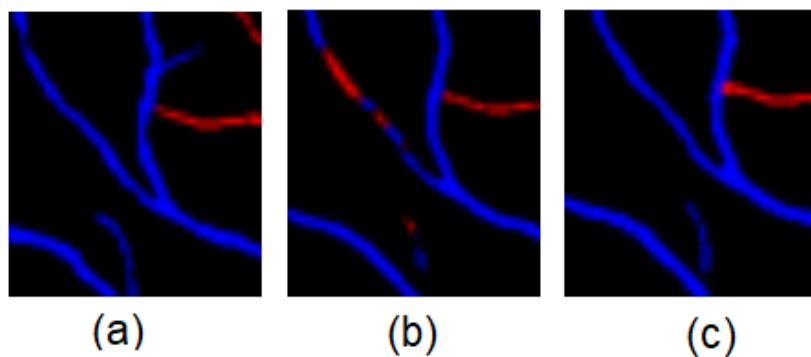


Figure 8: Segment label unification. (a) Ground truth (b) Model output where pixels in one vessel segment are partly classified as arteries and partly as veins (c) Result of segment label unification on the model output - all the pixels in a vessel segment are assigned the class with the highest confidence.

Evaluation Metrics

The metrics that were used to quantify the performance of the models are precision, recall, accuracy, and dice score. Precision gives the probability of a positive prediction to be correct while recall gives the probability of an actual positive to be predicted correctly as positive by the models. The mentioned evaluation metrics were calculated based on the below equations (4) to (7). TP, TN, FP and FN in the equations denote true positives, true negatives, false positives and false negatives respectively. Along with these metrics, the area under the ROC curve was also used for evaluation. All the metrics were calculated on the whole image.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Dice Score} = \frac{2TP}{2TP + FP + FN} \quad (7)$$

Results

Table 4 summarizes the performance of the BEFD-Net model on the test set. The test set was created using the CHASE-DB1 and DRIVE datasets and the performances of the model on these two datasets are also listed separately in the table. To act as a baseline, an equivalent U-Net (4 layers deep with 64 channels and kernel size 3) was trained for vessel segmentation on the training data. The performance of this U-Net is also given in table 4. A significant improvement in the average recall, accuracy and dice score can be noticed. The table also shows the cross annotator similarity for CHASE-DB1 dataset. The dataset provides two ground truths for each fundus image in it and the cross annotator similarity represents the level of agreement between these two ground truths. This can be considered as another baseline and it can be seen that the metrics evaluated on the BEFD-Net outputs come close to it. The vessel segmentation model outputs for samples from the test set can be found in figure 9. Magnified view of selected regions of interest in the model outputs can be found in figure 11 in the appendix.

Table 4: Performance of the BEFD-Net model on the vessel segmentation task and the performance of the U-Net for comparison.

Dataset	Model	Precision	Recall	Accuracy	AUC	Dice Score
Test	U-Net	0.8804	0.5662	0.9540	0.9694	0.6892
	BEFD-Net	0.8359	0.6962	0.9603	0.9672	0.7597
DRIVE	U-Net	0.8932	0.6052	0.9542	0.9703	0.7215
	BEFD-Net	0.8679	0.6687	0.9575	0.9637	0.7554
CHASE-DB1 (Annotator 1)	Cross annotator similarity	0.7925	0.7593	0.9655	-	0.7755
	U-Net	0.8538	0.4966	0.9538	0.9677	0.6280
	BEFD-Net	0.7894	0.7453	0.9644	0.9748	0.7667

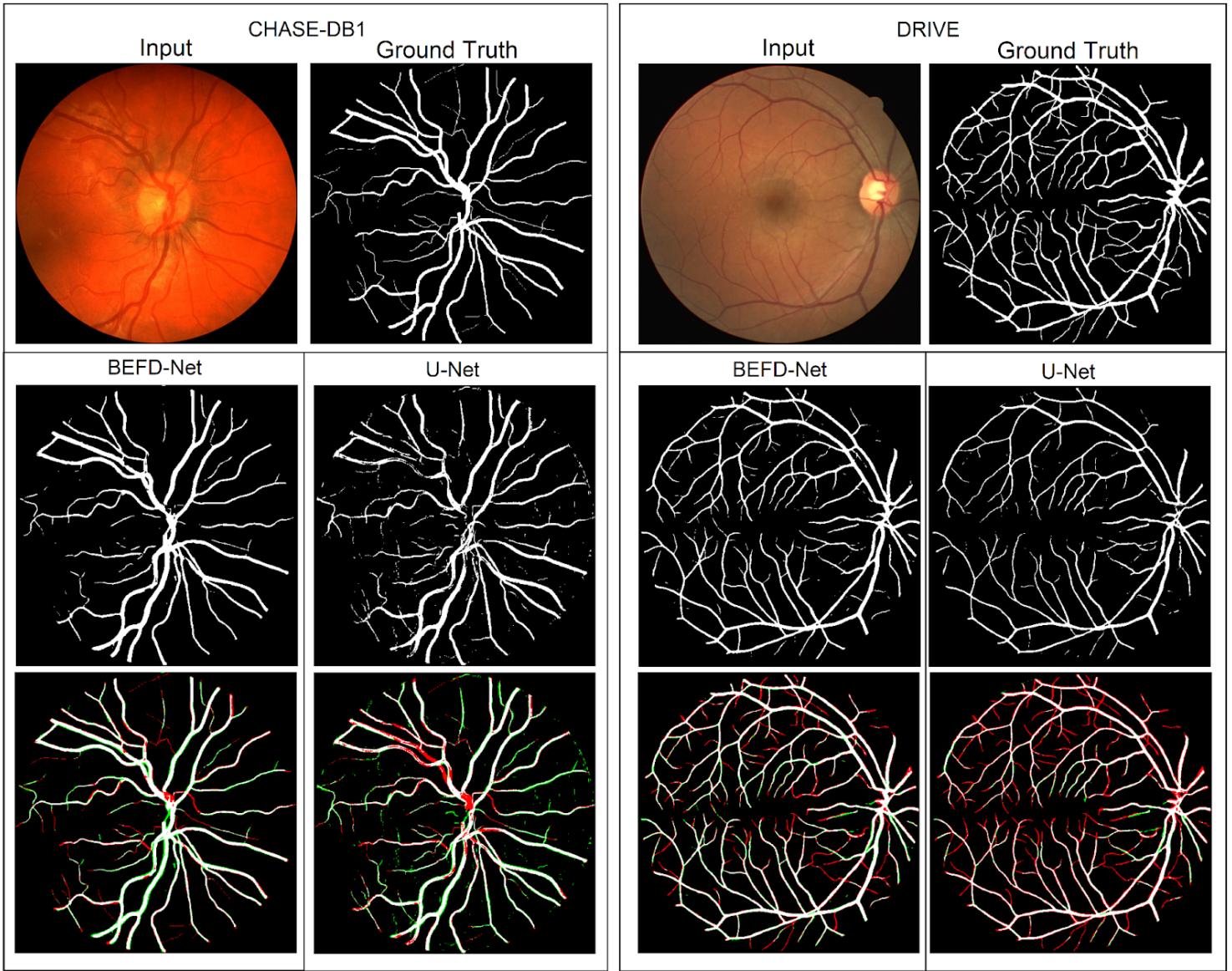


Figure 9: Vessel segmentation outputs for samples from test set. For comparison, outputs from an U-Net model are shown along with the outputs from the BEFD-Net model. On the last row, are model outputs with the false negatives marked red and false positives marked green.

In table 5 are the results of the artery vein classification model on its test set. The metrics were evaluated for each of the three classes - artery, vein and background. The inverse of the background ground truth will give the vessels ground truth. Thus instead of calculating the metrics on the background, it was calculated for the vessel and the results thus obtained are shown in the table. Similar to vessel segmentation, an equivalent U-Net was trained for the artery vein classification task. The results on the test set that were obtained using the U-Net are also given in table 5 for comparison. It could be noticed that when compared to the U-Net, the artery vein classification model yielded higher recall and dice score for all the three classes but the average accuracy remains more or less the same and there is also a slight drop in average precision. In figure 10 is the output of the artery vein classification model for a sample from its test set. The output for the same from the U-Net model is also given for comparison. Magnified view of selected regions of interest in the model outputs can be found in figure 12 in the appendix. Model outputs in both figure 10 and 12 are without any post processing. In this way the actual prediction accuracy of the models could be compared.

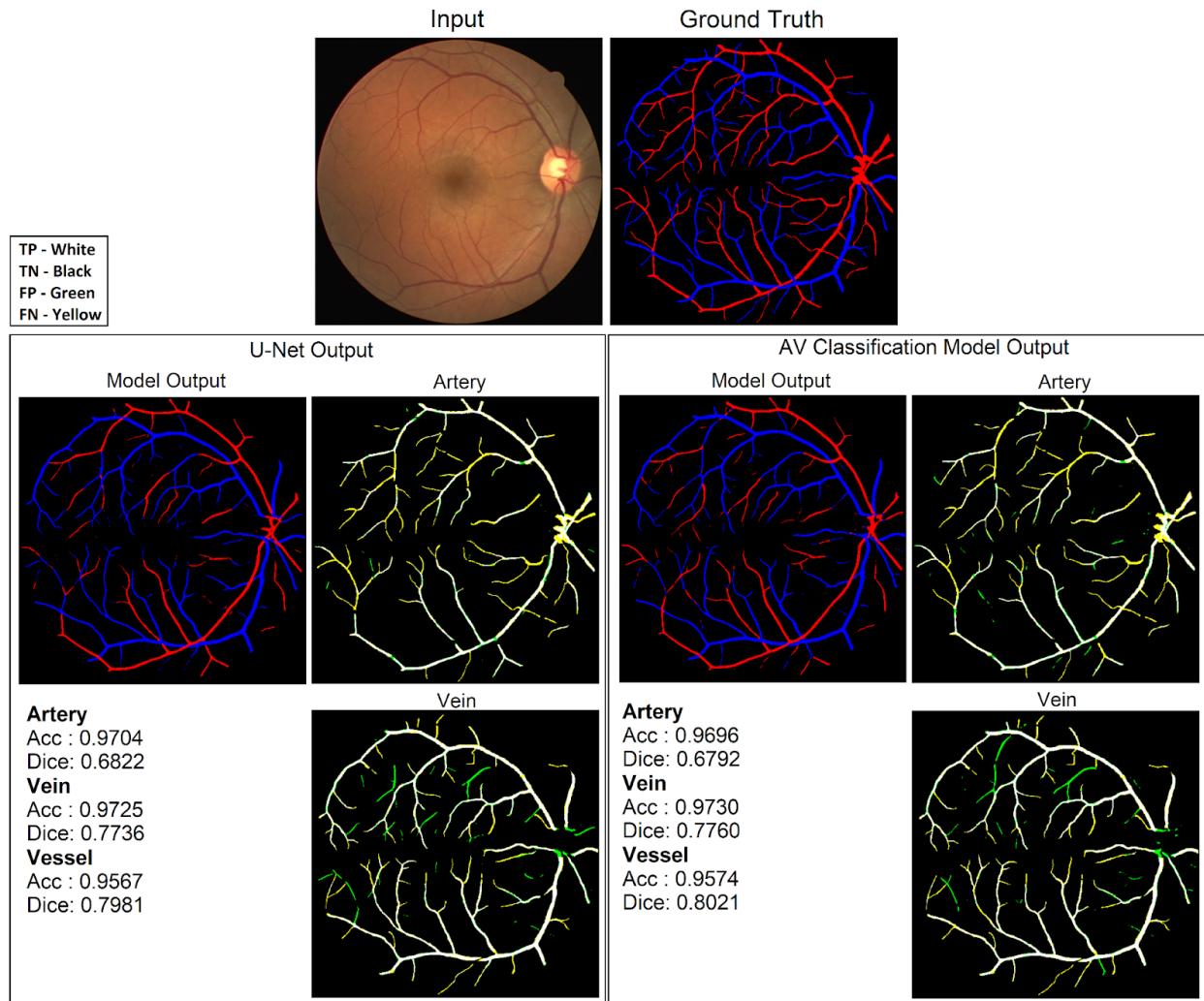


Figure 10: Artery vein classification model output for a test sample. For comparison, output from an U-Net model is also shown. For each model output, there are two images that show false positives (green) and false negatives (yellow) in artery classification and in vein classification.

Table 5: Performance of the artery vein classification model and the performance of the U-Net for comparison.

Dataset	Model	Class	Precision	Recall	Accuracy	Dice Score
Test	U-Net	Artery	0.7280	0.5998	0.9756	0.6499
		Vein	0.7610	0.6721	0.9752	0.7106
		Vessel	0.8223	0.7107	0.9623	0.7587
	AV-classification Model	Artery without post processing	0.7069	0.6235	0.9751	0.6565
		Artery with post processing	0.6961	0.6351	0.9748	0.6578
		Vein without post processing	0.7505	0.6887	0.9752	0.7152
		Vein with post processing	0.7477	0.6901	0.9750	0.7143
		Vessel	0.8074	0.7352	0.9627	0.7666

Discussion

Notable results were obtained at the end of the vessel segmentation task. The BEFD-Net was able to outperform its U-Net baseline. Additionally, its performance was also comparable to cross-annotator similarity metrics that were calculated on images that were provided with two ground truths. However, the model still can be seen to miss detecting a considerable number of fine capillaries. The speculated causes for this behavior are the difference in granularity across the different retinal datasets that were used and the information loss caused by the resizing operation that was done as preprocessing. There are also instances where the vessel boundaries are inconsistent with the ground truths. The vessels in the model output are either thicker or thinner than the annotations owing to false positives and false negatives. This could be seen in the images in the last row of figure 9 where green (false positives) or red (false negatives) pixels can be seen along the vessel boundaries in the model outputs. The reason behind the difference in vessel thicknesses between the model outputs and the corresponding ground truths could be the difference in annotation styles across different retinal datasets potentially leading to cases of over-segmentation or under-segmentation in ground truths.

Artery vein classification, when compared to vessel segmentation, is a more complicated task. The model proposed in this work overperforms the U-Net but only marginally. Given that even human observers can find it difficult to separate the vessels that are away from the optic disc into arteries and veins based just on its local appearance, instead of investing on deeper or more complex deep learning models, opting for classical approaches might be more advantageous in this case. Although there was a significant need for improvement in the model performance, it was observed that the vessels close to the optic disc are usually correctly classified. To correct the misclassifications in the vessels that are away from the optic disc, graph-based classical algorithms that allow vessel tracing can be adopted. The segment label unification algorithm that was discussed as a post-processing step is a very simple technique to deal with misclassifications and can be considered as a starting step that can correct a limited number of misclassifications. Due to limited time, more sophisticated graph-based post processing steps to identify and correct misclassifications are suggested as future work.

Conclusion

U-Net based vessel segmentation and artery vein classification models for retinal fundus images were trained and tested in this work. A BEFD-Net model was used for the vessel segmentation task and a similar model adapted for artery vein classification was used for that task. Noteworthy results were obtained for vessel segmentation but the artery vein classification model needs further improvement. Since the model was able to correctly classify the vessels close to the optic disc, as a future work recommendation, graph-based vessel tracing algorithms are suggested for performance improvement.

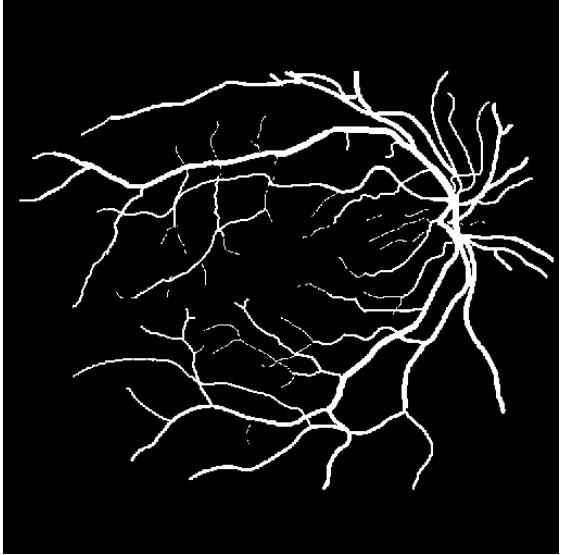
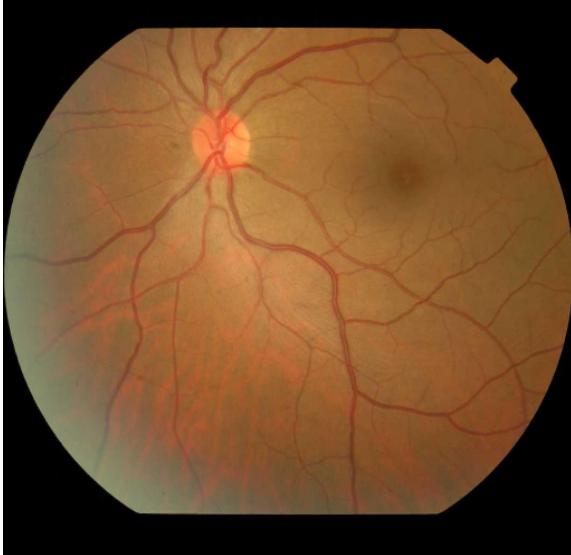
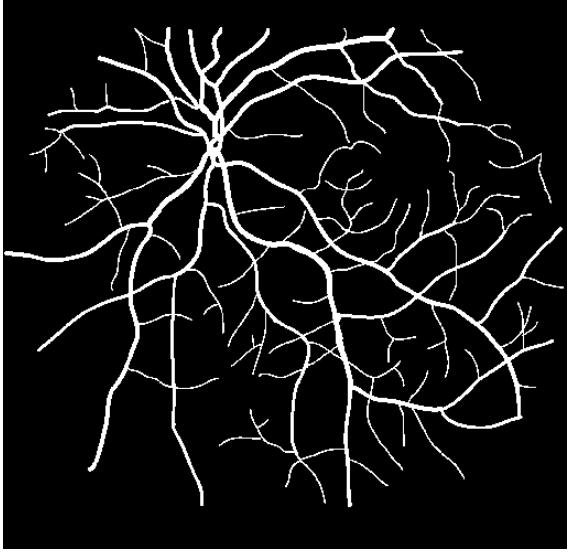
References:

- [1] Dou, G.R., Wang, L., Wang, Y.S. and Han, H., 2012. Notch signaling in ocular vasculature development and diseases. *Molecular Medicine*, 18(1), pp.47-55.
- [2] Wong, T.Y., Klein, R., Sharrett, A.R., Duncan, B.B., Couper, D.J., Klein, B.E., Hubbard, L.D. and Nieto, F.J., 2004. Retinal arteriolar diameter and risk for hypertension. *Annals of internal medicine*, 140(4), pp.248-255.
- [3] Rousso, L. and Sowka, J., 2017. Recognizing abnormal vasculature: a guide to following and educating patients who face this class of sight-threatening diagnoses. *Review of Optometry*, 154(1), pp.82-87. [Recognizing Abnormal Vasculature \(reviewofoptometry.com\)](http://www.reviewofoptometry.com)
- [4] Nguyen, U.T., Bhuiyan, A., Park, L.A. and Ramamohanarao, K., 2013. An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern recognition*, 46(3), pp.703-715.
- [5] Yavuz, Z. and Köse, C., 2017. Blood vessel extraction in color retinal fundus images with enhancement filtering and unsupervised classification. *Journal of healthcare engineering*, 2017.
- [6] Dastgheib, M.A. and Seyedin, S., 2018, December. Vessel Segmentation in Coloured Retinal Fundus Images Based on Multi-scale Analysis. In 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) (pp. 201-205). IEEE.
- [7] Zhu, C., Zou, B., Zhao, R., Cui, J., Duan, X., Chen, Z. and Liang, Y., 2017. Retinal vessel segmentation in colour fundus images using extreme learning machine. *Computerized Medical Imaging and Graphics*, 55, pp.68-77.
- [8] Akbar, S., Sharif, M., Akram, M.U., Saba, T., Mahmood, T. and Kolivand, M., 2019. Automated techniques for blood vessels segmentation through fundus retinal images: A review. *Microscopy research and technique*, 82(2), pp.153-170.
- [9] Chen, C., Chuah, J.H., Raza, A. and Wang, Y., 2021. Retinal Vessel Segmentation Using Deep Learning: A Review. *IEEE Access*.
- [10] Fan, Z. and Mo, J.J., 2016, July. Automated blood vessel segmentation based on de-noising auto-encoder and neural network. In 2016 International Conference on Machine Learning and Cybernetics (ICMLC) (Vol. 2, pp. 849-856). IEEE.
- [11] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- [12] Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M. and Asari, V.K., 2018. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation.
- [13] Wu, Y., Xia, Y., Song, Y., Zhang, D., Liu, D., Zhang, C. and Cai, W., 2019, October. Vessel-Net: retinal vessel segmentation under multi-path supervision. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 264-272). Springer, Cham.
- [14] Kamran, S.A., Hossain, K.F., Tavakkoli, A., Zuckerbrod, S.L., Sanders, K.M. and Baker, S.A., 2021. RV-GAN: retinal vessel segmentation from fundus images using multi-scale generative adversarial networks. *arXiv preprint arXiv:2101.00535*.
- [15] Hu, J., Wang, H., Cao, Z., Wu, G., Jonas, J.B., Wang, Y.X. and Zhang, J., 2021. Automatic Artery/Vein Classification Using a Vessel-Constraint Network for Multicenter Fundus Images. *Frontiers in Cell and Developmental Biology*, 9.
- [16] Joshi, V.S., Reinhardt, J.M., Garvin, M.K. and Abramoff, M.D., 2014. Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks. *PloS one*, 9(2), p.e88061.
- [17] Zamperini, A., Giachetti, A., Trucco, E. and Chin, K.S., 2012, June. Effective features for artery-vein classification in digital fundus images. In *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 1-6). IEEE.
- [18] Niemeijer, M., van Ginneken, B. and Abràmoff, M.D., 2009, February. Automatic classification of retinal vessels into arteries and veins. In *Medical Imaging 2009: Computer-Aided Diagnosis* (Vol. 7260, p. 72601F). International Society for Optics and Photonics.

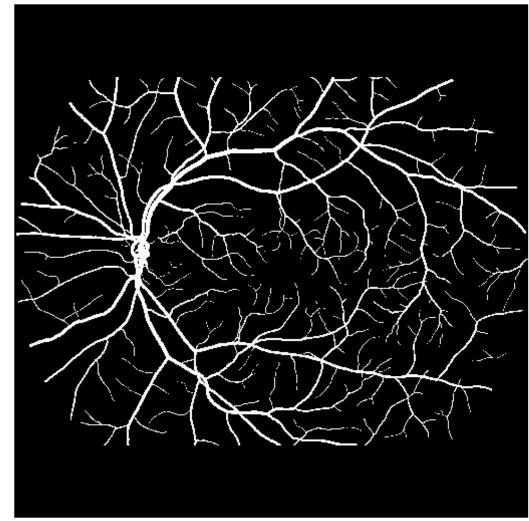
- [19] Hemelings, R., Elen, B., Stalmans, I., Van Keer, K., De Boever, P. and Blaschko, M.B., 2019. Artery–vein segmentation in fundus images using a fully convolutional network. *Computerized Medical Imaging and Graphics*, 76, p.101636.
- [20] Hu, J., Wang, H., Cao, Z., Wu, G., Jonas, J.B., Wang, Y.X. and Zhang, J., 2021. Automatic Artery/Vein Classification Using a Vessel-Constraint Network for Multicenter Fundus Images. *Frontiers in Cell and Developmental Biology*, 9.
- [21] Zhang, M., Yu, F., Zhao, J., Zhang, L. and Li, Q., 2020, October. BEFD: Boundary Enhancement and Feature Denoising for Vessel Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 775-785). Springer, Cham.
- [22] Wang, B., Qiu, S. and He, H., 2019, October. Dual encoding u-net for retinal vessel segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 84-92). Springer, Cham.
- [23] Wang, W., Zhong, J., Wu, H., Wen, Z. and Qin, J., 2020, October. Rvseg-net: An efficient feature pyramid cascade network for retinal vessel segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 796-805). Springer, Cham.
- [24] Zhuang, J., 2018. LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv preprint arXiv:1810.07810*.
- [25] Li, L., Verma, M., Nakashima, Y., Nagahara, H. and Kawasaki, R., 2020. Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3656-3665).
- [26] Farnell, D.J., Hatfield, F.N., Knox, P., Reakes, M., Spencer, S., Parry, D. and Harding, S.P., 2008. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *Journal of the Franklin institute*, 345(7), pp.748-765.
- [27] Akram, M.U., Akbar, S., Hassan, T., Khawaja, S.G., Yasin, U. and Basit, I., 2020. Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema. *Data in brief*, 29, p.105282.
- [28] Budai, A., Bock, R., Maier, A., Hornegger, J. and Michelson, G., n.d. *High-Resolution Fundus (HRF) Image Database*. [online] Available at: <<https://www5.cs.fau.de/research/data/fundus-images/>> [Accessed 15 November 2021].
- [29] Budai, A., Bock, R., Maier, A., Hornegger, J. and Michelson, G., 2013. Robust vessel segmentation in fundus images. *International journal of biomedical imaging*, 2013.
- [30] Shulin Zhang, Rui Zheng, Mingzhai Sun, 2019. DualModal2019 dataset. Available at: <https://dx.doi.org/10.21227/wbpy-mf57>.
- [31] Hoover, A., The STARE Project. [online] Available at: <<http://cecas.clemson.edu/~ahoover/stare/>> [Accessed 15 November 2021].
- [32] Hoover, A.D., Kouznetsova, V. and Goldbaum, M., 2000. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical imaging*, 19(3), pp.203-210.
- [33] DRIVE - Grand Challenge. [online] Available at: <<https://drive.grand-challenge.org/>> [Accessed 15 November 2021].
- [34] Xie, C., Wu, Y., Maaten, L.V.D., Yuille, A.L. and He, K., 2019. Feature denoising for improving adversarial robustness. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 501-509).
- [35] Hu, Q., Abràmoff, M.D. and Garvin, M.K., 2013, September. Automated separation of binary overlapping trees in low-contrast color retinal images. In *International conference on medical image computing and computer-assisted intervention* (pp. 436-443). Springer, Berlin, Heidelberg.
- [36] Hu, Q., Abràmoff, M.D. and Garvin, M.K., 2013, *RITE dataset | Department of Ophthalmology and Visual Sciences*. [online] Available at: <<https://medicine.uiowa.edu/eye/rite-dataset>> [Accessed 18 November 2021].
- [37] Hemelings, R., 2021. *GitHub - rubenhx/av-segmentation*. [online] GitHub. Available at: <<https://github.com/rubenhx/av-segmentation>> [Accessed 18 November 2021].

Appendix

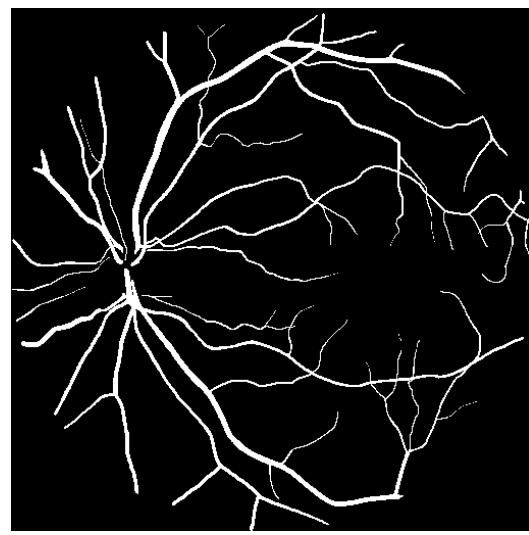
Sample fundus images from each dataset and their ground truth vessel segmentations.

Dataset	Retinal Fundus Image	Vessel Segmentation
ARIA		
AFIO		

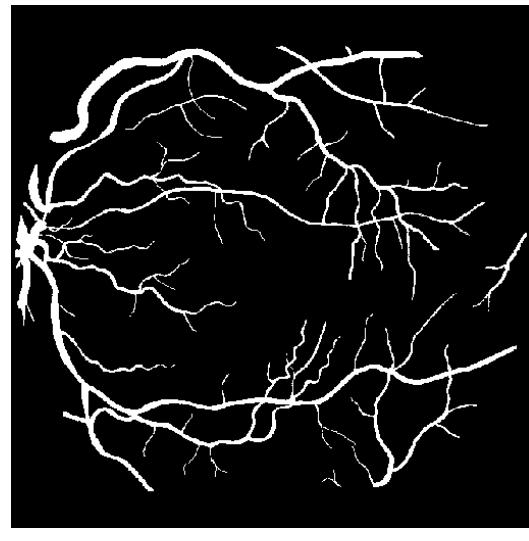
HRF



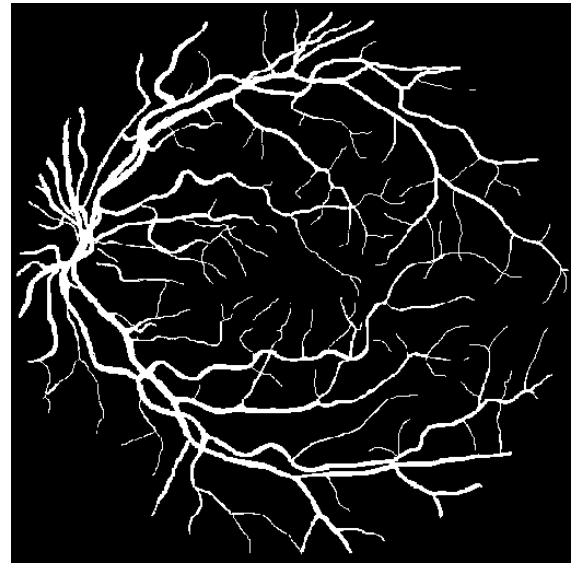
DualModal2019



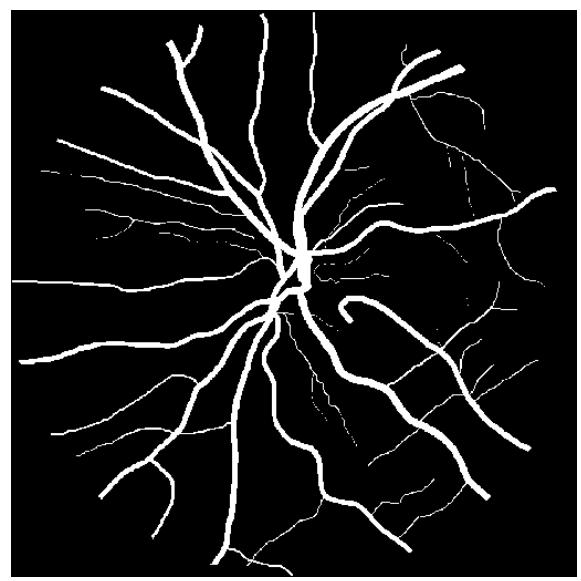
STARE



DRIVE



CHASE-DB1



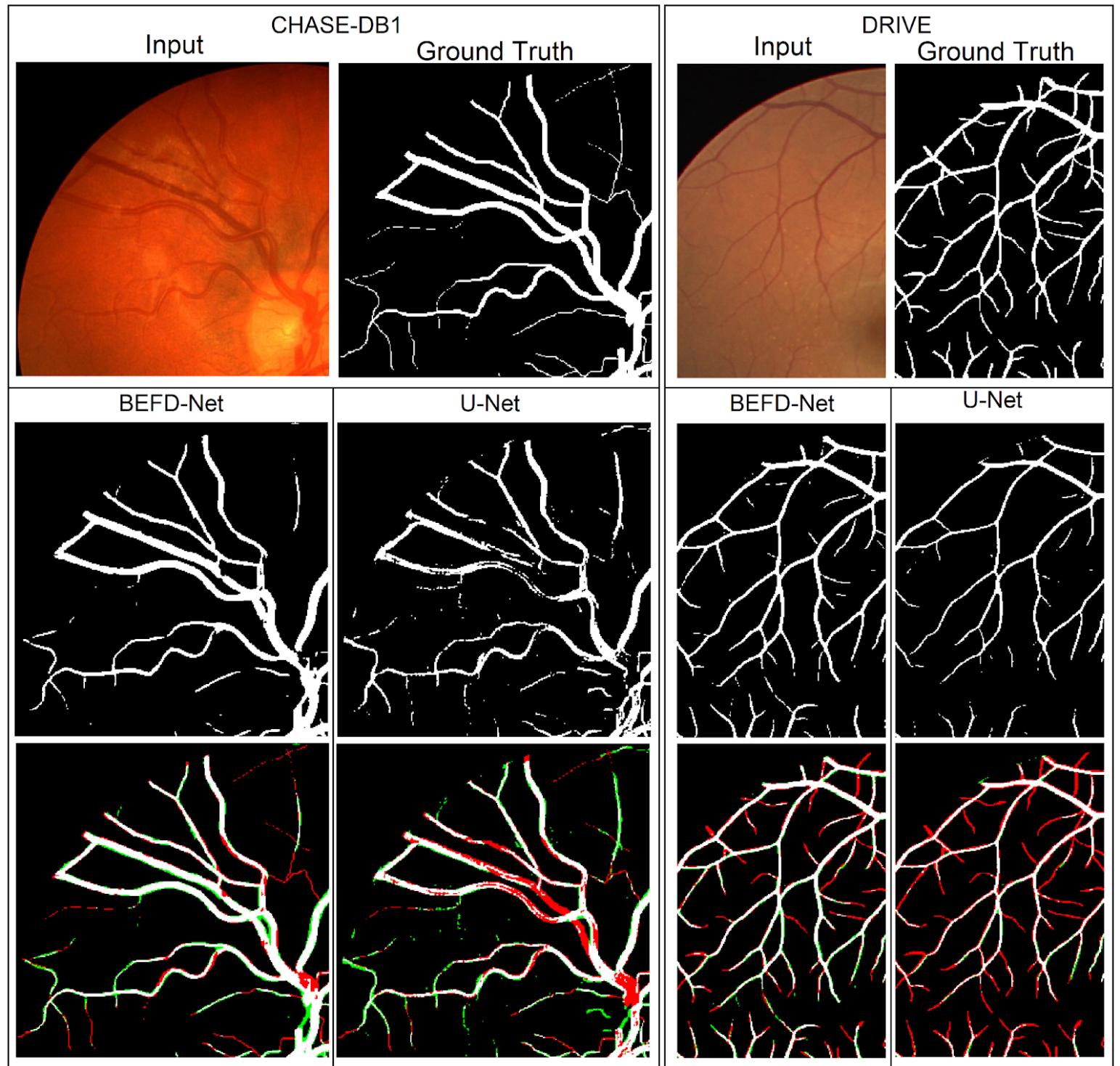


Figure 11: Magnified view of selected regions of interest from the vessel segmentation model outputs. On the last row is the same magnified view with the false negatives marked red and the false positives marked green.

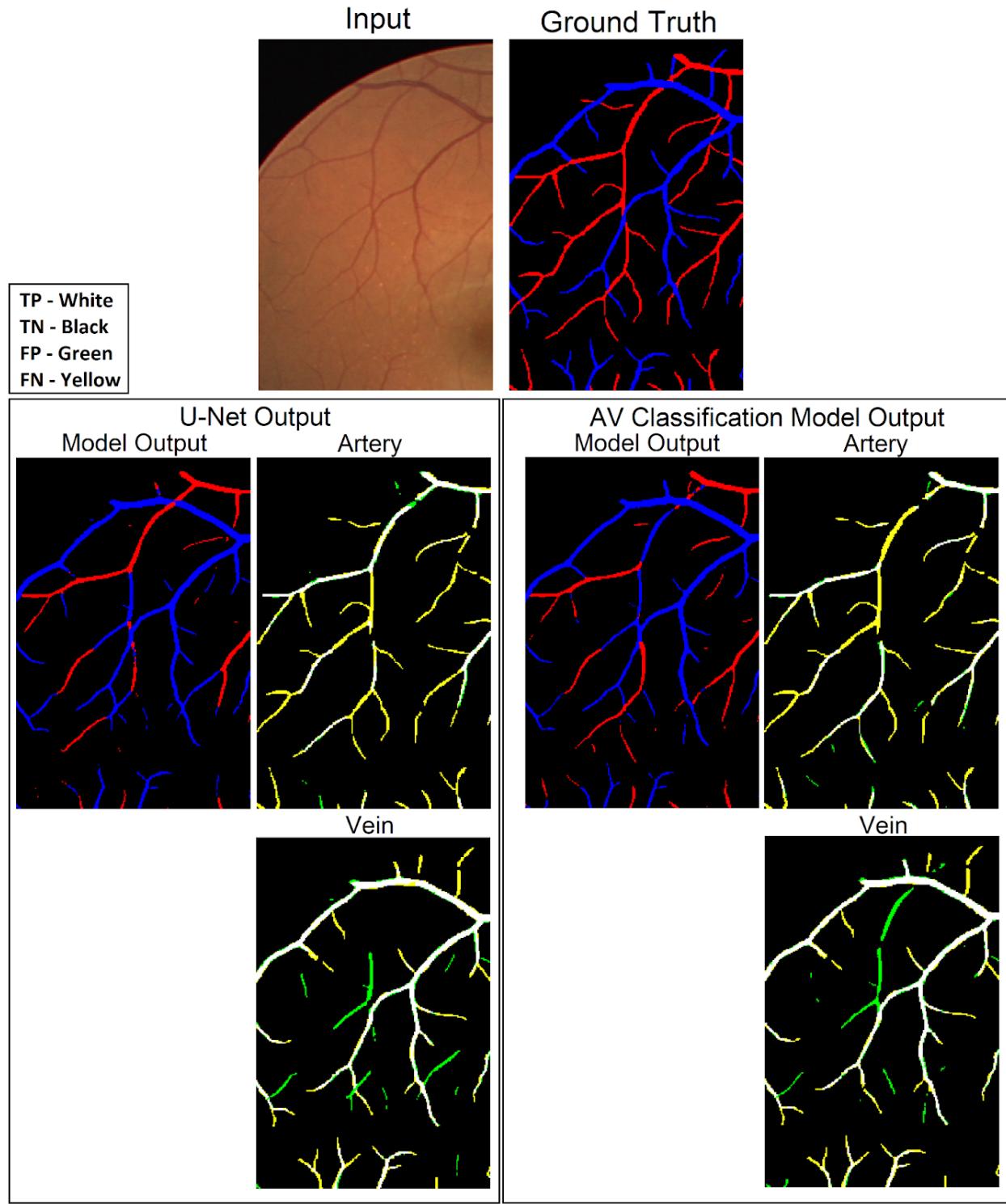


Figure 12: Magnified views of the artery vein classification models. For each model output, there are two separate images that show false positives (green) and false negatives (yellow) in artery classification and in vein classification.