Diabetic retinopathy grading with respect to the segmented lesions

by

Hoda Kheradfallah

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Vision Science and Systems Design Engineering

Waterloo, Ontario, Canada, 2022

Author’s Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The following publications have resulted from the work presented in this thesis and some parts of this thesis are put from these published articles, my role in each of them is the first author:

1. Lakshminarayanan, Vasudevan, Hoda Kheradfallah, Arya Sarkar, and Janarthanam Jothi Balaji. “Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey.” Journal of Imaging 7, no. 9 (2021): 165.

2. Kheradfallah, Hoda, Balaji, Janarthanam Jothi, Jayakumar, Varadharajan, Abdul Rasheed, Mohammed and Lakshminarayanan, Vasudevan. “Annotation and Segmentation of Diabetic Retinopathy Lesions: An Explainable AI Application.” SPIE Journal of Imaging, (2022):502-511.

Abstract

Diabetic Retinopathy (DR) is a major cause of visual loss among the working-age population and has a globally high prevalence rate. This disease is caused by the capillary damage due to the chronic high blood glucose level that can progress to proliferated levels. DR affects retina, and its severity can be observable on fundoscopy or optical coherence tomography images. There are 10 major DR-associated lesions defined by the International Clinical Diabetic Retinopathy Scale (ICDRS). This system categorizes DR depending on its severity level and for each lesion it defines the initial presence in a certain level. Then, the grading is done through detection of each lesion type and number according to the grading system criteria.

The diagnosis process is currently done manually; however, it is time consuming and requires a trained clinician. Thus, several automated alternatives have been proposed as a partial substitute in the diagnosis process. Deep Convolutional Neural networks (DCNNs) are among the most successful Computer-Aided Diagnosis methods in terms of performance. Even though these state-of-the-art architectures reach high performance scores, but they act as black boxes and their decision process and learned features are not clinically interpretable which becomes a major barrier for their adoption for ophthalmological purposes. This problem was initially addressed by eXplainable Artificial Intelligence (XAI) solutions. XAI methods can visualize the critical regions of an input image for a given DR grade. Initially, we applied some common fundamental XAI methods on customized trained DCNN model outputs. The attention map results of applied XAI solutions are either generic or sparse and do not provide sufficient interpretation for a predicted grade.

Hence, we applied a publicly available dataset, FGADR, which has sufficient fundus images with pixel-wise annotation of six DR lesions. We selected 143 samples of FGADR database and annotated the missing lesions including vitreous-preretinal hemorrhage, intraretinal hemorrhage, venous beading (VB) and fibrous proliferation (FP) that are annotated for the first time as a public dataset.

Then, we applied distinct DCNNs with similar architectures of holistically nested edge detector network (HEDNet) and pretrained weights of backbone on ImageNet. These models were fine-tuned to segment each lesion, separately. Our net plan was to apply these segmentation outputs in grading the disease severity based on ICDRS criteria.

Finally, we found that 4 out of 9 model outputs related to vascular abnormalities were not satisfying a defined level to use them in the grading step. these lesions include VB, IRMA,MA and FP with the mAP scores of 19%, 21%, 26% and 22%, respectively. The suggested solutions to improve their performance are reducing diversity of lesion morphological features in the image sets for training, increase number of dataset samples and try other network architectures related to graph patterns due to similarity to vessel patterns such as hierarchical networks.

Overall, this study could provide a novel and comprehensive dataset of pixel-wise annotations of DR-related lesions, and it could be used for further research with focus on DR lesion segmentation. In addition, this study extended the use of HEDNet model to segment the newly annotated lesions.

Acknowledgements

I am grateful of my supervisors Dr. Vasudevan Lakshminarayanan and Dr. John Zelek for their unwavering support and insightful guidance in my field of study. I am so grateful of Dr. Lakshminarayanan for giving me the freedom to choose a novel intriguing research topic and provide all requirements to make it more attainable reaching my goals. Special thanks to my committee members Dr.Kaamran Rahemifar and Dr.Jeff Hovis for their valued knowledge and bight advices on my research. I appreciate the help of my colleagues Jothi, Abdul and Varadhu for sharing their experience and assistance in my research.

Dedication

With my whole heart to my loving parents.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	artificial intelligence
AP	average precision
PR-AUC	area under the precision recall curve
CAM	class activation mapping
CAD	computer aided diagnosis
CE	cross entropy
CNN	convolutional neural network
CWS	cotton wool spots
DCNN	deep convolutional neural network
DM	diabetes mellitus
DR	diabetic retinopathy
Ex	hard exudate
FP	fibrous proliferation
FOV	field of view
GAN	generative adversial network

GradCAM	gradient-weighted class activation mapping
HEDNet	holistically nested edge detector network
IG	integrated gradients
IHE	intraretinal haemorrhage
IRMA	intraretinal microvascular abnormalities
LIME	local interpretable model-agnostic explanations
LRP	layer wise relevance propagation
MA	microaneurysm
NV	neovascularization
PHE	preretinal haemorrhage
XAI	explainable AI
VB	venous beading
VHE	vitreous haemorrhage
VPHE	vitreous preretinal haemorrhage

Chapter 1

Automated Diabetic Retinopathy Diagnosis

Based on:

- **Lakshminarayanan, Vasudevan, Kheradfallah, Hoda, Sarkar, Arya, and Balaji, Janarthanam Jothi. "Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey." Journal of Imaging 7, no. 9 (2021): 165.**

1.1. Introduction

Diabetic retinopathy (DR) is a leading cause of visual impairment with a prevalent rate. This disease is screenable on retinal images and could be detectable through automated diagnosis solutions. Advancements in deep convolutional neural networks (DCNNs) in automating DR diagnosis has been proved to be promising and is the focus of this research [1] . The major barrier to adoption of this sort of methods is the lack of reasoning behind their decisions [2]. To add on their interpretability, explainable artificial intelligence (XAI) has proposed some methods to validate before percolation to healthcare systems. The major contributions of this thesis in the scope of model evaluation are:

- Provide summaries over the most unique automated DR diagnosis approaches and their categorization to DCNN based solutions and others.
- Introduce our DCNN model for disease grading, analysis of its main structure.
- Implement a selection of explainability methods to generate explanation heatmaps of input images.
- Evaluation of the XAI results from a clinician point of view.
- Propose the alternative solution which is segmentation-based DR diagnosis
- Introduce our collected FGADR 143-9 database
- Evaluate existing DR segmentation networks
- Segmentation results on our proposed dataset and the directions of future research

In this thesis, chapter 1 serves as a brief overview to DR, its screening and computerized diagnosis methods. In chapter 2 deep learning-based solutions in DR diagnosis are discussed. Chapter 3 covers the most common XAI approaches on DR diagnosis and demonstrate deficiencies of XAI solution with a clinical insight. Chapter 4 and 5 discuss on the existing DR segmentation databases, the proposed database and the deep segmentation model with implementation details. A conclusion over the whole project and future improvement directions are also presented in chapter 6.

1.2. CAD methods in DR detection

1.2.1. Diabetic Retinopathy

Diabetes mellitus (DM) is mainly caused by insulin resistance which directly affects blood glucose level and causes hyperglycemia. Hyperglycemia can have multiple effects on vessels. Among them, diabetic retinopathy is highly prevalent [3]. Among individuals with DM, DR had 22.27% prevalence rate in 2021, globally [4]. This disease is also affecting a higher portion of the global population, as it was reported to be 103.12 million in 2021 and estimated to be 160.5 million by 2045 [4]. This disease begins at a mild level and can progress to advanced levels with a progressive vision affection [1]. There are some clinical characteristics that indicate the presence of DR on retina that are observed through retinal imaging techniques.

DR progression state can be graded using a standard grading system such as the Early Treatment Diabetic Retinopathy Study (ETDRS) [5] that separates DR characteristics in high details to over 80 severity levels [1]. In addition, this system needs information from all seven fields of view (FOV) through retinal imaging to be qualified for grading. There are also more generic and widely used grading systems such as the International Clinical Diabetic Retinopathy Scale (ICDRS) [6] that defines 10 major DR-associated lesions: Microaneurysms (MA), Intraretinal Haemorrhages (IHE), Hard Exudates (Ex), Cotton Wool Spots (CWS), Venous Beading (VB), Intraretinal Microvascular Abnormality (IRMA), Preretinal Haemorrhages (PHE), Vitreous Haemorrhages (VHE), Neovascularization (NV) and Fibrous Proliferation (FP). ICDRS also has five severity levels for DR as follows:

1. No Retinopathy

2. Mild Non-Proliferative Diabetic Retinopathy (NPDR): As the first stage of diabetic retinopathy, it includes tiny areas of swelling in retinal blood vessels known as microaneurysms (MA) [1, 3] (Figure 1.2A).

3. Moderate NPDR: When left unchecked, mild NPDR progresses to a moderate stage when bleeding starts from the blocked retinal vessels. IHE signs should be less than 20 in at least one quadrant. At this level, hard exudates (Ex) may also exist (Figure 1.2B). Venous Beadings (VB) are signs on retinal funduscopy images that are caused as a result of the dilation and constriction of venules in the retina [1, 3]. At moderate NPDR, VB might be detected but it should be observed in less than 2 quadrants.

4. Severe NPDR: In this level, in addition to MA and Ex, any of intra-retinal hemorrhages (IHE), Intra-Retinal Microvascular Abnormalities (IRMA) and VB could occur but none of lesions specific for the proliferative DR should be observed. As a threshold for the number of IHE, ICDR considers the images with over 20 IHE on all four fundus

quadrants as severe level (Figure 1.2C). There may also be IRMA which can be seen as bulges of thin vessels. IRMA could appear as small and sharp-bordered red spots in at least one quadrant. VB also should exist in over two quadrants [1, 3].

5. Proliferative Diabetic Retinopathy (PDR): Different functional visual problems occur in PDR, such as blurriness, reduced field of vision, and even complete blindness in some cases. This progressive stage of DR mainly occurs when the retinal examination is left unchecked. At this level, which is also called vision threatening level of DR could have one or more of the following lesions. The creation of new blood vessel networks on retina to feed the areas of damaged blood vessels which is termed as Neovascularization (NV), Blood leakage from the tiny abnormal blood vessel networks and proliferation of fibrous tissue as a natural eye tissue recovery mechanism [1, 3] (Figure 1.2D).

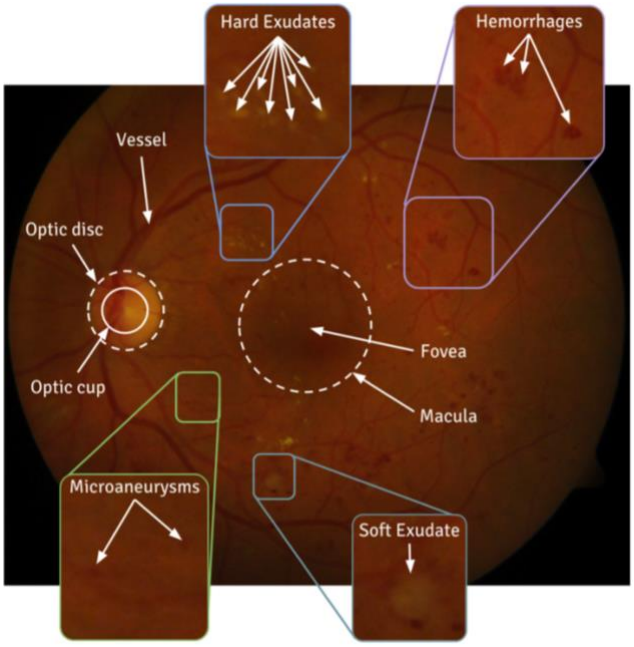

Figure 1.1.DR lesions on a sample retinal funduscopy image.

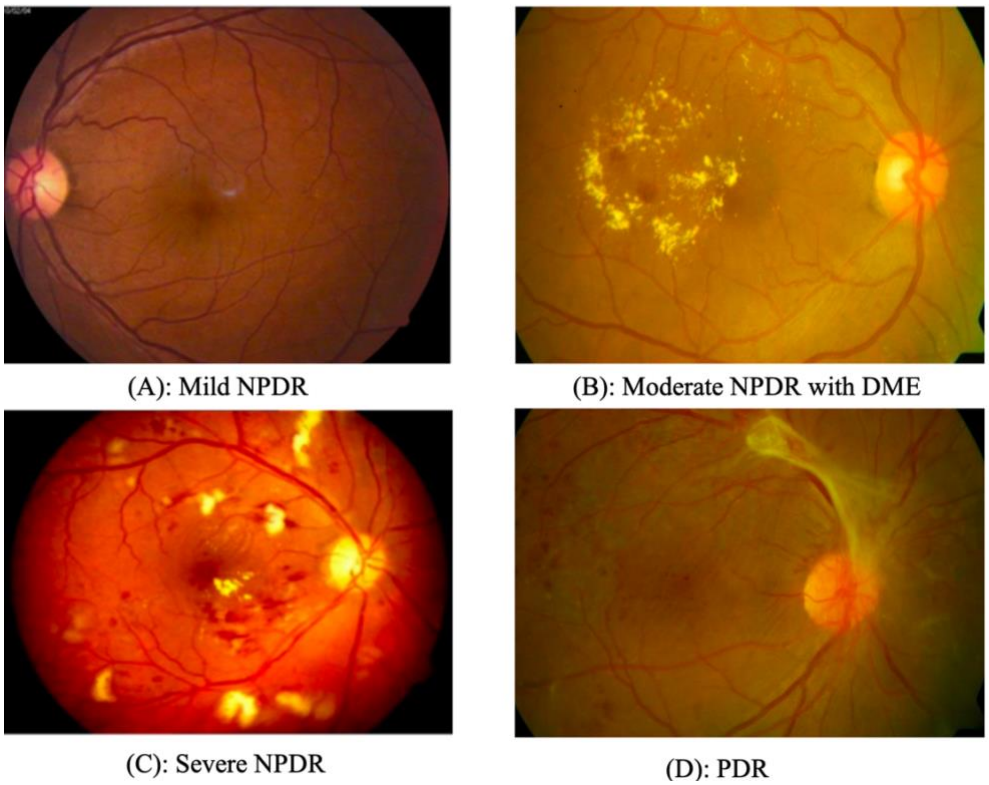

Figure 1.2. One FOV retinal funduscopy images of level 2-5 of DR (images courtesy of Rajiv Raman et al., Sankara Nethralaya, India).

1.2.2.DR screening methods

DR diagnosis requires screening on which fine pathognomonic DR signs in the initial stages are detectable. In this scope, dilated stereoscopic fundoscopy, fundus analogue photography and optical coherence tomography (OCT) could be applied [7, 8]. The funduscopy systems could be divided to pupil dilation-required (mydriatic)or without pupil dilation (non-mydriatic). Since non-mydriatic fundus cameras may be unable to visualize DR with high quality in cases where cataract is also present, mydriatic fundus imaging is the most common solution in DR screening worldwide [9].

The academic gold standard DR grading system, ETDRS, requires seven- 30° FOV stereoscopic fundus photographs through a dilated pupil [1]. Then experienced clinicians apply ETDRS criteria to grade severity. ETDRS grading still remains the highest standard in academic research, however, it cannot be applied for practical screening among populations due to its time-consuming procedure and high number of funduscopy FOV images required for diagnosis [10]. Furthermore, seven-field stereoscopic fundus photographs require many trials per patient and are therefore not suitable for population screening with an incremental prevalence rate of the disease [11].

With equally effectiveness, examining retina with slit-lamp biomicroscopic funduscopy is also considered the clinical gold standard in DR diagnosis, but this method is not applicable for large-scale screening [7]. Hence, there is space to introduce a standard grading system that is both applicable for clinical and academic purposes. In this research, ICDRS can satisfy these requirements.

The detectability of DR colour fundus images was compared with ophthalmoscopy imaging in clinical application. The funduscopy camera detection rate was more than twice as high as ophthalmoscopy imaging through dilated pupils [7]. Hence, color fundus photography on dilated pupils is applied in this study.

The next step is diagnosis, which is manually done through finding DR-associated lesions and comparing them with the selected grading system criteria. Currently, the latest automated diagnosis technologies such as IDx-DR [12] and EyeArt AI [13] have a different way of grading the disease. The core artificial intelligence (AI) unit in these devices classifies the images using the previously trained data samples and the extracted patterns. In section 1.2.3 we will briefly explain the automated diagnosis approaches.

1.2.3.CAD history in DR studies

Recent developments in CAD techniques, which mainly belong to the scope of artificial intelligence (AI), are becoming more prominent in modern ophthalmology [14] as they can save time, cost, and human resources for routine DR screening and involve lower diagnostic error factors [14]. CAD can also efficiently manage the increasing number of afflicted DR patients [15] and diagnose DR in early stages when fewer sight threatening effects are present. AI based approaches could be divided into machine learning-based (ML) and deep learning-based (DL) solutions [1]. Overall DL-based methods outperform ML-based solutions in the diagnosis of more than mild DR with area under the receiver operating characteristic (AUROC) values of 0.98 and 0.96 with 95% confidence interval, respectively [16]. There are also further limitations that ML solutions have in practice.

ML methods apply hand-engineered image features which require a prior ophthalmological expert knowledge and extensive investigation of critical DR features. Furthermore, the ML solutions are unable to extract high level features such as the shape and texture of lesions as local objects on the image [1]. They are also vulnerable to direct bias if transferred to new imaging configuration [16]. Hence, there is space to look for other alternative solutions to address the DR diagnosis. DL-based models will be discussed with further details in chapter 2.

1.3. Discussion

DR is a major cause of visual blindness and there are several standard severity grading systems proposed for this disease. Among these systems ICDRS is the most convenient system in automated diagnosis studies which categorizes DR to five severity levels based on the type and number of detected lesions on one FOV retinal fundus image. The studies related to automated diagnosis of DR are categorized in the scope of CAD solutions which could include traditional image processing, ML and DL approaches. Generally, recent research of CAD solutions emphasizes on ML and DL and a combination of these two, however DL solutions proved to have higher performance in terms of sensitivity and AUROC [16], capable of learning high-level features in addition to low level features and transferable to new imaging system. We will further analyze the most relevant DL solution in addressing DR diagnosis in the next chapter.

Chapter 2

Deep learning as a promising approach in DR diagnosis

Based on:

- **Kheradfallah, Hoda, Balaji, Janarthanam Jothi, Jayakumar, Varadharajan, Abdul Rasheed, Mohammed and Lakshminarayanan, Vasudevan. "Annotation and Segmentation of Diabetic Retinopathy Lesions: An Explainable AI Application" SPIE Journal of Imaging, (2022):502-511.**

2.1. Introduction

In general, DL research on DR can be categorized into two approaches: lesion segmentation and image-based grading [1]. In image-based grading, retinal images will go through a DL model that is trained for classification. The reference labels per image come from the prior image-wise clinical annotations. In lesion segmentation, the DL model is trained on a certain lesion and will detect that lesion on images based on the learnt features and then, the information about lesions such as type, number, size and location are determined [17].

DL approaches can be divided into several common branches including convolutional neural networks (CNNs), autoencoders (AEs), recurrent neural networks (RNNs) and deep belief networks (DBNs) [18]. Among these approaches, CNNs are the most common DL paradigm to address DR diagnosis which will be illustrated in detail [1].

2.2. Deep Convolutional Neural Networks

CNNs include some interconnected layers of neurones similar to human visual system. Its applications span computer vision, robotics, financial and weather forecasting, and text analysis through neural language processing [19]. Three main types of CNN layers are: convolutional, pooling and fully connected (FC). Convolutional layers contain filters that convolve with the original image to extract local features related to the filter. The pooling layers reduce the size of feature maps and adds generalizability to the model. The FC layers are also connected to all previous filter layers to combine and correlate their extracted information [20].

The number of convolutional windows in a layer and filter sizes are the adjustable parameters in a convolutional layer. The final part of a convolutional layer is the activation function which controls the neuron's activation through a unique mapping and a trained firing threshold. Trained weights are multiplied by the output of the previous neuron and accumulated as the input of the next layer to form neural connections to neighbouring layers. Depending on the type of problem (classification or segmentation), the generated CNN predictions are either probabilities of each class or probabilities of being a part of an object.

A CNN model is trained using backpropagation in an end-to-end manner by learning the hierarchy of features automatically [21]. These models can also come with the previously trained weights and have the capability to be optimized for various applications through transfer learning. Commonly used CNN models have multiple layers called deep CNNs (DCNNs) including VGGNet, ResNet and Inception modules.

In this study, we initially started with a pretrained DenseNet121 over ImageNet dataset and transferred this model to fit for fundus image classification in ICDRS system [22]. This model was fine-tuned with the Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 database [22]. Initially, the APTOS dataset was applied in the blindness detection challenge. This dataset with 3662 retinal fundus images was taken with various imaging devices [23]. The images are graded manually on ICDRS. The number of retinal images available in the dataset per severity level is presented in Table 2.1. According to Table 2.1 most of the images are normal and very few images belong to sever NPDR. Hence, as a preprocessing step, this class imbalance in the dataset is resolved by randomly down sampling to an equal value in each class.

Severity level	Number of samples
Normal	1805
Mild-NPDR	370
Moderate-NPDR	999
Severe-NPDR	193
PDR	295

Table 2.1. Distribution of the APTOS 2019 dataset per severity level [24].

During the fine-tuning phase with the APTOS database, 80% and 20% of the data were used for training and validation, respectively. In the training phase, the loss function is cross entropy (CE) loss, optimizer is Adam, learning rate is 0.0001, batch size 16 and training was done in 20 epochs.

Figure 2.1 Shows the image classification model with detailed architecture of DenseNet121 model [24]. In this research, computations are all done on a 6 core Intel core i7 CPU at 2.6 GHz with 32GB RAM. The final mean classification accuracy, precision and recall scores over all DR severity levels was 73%, 70% and 68%, respectively.

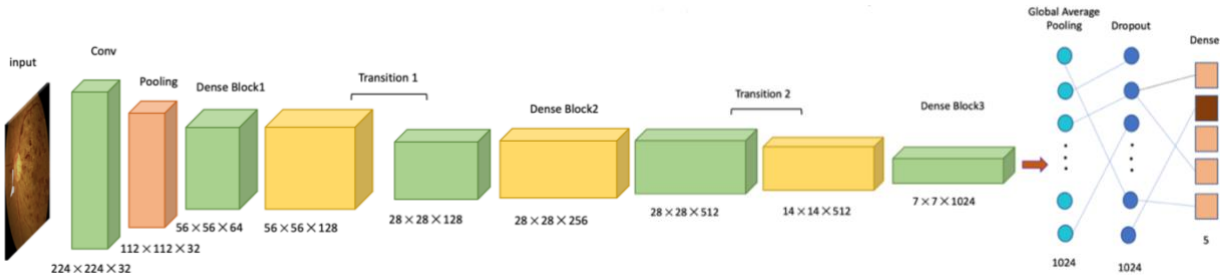

Figure 2.1. DenseNet121 architecture applied for the DR classification task.

A key barrier to the adoption of DCNNs for clinical applications in ophthalmology is the lack of medically supportable reasons behind decisions [2]. This makes it difficult to trust the medical system, both at clinical and health regulator

levels. Hence, they require either additional tools to explain their predictions or modification of the architecture to be intrinsically interpretable.

2.3. Discussion

In the previous sections, an overview of the screening techniques and the concept of deep convolutional networks were presented. The major difference between traditional CAD and deep learning methods is that the feature maps generated automatically with DL make the preparation phase easier compared to the manual feature engineering in traditional CAD algorithms. As computing systems improved and their computational power and capacity increased, the application of DL solutions in various topics became more feasible.

DL also escalated the automated diagnosis of retinal diseases, considerably. The major applications of DL on retinal disease include classification of AMD, DME, and DR as well as segmentation of retinal lesions, optic disc, and vessels [25]. Not limited to classification and segmentation, DL could be also used for denoising, image generation and super-resolution tasks [26]. Various types of neural network designs, such as CNN, encoder-decoder and generative adversial network (GAN) have been well performed on a wide variety of tasks [2]. The selected classification method we applied to classify APTOS 2019 database images is DenseNet121. Previously, this DCNN was applied by Chaturvedi et al.[22], on the whole database images without down sampling and using different data preprocessing reported unweighted mean f1, precision and recall values of 70%, 67% and 75%.

The major problem with the DCNN family is the difficulty of reasoning verification which is an essential step before its practical application for ophthalmological applications. The next section introduces the concept of explainability and its application to the DL model used for DR grading task.

Chapter 3

Explainable AI and its application in evaluating DR detection networks

3.1. Introduction

DCNNs are the most recently used CAD tools for the image analysis studies [1]. In the ophthalmology domain, DCNNs proved to have comparable performance to human experts in disease classification [13].

Despite their emergence as successful assistance in retinal image diagnosis, adoption for clinical applications requires further verification [11]. The main obstacles that cause a lack of trust in them are their black-box nature and the complexity of their models [28]. Complicated internal connections in deep models, on the other hand, can result in high-accuracy disease detection but cannot explain the logic behind their decision.

Simple ML models such as decision trees and k-nearest neighbors (KNNs) are self-explanatory as the decision boundary used for classification is visualizable [28]. But these simple ML models lack the required complexity for tasks such as the classification of 3D and most 2D medical images which include a high volume of detailed information [27].

The lack of tools to inspect the black-box behaviour of DCNN models creates a barrier to the application of deep learning in all domains where explainability, transparency and reliability are required to trust model decisions. Currently, newer regulations like the European General Data Protection Regulation (GDPR) are proposed which make it harder for the use of DCNNs in all businesses, including healthcare which requires decision retracability [2, 29].

3.2. Explainable AI methods

As mentioned in section 2.1, DCNNs require a determined amount of explainability to retrace the logic behind a certain decision. Presently, the two terms of interpretability and explainability are used interchangeably due to the lack of determined mathematical formulations [30].

Visualization of what a model focuses on while making a decision can make it trustable. It is also essential to verify critical features with the standard domain-specific features medical professionals apply. To this end, the explainability methods should include some sort of solution to provide reasoning for model decisions [31]. A majority of XAI methods are called attribution-based methods since they compute the contribution value of each image pixel with the model output [2].

In the first phase of this study, we will focus on attribution-based solutions since this group of methods has model invariant approaches and is available with open-source implementations. Hence, DL practitioners can apply these convenient and explainable tools to understand their designed DCNN model independent of the type of the task they use the model for [27].

3.3. Application of XAI in DR grading models

Attribution or relevance value assignment to the input feature of a network can be done in several ways. Attribution methods, in general, determine the contribution of an input feature to the target neuron. In classification problems, the target neuron is the output neuron of the correct class. Attributions of all features are visualized on the input image in the form of heatmaps known as the attribution maps [2]. There are some common types of attribution based XAI solutions that are listed in Table 3.1. Figure 3.1 also shows some of these mentioned using a VGG16 model.

XAI method	Description
Gradient [2]	As the simplest approach to XAI, this method computes the gradient of target neuron output compared to the input.
LRP *z* [32]	Apply backward proportional decomposition of LRP rule of the upper layer relevance value to the previous layers.
ε-Layer-wise relevance propagation (ε-LRP) [33]	Proportional redistribution of the prediction score on the network by adding a small constant value ε to the denominator of LRP rule.
Gradient × input [34]	Multiplication of the signed partial derivative of the output with the input which enhances sharpness of attribution maps.
Class activation map (CAM) and gradient-weighted class activation map (GradCAM) [35]	Uses the gradients of the target neuron as it flows to the final convolutional layer. Only models that end with a global average pooling and FC are eligible for CAM.
Integrated gradient (IG) [36]	The average gradient of target neuro output when the input value changes between the baseline (often zero) to the actual input value.
Deep Taylor [37]	On a certain input, a relevance score is assigned to the output prediction through Deep Taylor decomposition rule and then backpropagated to the input and produce a heatmap of relevance scores.

Saliency maps [27]	By making the least perturbation on input, this approach aims to find the most important features on output predictions. It computes the absolute value of partial derivatives of target and backpropagates to the input.
Local interpretable model agnostic explanations (LIME) [38]	As a perturbation- based model, in each perturbation, LIME turns off some super pixels of the image which are interconnected pixels with similar colors [39].

Table 3.1. Most recently used XAI methods with descriptions [2].

From the above list we applied 7 XAI methods on the trained DenseNet121 model: CAM, GradCAM, LIME, ε-LRP, IG, gradient × input and saliency maps. Figure 3.2 shows the output attribution maps of each method on input image that was initially annotated as moderate NPDR.

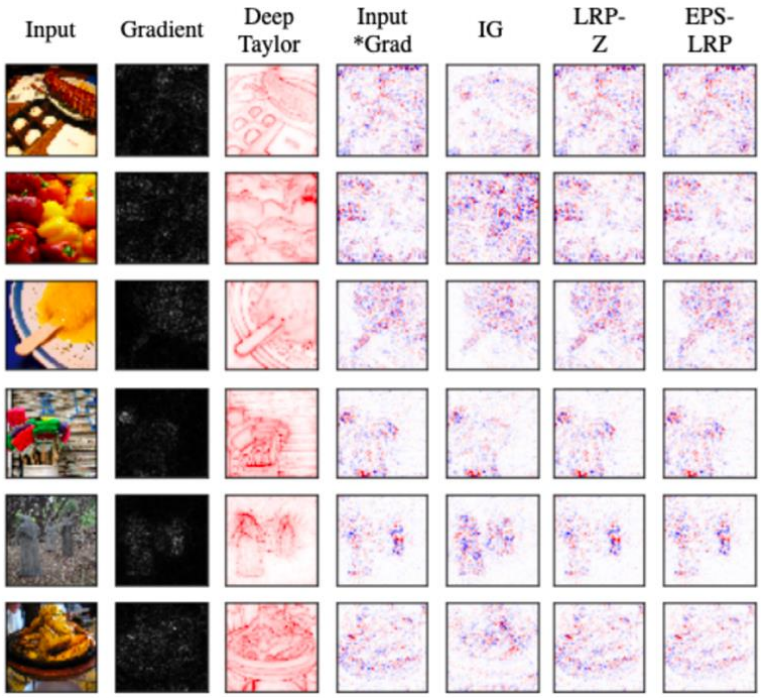

Figure 3.1. XAI maps of different inputs using a VGG16 model. the XAI solutions used to evaluate this classification model are Gradient, Deep Taylor, Input × Gradient, IG, LRP-Z and ε-LRP.

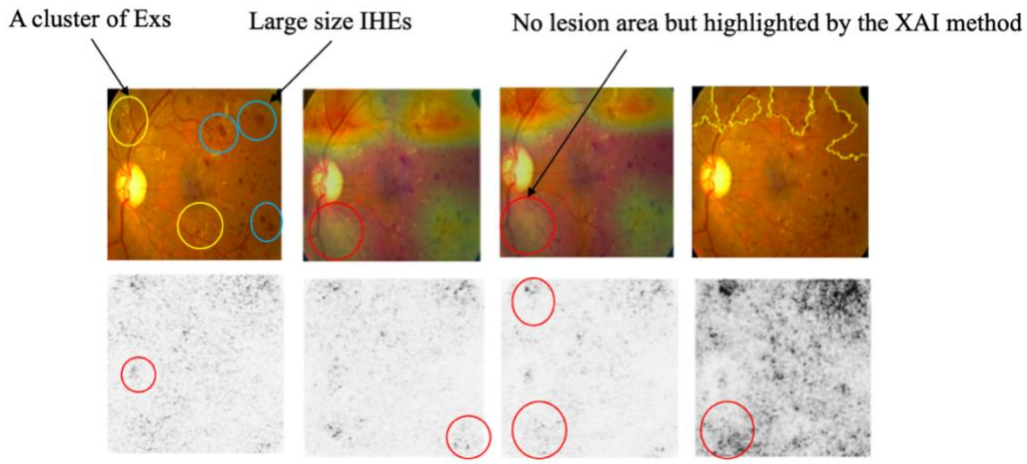

Figure 3.2. XAI maps of the last convolutional layer from a DR fundus image. The images were obtained from the APTOS 2019 dataset. The images on the top row from left to right are as follows: original, CAM, GradCAM, LIME, and on the bottom row from left to right are epsilon LRP, integrated gradient, gradient×input and saliency map. The yellow shapes on the original image around the lesions point to exudates and blue shapes point to the IHE regions which are more significant than other regions. The red shapes also show the regions that the method has highlighted, but there is no lesion.

On Figure 2.3, the first three XAI methods (CAM, GradCAM and LIME) provide generic explanation and miss the lesions that exist on the bottom right side of the image. The bottom four methods are also sparce and highlight the regions that do not have any lesion.

3.4. Discussion

DL is a promising CAD method in DR diagnosis. Despite remarkable performance results in the medical domain, DL methods are not widely deployed in clinical applications [11]. The reason behind this fact is the internal connections in DCNN architectures and the high volume of contributing parameters. Hence, the first approach of adding explainability to DCNN methods is addressed by XAI solutions. Some XAI methods can visualize DCNN models independent of the internal architecture of the model since they are applied at the final step of prediction [2].

Among XAI methods we selected seven methods and applied them to the proposed DCNN model. The outputs of these models were either heatmaps that highlighted the most important regions of the image that indicated DR or provided some sparse and not meaningfully related attribution maps. These two fundamental shortcomings of XAI methods indicate their inability to give sufficient medically meaningful information such as number and size of detected lesions that could be verifiable with standard grading criteria. Hence, we propose an alternative solution based on lesion segmentation that will be discussed more in section 4.

Chapter 4

Retinal lesion segmentation database

Based on:

- **Kheradfallah, Hoda, Balaji, Janarthanam Jothi, Jayakumar, Varadharajan, Abdul Rasheed, Mohammed and Lakshminarayanan, Vasudevan. "Annotation and Segmentation of Diabetic Retinopathy Lesions: An Explainable AI Application" SPIE Journal of Imaging, (2022):502-506.**

4.1. Introduction

Lesion segmentation is the second application of DCNNs in medical image analysis and needs databases that have lesion-wise annotation. According to ICDRS, DR has 10 distinct lesions. Some of these lesions including HE, EX, CWS and MA are annotated pixel-wisely in several public databases [40, 41], however, lesions such as NV and FP are not reported previously in any databases. There are some crucial parameters that directly impact the dataset quality such as verification by specialized optometrists or ophthalmologists, high intra-rater agreement on annotations, the precision and quality of annotations and resolution and readability of images which are further illustrated in chapter 4.2.

4.2. Existing DR lesion annotation datasets

In the scope of DR image analysis, lesion annotation could be done with these two approaches: pixel-wise lesion annotation over the whole image and labelling lesions. Hence, several datasets are publicly available as sorted in Table 4.1.

Dataset	# Images	Annotated lesions	Image size	Description
IDRID [40]	81	MA, HE, Ex, CWS	4288×2848	High annotation resolution, Not annotated with clinicians
Retinal Lesions [42]	1593	MA, IHE, VHE, PHE, Ex, CWS, NV, FP	896×896	Low annotation resolution, annotated by 45 ophthalmologists
eOphtha [43]	463	MA, Ex	Multiple sizes	High annotation quality, annotated by ophthalmologists
RC-RGB-MA [41]	250	MA	2595×1944	High annotation quality annotated with MA annotation tool (RC-MAT) by two experts
FGADR [44]	1842	MA, Ex, CWS, IRMA, NV, HE	1280×1280	High annotation quality with ICDRS-based grades

Table 4.1. Existing DR lesion segmentation databases.

This table shows that the databases may be small, as IDRID or not annotated with high resolution, as Retinal Lesions. Small databases make the method limited and biased. Furthermore, Low annotation resolution will cause to have a model that does not detect fine lesions such as MAs in initial stages even the size of a database cannot compensate for this effect. Furthermore, it makes the final severity grade questionable if deciding on the number of an existing lesion. There is also a lack of a database that covers all DR-

related lesions. A method can be used to diagnose disease if it performs well on all DR-related lesions, but there is no database available that can address all requirements.

4.3. FGADR 143-9 as a comprehensive DR lesion annotation database

In this study, we applied the Fine-Grained Annotated Diabetic Retinopathy (FGADR) database that contains two subsets: Seg-set (1842 images) and Grade-set (1000 images). This database is collected from UAE hospitals and six ophthalmologists annotated images with five severity levels according to ICDRS with high intra-rater consistency [44].

FGADR database in addition to image-wise grading offers 6 pixel-wise lesion annotation masks of MA, Ex, CWS, IRMA, NV and the combination of all three hemorrhage types as one lesion. This database does not include annotation of two lesions (FIP and VB) and does not distinguish between the three hemorrhage types. VHE and PHE could be present if DR is proliferated. Hence, we can annotate VHE and PHE together in one lesion mask. In this study, on a set of 143 images of FGADR database we added the annotation of VB, FIP, IHE and VHE - PHE and have complete set of annotations on this subset. The distribution of selected images based on the disease severity is shown on Table 4.2. To have sufficient evidence per lesion in our subset we also considered the number of images that contain each lesion as Table 4.3 presents.

The annotation part of our study is done by one specialized optometrist from India, JJB, as the annotation reference and two optometrists from Canada, MAR and VJ [45].

We had multiple equalization meetings about the features of each lesion and the technical points such as the monitor resolution and the image size during annotation which is the scale of 100% of original size (1280 × 1280 pixels). By consensus, we agreed to do annotation with ImageJ software and use free hand selection tool for all lesion annotations except VB which needs to annotate the Region of Interest (ROI) of vessel parts containing VB with the paintbrush tool. For the intra-rater variation evaluation, 10 samples of FGADR database were selected such that the four lesions were present in at least two of them, and they were assigned to three annotators equally.

Next, the quality of annotation compared with the reference and among other annotators was evaluated in terms of Dice coefficient, Jaccard index and pixel accuracy. The annotation and evaluation results of equalization phase were available for the group members to do comparisons or verification during the main annotation phase. Furthermore, specific agreements were made on annotating the critical lesions such as VB to enhance inter-rater consistency over the same 10 samples. For VB, the variation of vessel diameter should be more than 50% in a length of 1 to 3 times of vessel diameter

were considered as being VB. The final mean scores of three annotators are listed in Table 4.4. The whole data selection and annotation process is shown on Figure 4.1.

DR Grade	Number of images in the cluster
0, No DR	3
1, mild NPDR	9
2, moderate NPDR	24
3, severe NPDR	41
4, PDR	66

Table 4.2. Distribution of selected images per DR grade according to ICDRS.

Lesion Name	Number of qualified images for segmentation model	
	Train and evaluation	Test
MA	58	6
IHE	61/49	6
VPHE	31/79	7
NV*	12 / 35	6
VB	41/66	6
CWS*	34 / 206	11
Ex	66	4
IRMA*	22 / 57	5
FP*	31 / 78	5

*Table 4. 3. Number of qualified images used for each lesion detection. The lesions that are marked with * initially used a subset of our 143 images (determined before "/") and then added additional samples from the original FGADR set (after "/").*

Lesion type	Dice score	Jaccard Index	Pixel Accuracy
IHE	69%	12%	75%
VPHE	78%	15%	94%
VB	23%	6%	51%
FIP	57%	12%	84%

Table 4.4. inter-rater agreement evaluation using Dice score, Jaccard index and IOU score.

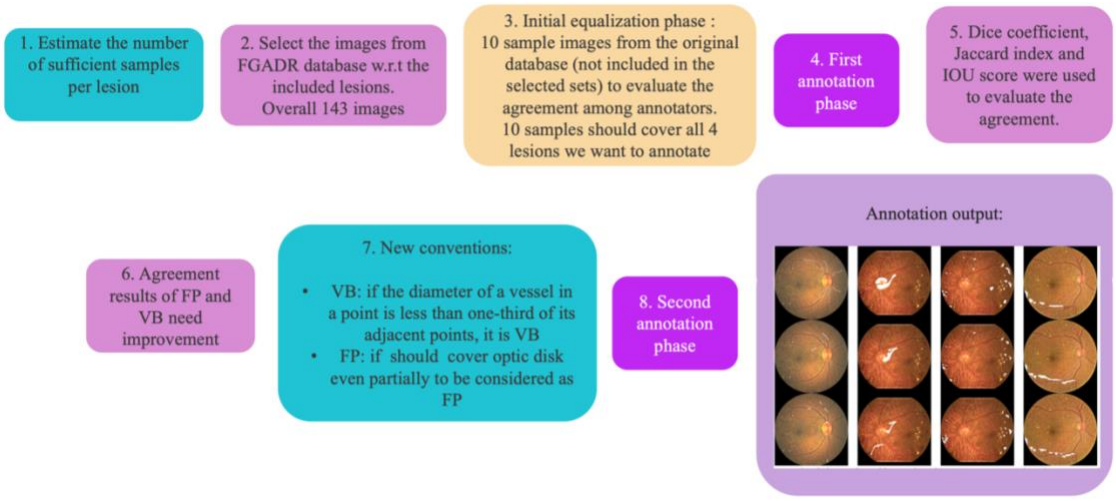

Figure 4.1. Block diagram of data annotation phase.

4.4. Discussion

In this study, the lesion segmentation is done upon FGADR database samples. The reason behind choosing this database as our reference are: the database images are graded based on ICDRS criteria, images all have equal resolution and are partially annotated with high precision on HE, MA, Ex, IRMA, NV and CWS.
Hence, we added the annotation of FP, VB, IHE and VPHE to a subset of 143 images of this database. Based on annotation evaluation metrics, the annotation of VB needs another type of annotation. This process is done by 3 optometrists with at least 5 years of experience. The database is publicly available on Github [1]and through email[2].

[1] https://github.com/hoda213/FGADR-143-9.git
[2] hkheradf@uwaterloo.ca

Chapter 5

Retinal lesion segmentation toolbox

Based on:

- **Kheradfallah, Hoda, Balaji, Janarthanam Jothi, Jayakumar, Varadharajan, Abdul Rasheed, Mohammed and Lakshminarayanan, Vasudevan. "Annotation and Segmentation of Diabetic Retinopathy Lesions: An Explainable AI Application" SPIE Journal of Imaging, (2022):507-511.**

5.1. Introduction

In the scope of DR-related lesion segmentation, there are some successful DCNN models that work well on certain lesions and have the potential to improve in architecture and be extended to further lesions. Son 2018 [46] proposed a modified UNet pipeline named VRT that achieves Area Under the Precision Recall Curve (PR-AUC) scores of 0.71, 0,68, 0.49, 0.69 on Ex, HE, MA and CWS respectively. According to the IDRID challenge website, Liu et al. 2020 [47] applied a combination of DenseNet and the dilation block of UNet called PATech. PATech has PR-AUC scores of 0.88, 0.64 and 0.47 on Ex, HE and MA which has better performance in Ex, but lower overall scores compared to VRT.

Furthermore, Wang et al. 2020 [47] proposed another UNet-based pipeline, IFLYTEK-MIG, for DR lesion segmentation. IFLYTEK-MIG has almost similar performance in Ex pixel-wise segmentation to PATech and like VRT on MA, but it performs weaker than VRT and PATech on HE and CWS with 0.55 and 0.65 PR-AUC scores, respectively.

Xue et al. 2019 [48] applied Mask-RCNN which maintains the information of previous layers in its residual architecture upon IDRID and eOphtha databases. They focused on the pixel-wise segmentation of MA and Ex and using IDRID, reported the sensitivity values of 76.4% and 77.9%, respectively.

On the IDRID, Xiao et al. 2019 [49] proposed a pipeline, incorporating Holistically Nested Edge Detection Network (HEDNet) into Conditional Generative Adversarial Network (HEDNet-cGAN) that applies a class-based GAN loss to the segmentation loss of the HEDNet architecture. This network, in terms of average precision (AP) score on MA, CWS, Ex, and HE received 43.92%, 48.39%, 84.05% and 48.12%, respectively. However, the results are not as good as some prior approaches on Ex and HE, but the pipeline overall has high average performance on all IDRID lesions, less implementation cost, and accessible results and technical details.

5.2. Proposed DR lesion segmentation toolbox

To obtain lesion segmentation, DCNN architectures have proved to be the highest performing method. Xiao et al. compared HEDNet-cGAN with two DCNN architectures: UNet, modified HEDNet. Initially, the HEDNet model was used as a successful edge detector for semantic segmentation purposes. According to Xiao et al.'s results, on all IDRID lesions, MA, CWS, Ex and HE, HEDNet outperformed UNet with AP scores of 44.03%, 43.07%, 83.98% and 45.69% compared to UNet AP scores of 41.84%, 42.22%, 79.05% and 41.93%, respectively. A comparison between HEDNet and

HEDNet-cGAN, showed that the HEDNet-cGAN resulted in a higher AP on CWS, Ex and HE, however, the time and computational cost of this model indicate that it is not efficient. The segmentation output of the three models based on Xiao et al's study are shown in Figure 5.1.

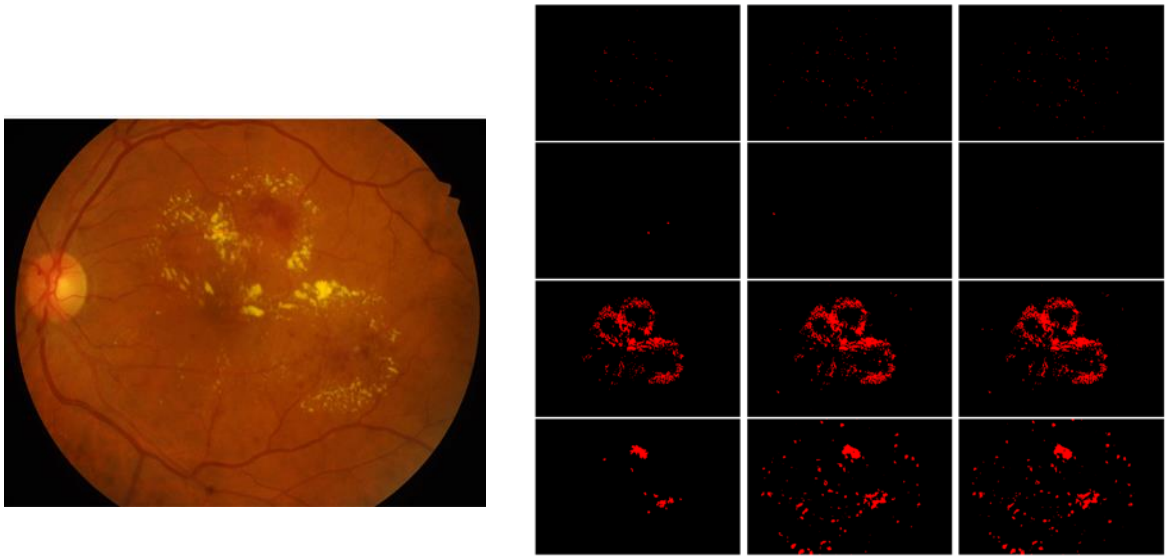

Figure 5.1. Left: a sample image of IDRID reported by Xiao et al. [50]. Right: the segmentation masks on each row from top to bottom belong to MA, CWS, EX, and HE. Each column from left to right shows Ground truth, the segmentation output of HEDNet and the output of HEDNet-cGAN.

The complete HEDNet architecture is shown in Figure 5.2. In this study, HEDNet is applied with modified VGG 16 backbone. The backbone is pre-trained with ImageNet and the weights are loaded in the segmentation model. This helps the model to improve better during training since it is a deep model it will be difficult to start with all random connection weights. After upsampling to the original image size, the model could get five side outputs that their difference indicates the prediction performance variation between convolutional layers.

During implementation, we divided the training set to 80% and 20%. This 20% is randomly selected as the validation set. As the data preprocessing on both train and test sets, three tools could be applied: 1. brightness balancing, 2. contrast enhancement through CLAHE [50] method with the grid size of 8×8 used for histogram equalization. 3. Image denoising. In Table 5.1, we put the exact preprocessing steps per lesion.

Lesion Name	Applied preprocessing steps		
	Brightness balance	CLAHE	Denoising
MA	✓	✓	✓
IHE	-	✓	✓
VPHE	-	✓	✓
NV	✓	✓	✓
VB	-	✓	✓
CWS	-	✓	✓
Ex	✓	✓	✓
IRMA	✓	-	✓
FP	-	✓	✓

Table 5.1. The preprocessing tools used per lesion

To have sufficient samples to train the model, the data augmentation is applied on the preprocessed training images. The optimum steps for augmentation and preprocessing are adjusted experimentally and depending on the total lesion set properties. Hence, as the data augmentation steps, random rotation with maximum 20-degree, random crop to the size of 256 × 256, and normalization of image colors based on each lesion set's properties (through measuring the mean and standard deviation of each lesion set). The loss function is CE and the wight of lesion pixels to non-lesion pixels is set to 10 according to equation 1:

$$BCE\ Loss = -(w\ gt \log p\ +\ (1 - gt) \log(1 - p)) \quad (1)$$

where *w* is the weight of positive lesion prediction, *gt* is the ground truth label of a certain pixel and *p* is the prediction on the same pixel. The final hyperparamer values is mentioned in Table 5.2 for which the optimizer is SGD. The weight decay of SGD optimizer acts as a regularizer that could improve efficient training [51]. The images are given in the batches of 4 to train the model and 5-fold cross validation is used to get more accurate test error estimates.

This phase of our study including training segmentation model and performance evaluations, were executed on a Intel Core i7 9700K 3.60GHz 8 core CPU, 64GiB RAM, Nvidia Titan V/PCIe/SSE2, 12GiB GPU and the training of each model takes 14400 seconds on average. All the codes and implementation details will be accessible on Github[3].

[3] https://github.com/hoda213/FGADR-143-9.git

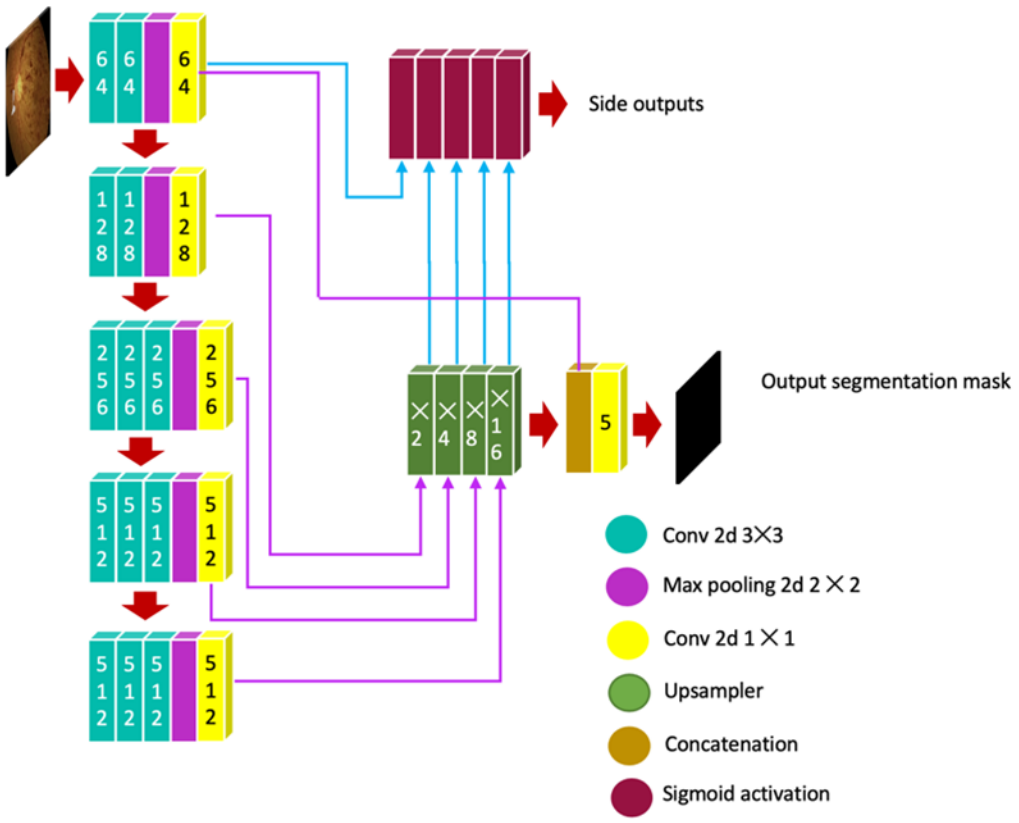

Figure 5.2. HEDNet model architecture based on VGG16 backbone. The numbers on the convolutional blocks are the number of corresponding output channels, and after each 3 ×3 convolutional block there is an activation layer. The output of each sigmoid channel indicates one side network output.

5.3. Quantitative performance evaluation of segmentation models

The metrics applied to evaluate performance were PR-AUC and mean AP (mAP) over test samples. Table 5.2 shows the performance scores of all nine models in terms of PR-AUC and mAP on four test images. Our results, show that HEDNet works well on segmentation of IHE, VPHE, Ex, CWS and NV and outperforms the results of Xiao et al's study [49]. Our model needs improvement on MA since there is a gap of 18% with the results reported by Xiao et al. Lesions such as IRMA and VB that are related to vessel abnormalities and FP could not be detected properly in terms of mAP and PR-AUC metrics. Figure 5.3 shows the segmentation results of all nine HEDNet models on a random sample image. Figures 5.4, 5.5 show the Precision-Recall curves of the study done by Xiao et al. [49] and our results, respectively. The training loss variation per epoch is also shown on Figure 5.6.

Lesion model	Hyperparameter values					
	Learning rate	Weight decay	Momentum	Loss function, Weights(lesion/non-lesion)	Epochs	K-Fold cross-validation
MA	$1\ e^{-3}$	$1\ e^{-3}$	0.93	CE (20)	80	5
IHE	$1\ e^{-4}$	$5\ e^{-3}$	0.9	CE (10)	100	5
VPHE	$1\ e^{-3}$	$5\ e^{-3}$	0.93	CE (10)	100	5
NV	$1\ e^{-3}$	$5\ e^{-3}$	0.93	CE (10)	100	5
VB	$1\ e^{-3}$	$1\ e^{-4}$	0.93	CE (20)	60	5
CWS	$1\ e^{-3}$	$5\ e^{-3}$	0.9	CE (10)	200	5
Ex	$1\ e^{-3}$	$5\ e^{-3}$	0.9	CE (10)	200	5
IRMA	$1\ e^{-5}$	$1\ e^{-2}$	0.93	CE (20)	200	5
FP	$1e^{-4}$	$1\ e^{-2}$	0.9	CE (20)	60	5

Figure 5.2. Hyper parameter values per lesion model.

Lesions	mAP %	PR-AUC %	AP% results of Xiao et al.[45] using HEDNet
MA	26	41	44.03
IHE	53	59	45.69
VPHE	61	86	45.69
NV	43	56	-
VB	19	34	-
CWS	48	51	43.07
Ex	78	93	83.98
IRMA	21	56	-
FP	22	34	-

Table 5. 3. Performance scores of each lesion's segmentation model over four of the test images of each lesion set. The lesions such as VB, IRMA and FP are detected with both low mAP and PR-AUC results and require improvement.

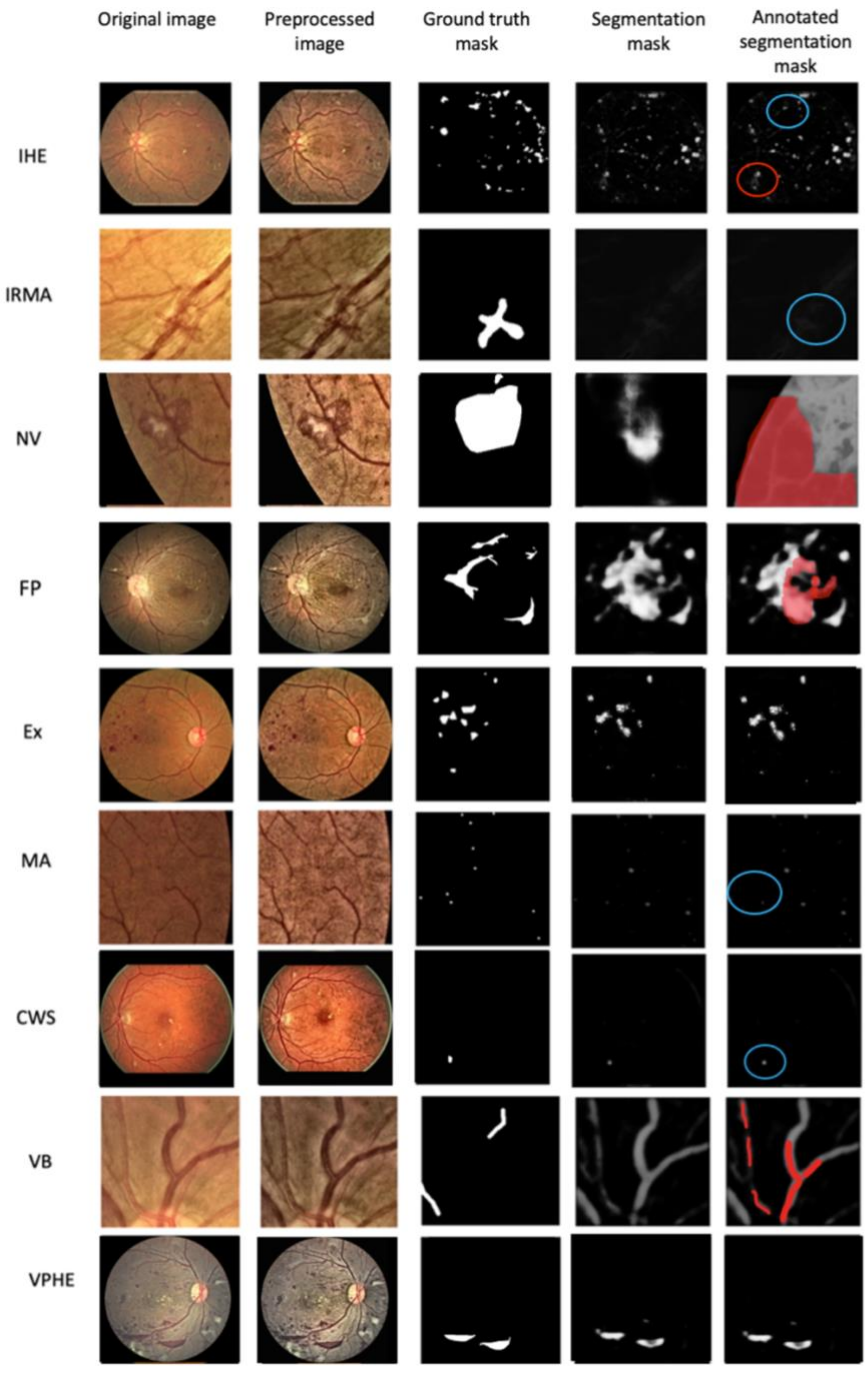

Figure 5.3. Segmentation outputs of all nine HEDNet models for each lesion on random test images. The columns from left to right belong to original images, preprocessed images (color enhanced with CLAHE), ground truth annotations model predictions and highlighted segmentation mask, respectively. In the last column, blue marks show the false negative predictions and red marks show the false positive predictions.

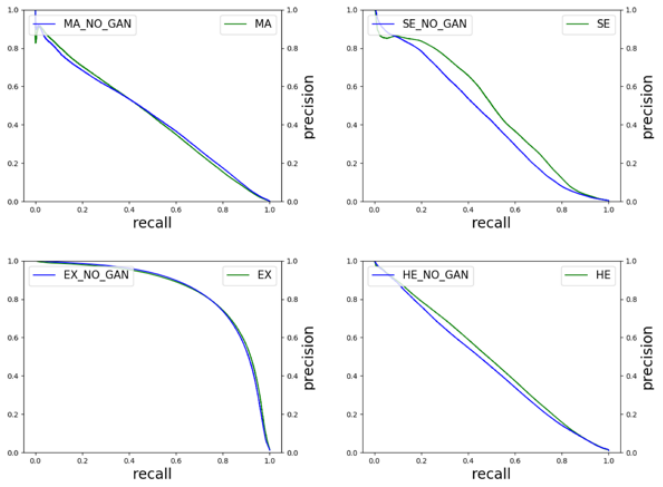

Figure 5.4. Precision Recall curves of Xiao et al work. The charts on top row from left belong to MA and CWS. On the bottom row from left the charts belong to EX and HE (combination of IHE and VPHE). The blue and green lines show the HEDNet and HEDNet-cGAN curves.

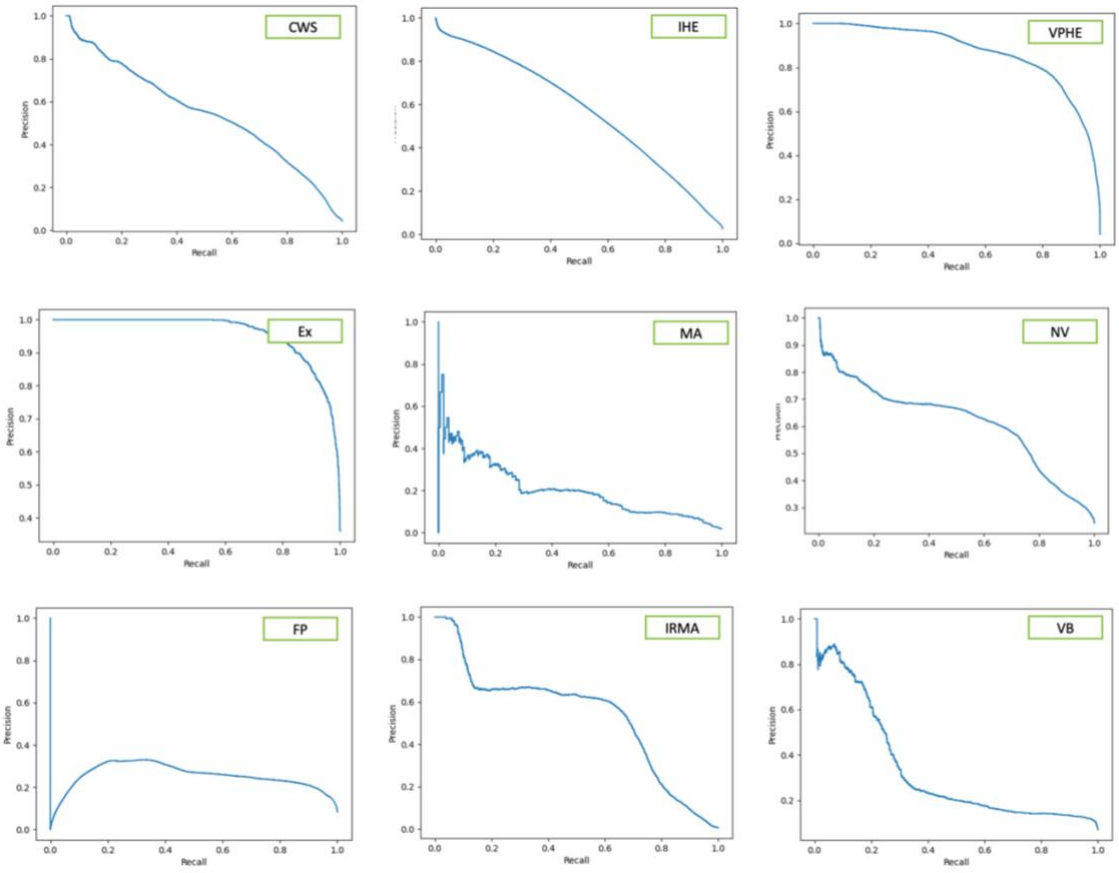

Figure 5.5. Precision Recall curve of the predictions shown in Figure 5.3 using HEDNet model.

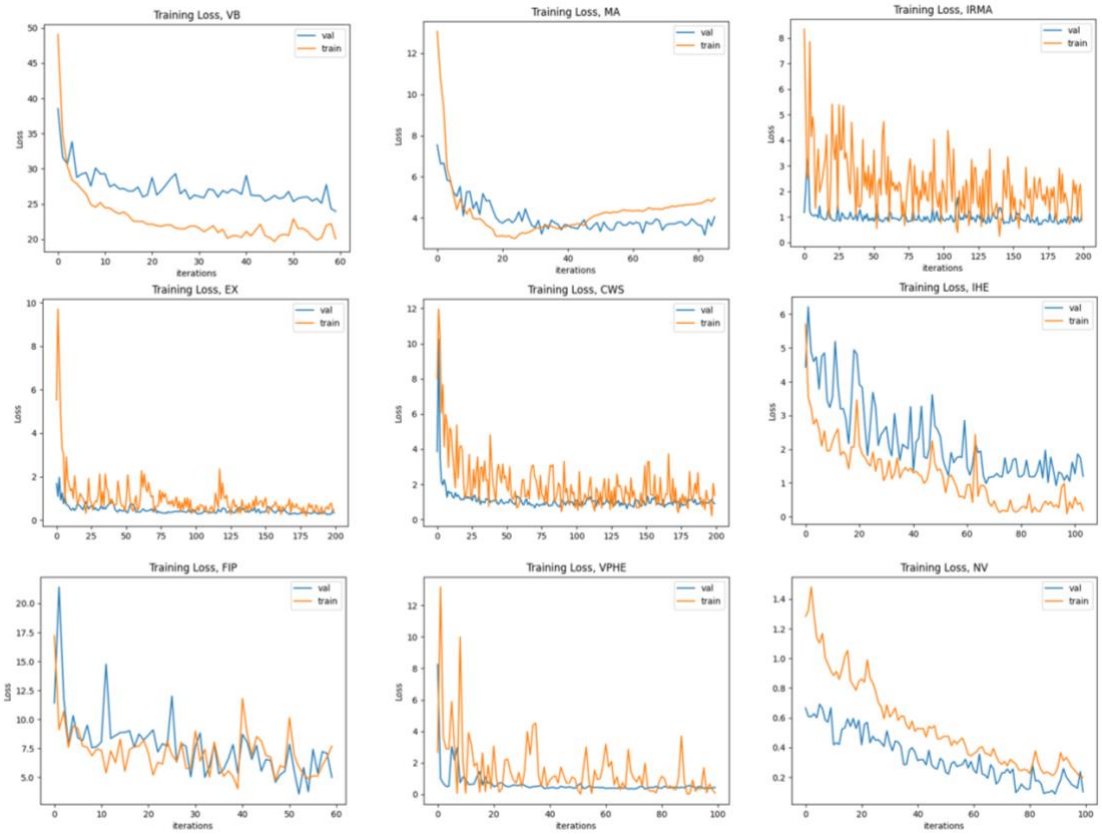

Figure 5.6. Training loss variation per epoch.

5.4. Discussion

In our proposed approach, we prepared the first comprehensive segmentation and grading dataset that covers all DR- associated lesions with high annotation quality and intra-rater agreement. This study offers a unique perspective on DR grading which is the most comparable to the clinical decision-making process. According to ICDRS as the reference grading system, we planned to grade DR based on the type and number of detected lesions compared to their range in ICDRS. The side outputs of the HEDNet model make verification of the segmentation process easier compared to a classification network. The classification part also would no longer need additional explainable tools due to its compatibility with ICDRS considerations.

The reason behind the model's inability to segment VB, IRMA, MA and FP, according to Table 5.2, could be the limited number of samples or the visual variation of some of these lesions in our dataset, so the model cannot find a unique morphological feature to detect these lesions. Furthermore, the anatomical semantics around retinal vessels are quite complicated. Perplexing structures and areas in retinal fundus images, including optic disc regions, pathological areas, hemorrhage, exudates, and low image contrast in some areas between damaged vessel and background, may easily result in false segmentation of abnormal vessels and lesions.

Chapter 6

Conclusion and future steps

Currently, DCNNs have proved to be the most promising approach in DR diagnosis and grading, which has comparable performance to clinicians in terms of accuracy, sensitivity, and specificity. The major barriers to the clinical application of DCNNs are the black box behavior and the inability of DCNN models to explain how the network comes to a certain decision due to complicated architecture and internal connections.

One approach to remove this barrier is to apply XAI methods that could be categorized to three types based on how to correlate the original image and output predictions. Initially, seven common XAI methods were selected from each of the three categories. The output attention maps show two main problems about the methods: the answers were either general or sparse. We found that the lack of detailed information on three of them and sparsity and high sensitivity of four of these methods results in less validity of produced attention maps for ophthalmology application.

Our solution, as an interpretable approach, is like a clinician's diagnosis process using fundus images: first look for the existing lesions and their significance, and next grade disease severity based on a standard grading system. Our optometrist group added annotation of lesions such as FP, IHE, VPHE and VB on a set of 143 images obtained from the FGADR fundus database. Then, the DCNN model used for segmentation, which is HEDNet, was trained over each lesion to segment all nine lesion masks, separately.

Our models on IHE, VPHE, CWS, MA, NV and Ex have higher performance in terms of mAP and PR-AUC than the models tuned for FP, VB, and IRMA. The reason

behind it might be the diversity of morphological features that the latest lesions could have or that other annotation methods should be used. For instance, on VB, we annotated the beading part with a line. IRMA in the collected database has a wide variety of morphological shapes and is not as easily distinguishable from HE and NV as Ex is.

One further modification that might help is to apply hierarchical networks to segment graphs which are vessel-related patterns in this case, such as IRMA, and VB. Aside from that, we should continue covering more images in our dataset that will add to the evidence and could directly affect model performances. Consequently, with sufficiently high segmentation performance over all DR lesions, we can move to the grading phase with ICDRS considerations.

References

1. Lakshminarayanan, V., et al., *Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey.* Journal of Imaging, 2021. **7**(9): p. 165.
2. Brar, A.S., *Explainable AI for retinal OCT diagnosis*. 2021, University of Waterloo.
3. Qureshi, I., J. Ma, and Q. Abbas, *Recent development on detection methods for the diagnosis of diabetic retinopathy.* Symmetry, 2019. **11**(6): p. 749.
4. Teo, Z.L., et al., *Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis.* Ophthalmology, 2021. **128**(11): p. 1580-1591.
5. Group, E.T.D.R.S.R., *Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10.* Ophthalmology, 1991. **98**(5): p. 786-806.
6. Wilkinson, C., et al., *Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales.* Ophthalmology, 2003. **110**(9): p. 1677-1682.
7. Gangwani, R.A., et al., *Diabetic retinopathy screening: global and local perspective.* Hong Kong Medical Journal, 2016.
8. Fujimoto, J.G., et al., *Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy.* Neoplasia, 2000. **2**(1-2): p. 9-25.
9. Salz, D.A. and A.J. Witkin, *Imaging in diabetic retinopathy.* Middle East African journal of ophthalmology, 2015. **22**(2): p. 145.
10. Abramoff, M.D. and M. Niemeijer, *Mass screening of diabetic retinopathy using automated methods*, Teleophthalmology in Preventive Medicine. 2015, Springer. p. 41-50.
11. Abràmoff, M.D., et al., *Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning.* Investigative ophthalmology and visual science, 2016. **57**(13): p. 5200-5206.
12. Savoy M. *IDx-DR for diabetic retinopathy screening*. American Family Physician, 2020. **101**(5): p. 307-308.
13. Bhaskaranand, M., et al., *The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes.* Diabetes technology and therapeutics, 2019. **21**(11): p. 635-643.
14. Majumder, S., et al. *A deep learning-based smartphone app for real-time detection of five stages of diabetic retinopathy*. in *Real-Time Image Processing and Deep Learning 2020*. 2020. International Society for Optics and Photonics.
15. Bilal, A., et al., *Diabetic retinopathy detection and classification using mixed models for a disease grading database.* IEEE Access, 2021. **9**: p. 23544-23553.
16. Wu, J.-H., et al., *Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis.* Journal of medical Internet research, 2021. **23**(7): p. e23863.
17. Dai, L., et al., *A deep learning system for detecting diabetic retinopathy across the disease spectrum.* Nature communications, 2021. **12**(1): p. 1-11.

18. Das, S.K., P. Roy, and A.K. Mishra, *Deep learning techniques dealing with diabetes mellitus: a comprehensive study*, in *Health Informatics: A Computational Perspective in Healthcare*. 2021, Springer. p. 295-323.
19. Chen, Y., *Convolutional neural network for sentence classification*. 2015, University of Waterloo.
20. O'Shea, K. and R. Nash, *An introduction to convolutional neural networks.* arXiv preprint arXiv:1511.08458, 2015.
21. Asiri, N., et al., *Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey.* Artificial intelligence in medicine, 2019. **99**: p. 101701.
22. Chaturvedi, S.S., et al., *Automated diabetic retinopathy grading using deep convolutional neural network.* arXiv preprint arXiv:2004.06334, 2020.
23. Bodapati, J.D., et al., *Blended multi-modal deep convnet features for diabetic retinopathy severity prediction.* Electronics, 2020. **9**(6): p. 914.
24. Ji, Q., et al., *Optimized deep convolutional neural networks for identification of macular diseases from optical coherence tomography images.* Algorithms, 2019. **12**(3): p. 51.
25. Sengupta, S., et al., *Ophthalmic diagnosis using deep learning with fundus images–A critical review.* Artificial Intelligence in Medicine, 2020. **102**: p. 101758.
26. Kaji, S., et al., *Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging.* Radiological physics and technology, 2019. **12**(3): p. 235-248.
27. Singh, A., S. Sengupta, and V. Lakshminarayanan, *Explainable deep learning models in medical image analysis.* Journal of Imaging, 2020. **6**(6): p. 52.
28. Abd AL-Nabi, D.L. and S.S. Ahmed, *Survey on classification algorithms for data mining: comparison and evaluation.* International Journal of Computer Engineering and Intelligent Systems, 2013. **4**(8): p. 18-27.
29. Holzinger, A., et al., *What do we need to build explainable AI systems for the medical domain?* arXiv preprint arXiv:1712.09923, 2017.
30. Linardatos, P., V. Papastefanopoulos, and S. Kotsiantis, *Explainable ai: A review of machine learning interpretability methods.* Entropy, 2020. **23**(1): p. 18.
31. Mahbooba, B., et al., *Explainable artificial intelligence (xai) to enhance trust management in intrusion detection systems using decision tree model.* Complexity, 2021. **2021**.
32. Kohlbrenner, M., et al. *Towards best practice in explaining neural network decisions with LRP*. in *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020. IEEE.
33. Bach, S., et al., *On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation.* PloS one, 2015. **10**(7): p. e0130140.
34. Shrikumar, A., et al., *Not just a black box: Learning important features through propagating activation differences.* arXiv preprint arXiv:1605.01713, 2016.
35. Selvaraju, R.R., et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. in *Proceedings of the IEEE international conference on computer vision*. 2017.
36. Sundararajan, M., A. Taly, and Q. Yan. *Axiomatic attribution for deep networks*. in *International conference on machine learning*. 2017. PMLR.
37. Montavon, G., et al., *Explaining nonlinear classification decisions with deep taylor decomposition.* Pattern recognition, 2017. **65**: p. 211-222.

38. Ribeiro, M.T., S. Singh, and C. Guestrin. *" Why should i trust you?" Explaining the predictions of any classifier*. in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 2016.
39. Haunschmid, V., S. Chowdhury, and G. Widmer, *Two-level explanations in music emotion recognition.* arXiv preprint arXiv:1905.11760, 2019.
40. Porwal, P., et al., *Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research.* Data, 2018. **3**(3): p. 25.
41. Team, R., *RC-RGB-MA: RetinaCheck RGB Microaneurysm dataset*. 2016.
42. Wei, Q., et al. *Learn to segment retinal lesions and beyond*. in *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021. IEEE.
43. Decenciere, E., et al., *TeleOphta: Machine learning and image processing methods for teleophthalmology.* Irbm, 2013. **34**(2): p. 196-203.
44. Zhou, Y., et al., *A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability.* IEEE Transactions on Medical Imaging, 2020. **40**(3): p. 818-828.
45. Kheradfallah, H., et al., *Annotation and segmentation of diabetic retinopathy lesions: an explainable AI application.* SPIE Journal of Medical Imaging, 2022. **12033**: p. 502-511.
46. Son, J., et al., *Classification of findings with localized lesions in fundoscopic images using a regionally guided cnn*, in *Computational Pathology and Ophthalmic Medical Image Analysis*. 2018, Springer. p. 176-184.
47. Porwal, P., et al., *Idrid: Diabetic retinopathy–segmentation and grading challenge.* Medical image analysis, 2020. **59**: p. 101561.
48. Xue, J., et al., *Deep membrane systems for multitask segmentation in diabetic retinopathy.* Knowledge-Based Systems, 2019. **183**: p. 104887.
49. Xiao, Q., et al. *Improving Lesion Segmentation for Diabetic Retinopathy using Adversarial Learning*. in *International Conference on Image Analysis and Recognition*. 2019. Springer.
50. Kurosaka, T., et al., *CLAHE (Contrast limited adaptive histogram equalization) image processing to improve the CR Portal Image in radiation therapy.* Igaku Butsuri. Supplement, 2008. **28**(suppl. 2): p. 158-159.
51. Loshchilov, I. and F. Hutter, *Decoupled weight decay regularization.* arXiv preprint arXiv:1711.05101, 2017.