



HAL
open science

Diagnostic automatisé des pathologies de la rétine à l'aide des volumes OCT

Rami Safarjalani

► **To cite this version:**

Rami Safarjalani. Diagnostic automatisé des pathologies de la rétine à l'aide des volumes OCT. Traitement des images [eess.IV]. Université Bourgogne Franche-Comté; Université de Balamand (Tripoli, Liban), 2020. Français. NNT: 2020UBFCK039 . tel-03162545

HAL Id: tel-03162545

<https://theses.hal.science/tel-03162545>

Submitted on 8 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université Bourgogne
Franche-Comté
UBFC



THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE
PREPAREE au Laboratoire ImVia/IFTIM, Univ. Bourgogne Franche-Comté, Le Creusot 71200, France
En association avec le laboratoire LIA, EDST, Univ. Libanaise, Tripoli, Liban

Ecole doctorale n°37
ÉCOLE DOCTORALE «SPIM»

Doctorat de l'instrumentation et informatique de l'image

Par

M. Rami SAFARJALANI

**« Diagnostic automatisé des pathologies
de la rétine à l'aide des volumes OCT »**

**« Application de classification et de détection de
rétinopathie diabétique et d'œdème maculaire »**

Thèse présentée et soutenue à « Le Creusot », le « 15/12/2020 »

Composition du Jury:

Mme Fan YANG	Professeur, Univ. Bourgogne Franche-Comté	Présidente
M. Jocelyn CHANUSSOT	Professeur, Grenoble INP	Rapporteur
Mme Caroline PETITJEAN	Maître de Conférences HDR, Université de Rouen Normandie	Rapporteuse
M. Ahmad SHAHIN	Professeur, Université Libanaise	Examineur
Mme Samia AÏNOUZ	Professeur, INSA de Rouen Normandie	Examinatrice
M. Fabrice MÉRIAUDEAU	Professeur, Univ. Bourgogne Franche-Comté	Directeur de thèse
M. Désiré SIDIBÉ	Professeur, Université d'Evry - Paris Saclay	Codirecteur de thèse

Acknowledgment

This thesis was prepared in the Laboratory ImVia with the IFTIM Team (Functional and Molecular Imaging, Medical Image Processing) under the direction of Prof. Fabrice Meriaudeau, at Univ. Bourgogne Franche-Comté.

First of all, I would like to thank my thesis supervisor, Prof. Fabrice, for welcoming me to his team, for his encouragement, his repetitious help, his relevant advice, and his human spirit. I am also grateful to him for the significant time he has given me, for his support and his understanding. I learned a lot from him and I am honored to work under his surveillance. May he be convinced of my deep respect.

In addition, my appreciation goes to all the professors and doctors who supported me during the accomplishment of this thesis:

To Prof. Desire Sidibe, Co-supervisor of this thesis, for the consideration he brought to me during my thesis. I am very grateful for his aid and his advice.

To Prof. Samia Ainouz, for her care and assistance; her remarks and words are and will be always appreciated. All my appreciation is addressed to her for the trust she has shown in me.

I offer my sincere thanks to Prof. Ahmad Shahin, Professor of Lebanese University for his support and his advice. I will be always appreciated for his assistance.

To Prof. Salah El-Falou and Dr. Mohamad El-Falou for their encouragement and their support.

All my appreciation and gratitude will go also to my mother, for her continuous assistance, her encouragement, her back-up, and her feelings.

I especially thank my brothers and my friends...

Finally, I dedicate this work to my late father.

Remerciements

Cette thèse a été préparée au Laboratoire ImVia avec l'équipe IFTIM (Imagerie Fonctionnelle et Moléculaire, Traitement d'Images Médicale) sous la direction du Pr. Fabrice Meriaudeau, à l'Univ. Bourgogne Franche-Comté.

Tout d'abord, je tiens à remercier mon directeur de thèse, le Pr. Fabrice, de m'avoir accueilli dans son équipe, pour ses encouragements, son aide répétitive, ses conseils pertinents et son esprit humain. Je lui suis également reconnaissant pour le temps considérable qu'il m'a accordé, pour son soutien et sa compréhension. J'ai beaucoup appris de lui et je suis honoré de travailler sous sa surveillance. Qu'il soit convaincu de mon profond respect.

De plus, mes remerciements vont à tous les professeurs et les maîtres de conférences qui m'ont soutenu lors de la réalisation de cette thèse :

Au Pr. Desire Sidibe, codirecteur de cette thèse, pour la considération qu'il m'a apportée lors de ma thèse. Je lui suis très reconnaissant de son aide et de ses conseils.

Au Pr. Samia Ainouz, pour ses soins et son assistance; ses remarques et ses paroles sont et seront toujours appréciées. Toute ma gratitude est adressée à elle pour la confiance qu'elle m'a témoignée.

Je remercie sincèrement le Pr. Ahmad Shahin, professeur à l'Université Libanaise pour son soutien et ses conseils. Je serai toujours apprécié de son aide.

À Pr. Salah El-Falou et Dr. Mohamad El-Falou pour leurs encouragements et leur soutien.

Toute ma reconnaissance et ma gratitude iront aussi à ma mère, pour l'assistance continue, ses encouragements et ses sentiments.

Je remercie particulièrement mes frères et mes amis...

Enfin, je dédie ce travail à mon défunt père.



Titre : Diagnostic automatisé des pathologies de la rétine à l'aide des volumes OCT.

Mots clés : OCT, CNN, DR, DME, GNN.

Résumé :

La principale cause de cécité dans la population pourrait être surtout la détérioration de la rétine causée par les problèmes liés au diabète et la complication du vieillissement. La rétinopathie diabétique (DR) et l'œdème maculaire diabétique (DME) sont les principales causes directes de problèmes de vision chez les citoyens en âge de travailler de la plupart des pays avancés. Le nombre élevé de personnes diabétiques dans le monde indique que le DME et la RD resteront les principaux facteurs de perte de vision partielle ou totale, ce qui affecte la qualité de vie des patients pendant de nombreuses années et menace leur vie. Par conséquent, une détection précoce suivie de procédures de traitement rapide des personnes atteintes de maladies liées au diabète est importante pour prévenir les problèmes optiques et peut réduire le risque de cécité. De plus, les personnes de plus de 50 ans sont exposées à la dégénérescence maculaire liée à l'âge (AMD) qui attaque la rétine. Par conséquent, les chercheurs du monde entier sont attirés par les différences liées à plusieurs maladies rétinienne.

Plusieurs méthodes automatisées utilisant l'AI ont été appliquées pour la détection et le test des maladies rétinienne. Malheureusement, ces

modèles peuvent être confondus avec une incapacité de calcul, ce qui nécessite une intervention supplémentaire de la part de spécialistes. Cette thèse présente une méthode automatique - basée sur des algorithmes de réseaux de neurones d'apprentissage en profondeur - pour détecter DME et DR, ce qui permet de dépasser l'évaluation pratique subjective des ophtalmologistes.

Basé sur "Convolutional Neural Network", un modèle proposé est présenté avec un classificateur soft-max et entraîné de bout en bout pour la classification automatique des images rétinienne de tomographie par cohérence optique (OCT). Ce modèle a la capacité de détecter des caractéristiques permettant d'identifier la DR et le DME en utilisant ces images rétinienne avec une précision et une sensibilité améliorées. De plus, un modèle préformé a été affiné et réformé à l'aide d'un ensemble de données qui a été enrichi à l'aide de "Generative Adversarial Networks" (GAN). Contrairement au diagnostic manuel de la maladie rétinienne basé sur un examen clinique personnel et l'analyse des images OCT, cette méthode a montré la capacité de prédire automatiquement les cas atteints de DME par rapport aux cas sains. Les expériences ont été évaluées sur plusieurs ensembles

de données fournis par différentes institutions.

Le modèle, comparé à d'autres modèles CNN entraînés de bout en bout ou pre-entraînés et affinés, montre des fonctionnalités d'extraction efficaces, avec moins de temps, sur la base d'une étape

de prétraitement efficace des données. Les résultats expérimentaux ont montré une plus grande précision de classification, ce qui est prometteur dans le domaine de la détection précoce des maladies diabétiques pour aider les ophtalmologistes dans les technologies biomédicales.

Title: Automated Diagnostics of Retinal Pathologies Using OCT Volumes.

Keywords: OCT, CNN, DR, DME, GNN

Abstract:

The leading cause of blindness in the population could mostly be the degeneration of the retina caused by the diabetic-related problems and the aging issue. Diabetic retinopathy (DR) and diabetic macular edema (DME) are the main direct causes of vision problems in the labor age citizens of most advanced countries. The elevated number of diabetic people globally indicates that DME and DR will remain to be the principal factor to partial or total vision loss, which affects the lives quality of patients for many years to come and threaten their lives. Therefore, early detection followed by fast treatment procedures of persons with diabetic-related diseases is significant in preventing optical problems and can decrease the risk of blindness.

In addition, people above 50 are exposed to age-related macular degeneration (AMD) disease that hits the retina. Therefore, researchers over the world have attracted to the differences related to several retinal diseases.

Several automated methods using Artificial Intelligence (AI) (varying from traditional computer vision to advanced machine learning algorithms) have been applied for the detection and examination of retinal diseases. Unluckily, these models are able to be mistaken with computational inability, which necessitates additional interference from specialist personal. This thesis presents an automatic method - based on deep learning neural networks algorithms - to detect DME and DR, which allows overstepping the subjective handy evaluation of ophthalmologists.

Based on Convolutional Neural Network, a proposed model is presented with a soft-max classifier and fully trained from scratch for the automatic classification of Optical Coherence Tomography (OCT) retinal images where OCT screening techniques are applied as the current dependable assessment and measurement method to discover the existence of swallow in the retina. This

model has the ability to detect patterns for DR and DME using these retinal images with improved accuracy and sensitivity. Moreover, a pre-trained model has been fine-tuned and re-trained using a dataset that has been augmented using Generative Adversarial Networks (GANs). In opposite to manual retinal disease diagnosis based on personal clinical examination and analysis of OCT images, this method showed the capability to automatically predict DME diseased cases versus healthy cases. The experiments have been

evaluated over several datasets provided by different institutions.

The model, compared to other CNN end-to-end or transfer learned models, shows effective extracting features, with less time consumption, based on an efficient data pre-processing stage. The experimental results showed a higher accuracy of classification which is promising in the field of early detection of diabetic diseases to aid ophthalmologists in biomedical technologies.

Table of contents

Acknowledgment	2
Abstract.....	4
Table of Contents.....	7
List of Acronyms	11
List of Figures.....	15
List of Tables	18
List of Equations	19
Introduction.....	20
1- General context.....	20
2- Thesis objective.....	21
3- Thesis synthetic plan.....	21
Chapter 1	23
Overview of Diabetic Retinopathy pathologies.....	23
1.1. Introduction	24
1.2. Overview of eyeball and retinal-related pathologies	24
1.2.1. The Eyeball.....	24
1.2.2. Eye structure	25
1.2.3. Retinal layers.....	26
1.2.4. Retinal pathologies	27
1.2.4.1. Age-related macular degeneration (AMD).....	28
1.2.4.2. Cataracts	29
1.2.4.3. CytoMegalovirus Retinitis (CMV)	30
1.2.4.4. Retinal Detachment.....	30
1.2.4.5. Glaucoma.....	31
1.2.4.6. Cardiovascular Disease	31
1.2.4.7. Diabetic Retinopathy (DR).....	32
1.2.4.8. Diabetic macular edema (DME).....	34
1.3. Optical Imaging Modalities	36
1.3.1. Fundus Photography	37
1.3.2. Adaptive Optics (AO)	37
1.3.3. Scanning laser ophthalmoscopy (SLO)	38
1.3.4. Photoacoustic microscopy (PAM)	39
1.3.5. Ultrasound Biomicroscopy (UBM).....	40

1.3.6. Magnetic resonance imaging (MRI)	40
1.3.7. Optical Coherence Tomography (OCT).....	41
1.3.8. Optical coherence tomography angiography (OCTA)	45
1.4. Deep Neural Networks:.....	46
1.4.1. Deep learning applications	48
1.5. Conclusion.....	51
Chapter 2	52
Introduction to data pre-processing and classification in machine learning	52
2.1. Introduction	53
2.2. Data pre-processing in ML.....	53
2.2.1. Dataset acquiring:	53
2.2.2. Dataset Cleaning:.....	54
2.2.3. Dataset splitting:.....	54
2.3. Classification tasks in ML.....	54
2.3.1. Unsupervised classification:	55
2.3.2. Supervised classification:	55
2.3.2.1. Standard terms in classification algorithms.....	56
2.3.2.2. Types of classification:.....	56
2.3.2.3. Types of classification learners:	56
2.3.2.3.1. Lazy learners	56
2.3.2.3.2. Eager learners.....	56
2.3.2.4. Classification algorithms and techniques	56
2.3.2.4.1. Softmax classifier:	57
2.3.2.4.2. Artificial Neural Networks:	58
2.3.2.4.3. K-Nearest Neighbors (KNN)	59
2.3.2.4.4. Support Vector Machine (SVM)	59
2.3.2.4.5. Decision Tree	61
2.3.2.4.6. Random Forest.....	62
2.3.2.5 Classifier performance evaluation	62
2.3.2.5.1. Holdout method.....	62
2.3.2.5.2. Cross-validation.....	63
2.3.2.5.3. Confusion matrix (Precision and Recall)	63
2.3.2.5.4. Receiver Operating Characteristics (ROC curve).....	65
2.4. Conclusion.....	66

Chapter 3	67
State of the Arts.....	67
3.1. Introduction	67
3.2. Literature review.....	67
3.2.1. Conventional ML analysis.....	67
3.2.2. Deep learning and Neural networks models	70
3.2.2.1. Segmentation approach	70
3.2.2.2. Classification approach	73
3.3. Perspective and motivation	80
3.4. Conclusion.....	81
Chapter 4	82
Diabetic Retinal Tomographical Image Classification using CNN.....	82
4.1. Introduction	82
4.2. Image processing materials and methods.....	83
4.2.1. Dataset acquisition:	83
4.2.2. Dataset Pre-processing method.....	84
4.2.2.1. Dimension reduction:	84
4.2.2.2. Image denoising	84
4.2.2.3. Image cropping	84
4.3. The proposed method.....	86
4.3.1. CNN architecture	86
4.3.2. Setting-Up Hyper parameters:	89
4.4. Experimental performance evaluation	90
4.5. Discussion	94
4.6. Conclusion.....	95
Chapter 5	97
Convolutional Neural Network Classification Model for Diabetic Macular Edema using Spectral Domain Tomographical Scans	97
5.1. Introduction	97
5.2. Image processing methods	98
5.2.1. Dataset acquisition:	98
5.2.1.1. SERI Dataset.....	98
5.2.1.2. DUKE Dataset.....	98
5.2.1.3. Noor Eye Dataset.....	99

5.2.2. Dataset Pre-processing technique:.....	99
5.2.3. Synthetic data augmentation:	101
5.2.3.1. Generator:	104
5.2.3.2. Discriminator:	105
5.2.3.3. Adversarial Model:	106
5.2.3.4. GAN architecture:	106
5.2.3.5. Synthetic image evaluation	107
5.3. The proposed method.....	110
5.3.1. Fine-tune the pre-trained model.....	110
5.3.2. Re-train the model	110
5.4. Performance Analysis	113
5.5. Discussion	117
5.6. Conclusion.....	119
General Conclusion	120
Bibliography	126

List of Acronyms

1. **2D** (Two Dimension)
2. **3D** (Three Dimension)
3. **ACC** (Accuracy)
4. **Adagrad** (Adaptive Gradient)
5. **Adam** (Adaptive Moment Estimation)
6. **AE** (Auto-Encoder)
7. **AI** (Artificial Intelligence)
8. **AMD** (Age-related Macular Degeneration)
9. **ANN** (Artificial Neural Networks)
10. **AUC** (Area Under Curve)
11. **BM3D** (Block-Matching and 3D-Filtering)
12. **BoW** (Bag Of Words)
13. **CMV** (CytoMegalovirus)
14. **CNN** (Convolutional Neural Network)
15. **CNV** (Choroidal NeoVascularization)
16. **Convnet** (Convolutional Neural Network)
17. **CUHK** (Chinese University of Hong Kong)
18. **CV** (Computer Vision)
19. **CWS** (Cotton Wool Spots)
20. **DAGAN** (Data Augmentation Generative Adversarial Network)
21. **DCGAN** (Deep Convolutional Generative Adversarial Network)
22. **DenseNet** (Densely-Connected Network)
23. **DL** (Deep Learning)
24. **DME** (Diabetic Macular Edema)
25. **DR** (Diabetic Retinopathy)
26. **DT** (Decision Tree)
27. **ECG** (Electro-Cardiogram Impulse)
28. **EDST** (Doctoral School in Science & Technology)
29. **ELM** (External Limiting Membrane)
30. **ELM** (Extreme Learning Machine)

- 31.**EMG** (Electro-Myogram Waves)
- 32.**ERR** (Error Rate)
- 33.**EX** (Exudates)
- 34.**FC** (Fully Connected Layer)
- 35.**FCC** (Fully Connected Layer)
- 36.**FD-OCT** (Fourier domain Optical Coherence Tomography)
- 37.**FN** (False Negative)
- 38.**FP** (False Positive)
- 39.**FPR** (False Positive Rate)
- 40.**FT** (Fine-Tuning)
- 41.**GAN** (Generative Adversarial Network)
- 42.**GCL** (Ganglion Cell Layer)
- 43.**HC** (Hierarchical)
- 44.**HD** (High dimension)
- 45.**HEs** (Hemorrhages)
- 46.**HoG** (Histogram of Oriented Gradients)
- 47.**HR** (Hemorrhages)
- 48.**IFCNN** (Iterative Fusion CNN)
- 49.**IFTIM** (Functional and Molecular Imaging, Medical Image Processing)
- 50.**ILM** (Internal Limiting Membrane)
- 51.**ImVia** (Artificial Imaging and Vision)
- 52.**INL** (Inner Nuclear Layer)
- 53.**IPL** (Inner Plexiform Layer)
- 54.**IR** (Infra Red)
- 55.**IRF** (Intra-Retinal Fluid)
- 56.**IS/OS** (photoreceptor rod and cone Inner and Outer Segments)
- 57.**KNN** (K-Nearest Neighbors)
- 58.**LBP** (Local Binary Pattern)
- 59.**LIA** (Computer Science and Applications Laboratory)
- 60.**LOPO-CV** (Leave One Patient Out Cross-Validation)
- 61.**LTPO-CV** (Leave-Two-Patients-Out Cross-Validation)

- 62.**MA** (MicroAneurysms)
- 63.**MA**s (MicroAneurysms)
- 64.**ML** (Machine Learning)
- 65.**MMC** (Maximum Margin Classifier)
- 66.**MP** (Mega-Pixel)
- 67.**N** (Negative)
- 68.**NFL** (Nerve Fiber Layer)
- 69.**NHC** (Non-Hierarchical)
- 70.**NLM** (Non-Local Means)
- 71.**OCT** (Optical Coherence Tomography)
- 72.**ONL** (Outer Nuclear Layer)
- 73.**OPL** (Outer Plexiform Layer)
- 74.**P** (Positive)
- 75.**PCA** (Principal Component Analysis)
- 76.**PPV** (Positive Predictive)
- 77.**PREC** (Precision)
- 78.**RBF** (Radial Basis Function Kernel)
- 79.**RBF** (Radial Basis Function)
- 80.**RBM** (Restricted Boltzmann Machine)
- 81.**REC** (Recall)
- 82.**RELU** (Rectified Linear Activation Unit)
- 83.**ResNet** (Residual Neural Networks)
- 84.**RF** (Random Forest)
- 85.**RMSProp** (Root Mean Square Propagation)
- 86.**ROC** (Receiver Operating Characteristics)
- 87.**ROI** (Region On Interest)
- 88.**RPE** (Retinal Pigment Epithelium)
- 89.**RPE** (Retinal Pigmented Epithelium)
- 90.**SAE** (Stacked Auto-Encoder)
- 91.**SDA** (Stacked Denoising Auto-encoder)
- 92.**SD-OCT** (Spectral Domain Optical Coherence Tomography)

- 93.**Se** (Sensitivity)
- 94.**SERI** (Singapore Eye Research Institute)
- 95.**SEs** (Soft Exudates)
- 96.**SGD** (Stochastic Gradient Descent)
- 97.**SM** (Soft Margin)
- 98.**Sp** (Specificity)
- 99.**SPIM** (Engineering And Micro-technical Sciences)
- 100.**SVC** (Support Vector Classifier)
- 101.**SVM** (Support Vector Machine)
- 102.**Tanh** (Hyperbolic Tangent Function)
- 103.**TD_OCT** (Time Domain Optical Coherence Tomography)
- 104.**TL** (Transfer Learning)
- 105.**TN** (True Negative)
- 106.**TNR** (True Negative Rate)
- 107.**TP** (True Positive)
- 108.**TPR** (True Positive Rate)
- 109.**US** (Ultrasound)
- 110.**UV** (Ultra Violet)
- 111.**VGG** (Visual Geometry Group)

List of Figures

Fig. 1: Eye's anatomy.....	25
Fig. 2: Anatomy of the retina	26
Fig. 3: A representation of the 10 layers in the retina's OCT image	27
Fig. 4 : Partial blindness in the central vision caused by AMD	28
Fig. 5: Leakage representation caused by AMD	29
Fig. 6: An eye infected by cataracts	29
Fig. 7: Fundus image with CMV retinitis	30
Fig. 8: Retinal Detachment representation.....	30
Fig. 9: An Eye representation with Glaucoma	31
Fig. 10: Developpement of the Glucoma	31
Fig. 11: Retinal degeneration with yellow spots is a sign of heart problem	32
Fig. 12: A normal eye vs. DR eye showing abnormal leaking and vessels growth	32
Fig. 13: A display of healthy vision vs. same one infected by DR with dark spots.	33
Fig. 14: Fundus retinal image representation in the center. Pointing to normal formations (fovea, vessels and optic disc) and deformations related to DR: The left section (MAs), and HEs and in right(SEs), and (EXs)	33
Fig. 15: Representative OCT images of the different types of DME	34
Fig. 16: DR vs. DME eye where swelling vessels in DR leads into leaking fluid in macula.....	35
Fig. 17: Comparison between several sight being seen by eye infected with DR (Group A) and DME (Group B)	36
Fig. 18: Color Fundus photograph of a healthy person	37
Fig. 19: An illustration of 2 AO systems with its corresponding image	38
Fig. 20: Images of the normal retinal nerve fiber layer using AO-SLO.....	38
Fig. 21: PAM imaging presentation of a Rabbit	39
Fig. 22: An Ultrasonography presentation of retinal detachment	40
Fig. 23: A presentation of anatomical MRI of a Rat and Cat retina	41
Fig. 24: Operating method of TD-OCT: light emitting from the light source is divided into the reference wave and the middle wave. The echo light received is joined again and recorded by the detector	43
Fig. 25: Demonstration of 3D OCT-2000 SD-OCT machine and SPECTRALIS OCT device	43

Fig. 26: Optical implementation of spectrometer based OCT (SD-OCT) which contains a spectrometer for wave division	44
Fig. 27: The left image is a fundus optic nerve tissue captured by SPECTRALIS machine where the green line is represented on the right image as OCT cross-sectional B-scan . The Edema area is presented with blue arrow.....	44
Fig. 28: OCTA Fields of View.....	45
Fig. 29: An example of RBM network architecture.....	47
Fig. 30: An example of AE network architecture	47
Fig. 31: An example of CNN model architecture.	47
Fig. 32: Example of generated HD face images (Bottom) produced by handy drafts (top)	49
Fig. 33: : Softmax Regression	57
Fig. 34: A flowchart overview of ANN architecture	59
Fig. 35: Decision Tree classification algorithm architecture	61
Fig. 36: Random Forest algorithm outline	62
Fig. 37: A representation of 10-fold Cross-Validation	63
Fig. 38: A representation of ROC Curve plot.....	66
Fig. 39: Flowchart of the Gaussian mixtures model creation	68
Fig. 40: Overview of the algorithm for classifying SD-OCT volumes	68
Fig. 41: Outline of the algorithm for preprocessing and classifying	69
Fig. 42: Flowchart for OCT image.....	71
Fig. 43: Flowchart of multiclass fluid segmentation	72
Fig. 44: Details algorithm for preprocessing and classifying	74
Fig. 45: The preprocessing steps.....	74
Fig. 46: Experts and CNN gate network are fed by precise scales of the input model.....	74
Fig. 47: VGG16 Layer block model	76
Fig. 48: Overview of combined framework of feature extraction using AlexNet, VggNet and GoogleNet	77
Fig. 49: Architecture of the CNN model.....	78
Fig. 50: An example of OCT images of diabetic pathologies compared to normal case	83
Fig. 51: Example of NLM image denoising for a normal patient	84
Fig. 52: Cropping example where red rectangles indicates the area to eliminate 256 rows each (Group A) leading to a 256×256 pixels images (Group B).	85
Fig. 53: From original to cropped image, Example of OCT image.	85

Fig. 54: Flow chart diagram.....	86
Fig. 55: Steps to create a deep learning-based CNN model for DR diagnosis using SD-OCT	87
Fig. 56: Architecture of CNN model depth for CUHK dataset: the convnet contains 16 layers.	87
Fig. 57: Architecture of the CNN model. The input layer accepts B-scans with 256x256 pixels from CUHK dataset. The parameters (feature maps and size) are reported beneath every layer.	89
Fig. 58: Accuracy and Loss rate	91
Fig. 59: curve plot showing the performance of proposed model during testing.....	93
Fig. 60: Example of an OCT image pre-processing phase, from original to flattened, cropped and mirrored image for a healthy person (Group A) and a diabetic patient with DME (Group B).	100
Fig. 61: (a) Cropping simulation where red rectangles indicate the area to eliminate with a size of 256x256 pixels resulting a flattened and cropped image with a size of 256x256 (b).	101
Fig. 62: A representation of GAN model.....	102
Fig. 63: Discriminator training presentation where the discriminator learns to identify fake images.	103
Fig. 64: Generator training presentation where the generator learns to fool the discriminator.	104
Fig. 65: Representation of the proposed generator model.....	104
Fig. 66: Representation of the	106
Fig. 67: Generator and Discriminator Loss.....	108
Fig. 68: The proposed approach's workflow diagram for classifying.....	111
Fig. 69: Flowchart of the model trained over CUHK dataset	112
Fig. 70: Flowchart of the pre-trained model over SERI dataset in addition to newly classification (red section).....	113
Fig. 71: The ROC curve showed the classification performance.....	115
Fig. 72: Example B-scans from Normal, and DME	115
Fig. 73: Example B-scans from Normal, and DME.....	116

List of Tables

Table 1: Confusion Matrix for binary classification	63
Table 2: State of the art summary for retinal disorders detection methods using conventional ML analysis.....	70
Table 3: State of the art summary for segmentation methods using neural network.....	72
Table 4: State of the art summary for classification methods using neural network	79
Table 5: Layers values for the proposed architecture	88
Table 6: The distribution of images used in the system	90
Table 7: Evaluation results of the testing dataset	91
Table 8: Performance result of proposed model during training.....	92
Table 9 : Performance result of proposed model during validation	92
Table 10: Confusin matrix showing the testing rate.....	93
Table 11: Performance ratios of testing set.....	94
Table 12: Examples of generated DME images	107
Table 13: . Examples of generated images for normal cases.....	108
Table 14: Representaton of SERI preparation	109
Table 15: Images distribution for the training and validation sets.....	114
Table 16: The confusion matrix of our model’s classification on the validation	114
Table 17: Classification report showing the validation ratios.....	115
Table 18: A demonstration for all DME volumes where the number of slices showing DME prediction are colored in "green" compared to false negative colored in "red"	118

List of Equations

(1).....	57
(2).....	57
(3).....	58
(4).....	58
(5).....	59
(6).....	64
(7).....	64
(8).....	64
(9).....	65
(10).....	65
(11).....	65
(12).....	65

Introduction

1- General context

The lifestyle of recent civilization has improved significantly with the evolution of AI in recent years, which is a technology with multiple advanced components like ML and DL algorithms. These algorithms are awaited to supply the ophthalmologists with automated machines for early detection and diagnosis for the treatment of ocular diseases in the next years. DL has been applied in the ophthalmic domain to confirm the diagnosis of pathologies and scan images where the ophthalmic imaging presents a method to diagnose and detect the progress of plenty of diseases including DR, DME, and other ophthalmic.

There are two types of imaging applied as diagnostic techniques in the ophthalmic application: fundus photography and OCT is recently the assessment standard utilized to measure the fluid leaks in the retina in ophthalmology worldwide. Variations of lifestyle in communities, demographics differences, the extended average age, and the growing pattern of diabetic pathologies constitute an increasing need for such images. Moreover, the unavailability of many retina-trained human specialists is a principal obstacle in numerous advanced countries. Therefore, regarding the speed growth of the population, it is sure that examining a large number of images takes time and effort besides labor wages and human mistakes. Consequently, the early detection leading to suitable treatment of diabetic disorders via computerized systems will be assured in the coming future.

Any person with an uncontrolled high sugar level in the blood can lead to diabetes stage 1 or 2. Over time, this patient is likely infected by diabetic retinopathy in both eyes with different infections level. The blood vessels that fulfill blood to the eye's retina will be blocked which may lead to swallow and prevent the retina to get the blood it needs to work normally. Eventually, these blood vessels might begin to bleed a thick fluid into the vitreous, which affects the quality of sharp vision (dark spots) and can lead to permanent or partial vision loss. It is crucial in this stage to get immediate treatment. Otherwise, the patient is in danger of losing his sight.

DR and DME are the most generally diabetic eye pathologies that threaten the vision and they are the primary causes of blindness in the most advanced nations. Hence, early detection accompanied by urgent therapy systems is meaningful in limiting vision problems and can minimize the prospect of vision loss.

From old traditional methods to new deep learning methods, numerous papers have addressed the problem of the detection of pathologies in the retina using OCT scans. This retinal medical field has been covered using various feature detectors and

classifiers for the classification and segmentation of retinal disease layers. Certain models could be confused and necessitate intervention from ophthalmologists. Deep learning models for retinal disease detection over OCT images have received much attention in many research fields such as medical analysis and computer-aided diagnosis. This continuous growth can be related to the computational performance availability and accessibility of processing materials, which were not generally affordable 10 years ago. Besides, it has shown forefront achievement widely in image processing and computer vision, and especially in image detection object recognition, alongside other research fields.

Moreover, the effectiveness of deep learning methods has a powerful dependence on the structure of the proposed model, which raises the processing burdens. The progress and appropriateness of deep learning methods rely simultaneously on the design of the model and its adaptation. In this thesis, we propose a novel deep learning perspective applied to a medical domain with less processing overheads.

2- Thesis objective

In this thesis, the foremost objective is to detect the pathologies from scanned OCT images. Therefore, the major plan is to propose a classification model with the ability to detect patterns for diabetic macular edema and diabetic retinopathy with no intervention of clinical persons analyzing these retinal scans. The model must provide the capability to automatically detect pathology with high precision for predicting image classes as normal or infected. For that reason, this thesis focused on the 2 principal retinal pathologies (DR and DME) using an augmented dataset in order to train a CNN model which is conceivably considered the most efficient model of deep learning algorithms in computer vision.

3- Thesis synthetic plan

The first covered subject is an overview and description of the eye structure and the commonly diabetic ocular-related diseases that can harm human vision. This topic covers detailed information about the pathology besides its causes and symptoms where the detection requires subjective reviews of professionals to examine patient's volumetric OCT scans. Subsequently, deep learning methods are presented where the advantages of DL are shown regarding this topic, alongside with other domains.

The second chapter will introduce an overview of data pre-processing and classification. The first part will present the method to use for data manipulation as a pre-step in order to prepare it to be fed into ML algorithms, leading to the best result.

The second will present definitions of classification types, algorithms, approaches, terms, learners, and categories to be used partially in this thesis.

Based on relevant studies, the third chapter of this thesis will discuss the different methods and models used in related work. The first part covers the bottleneck of traditional ML including the different modeling techniques, the feature extractors used, and the classifiers applied for solving the detection problem. In the second part, we present the concept of neural networks and the performance of CNN when applied to medical images. In the third part, we discuss the effectiveness of data augmentation techniques.

The fourth chapter of this thesis will focus on solving the classification of diabetic retinopathy pathologies and detect particularly DR and DME. In the first part, we present a CNN model being trained from scratch over a pre-processed dataset. We show that the proposed model achieved advanced accuracy in detecting pathologies. The second part presents a well-architected CNN model by using hyper-parameters values that provide better results. The third part discussed the classification results.

The fifth chapter will also focus on solving the classification of DME. Based on the CNN model presented in the fourth chapter, this chapter presents a fine-tuning of the pre-trained model using a similar dataset from other institutions after being enlarged using both classical and GAN augmentation methods. This final model achieves advanced accuracy in detection of DME cases.

Finally, we conclude by a general conclusion presenting different perspectives covered by this thesis.

Chapter 1

Overview of Diabetic Retinopathy pathologies

OCT Scanning methods and deep learning applications

Summary

1.1. Introduction	24
1.2. Overview of eyeball and retinal-related pathologies	24
1.2.1. The Eyeball.....	24
1.2.2. Eye structure	25
1.2.3. Retinal layers.....	26
1.2.4. Retinal pathologies	27
1.2.4.1. Age-related macular degeneration (AMD).....	28
1.2.4.2. Cataracts	29
1.2.4.3. CytoMegalovirus Retinitis (CMV)	30
1.2.4.4. Retinal Detachment.....	30
1.2.4.5. Glaucoma.....	31
1.2.4.6. Cardiovascular Disease	31
1.2.4.7. Diabetic Retinopathy (DR).....	32
1.2.4.8. Diabetic macular edema (DME).....	34
1.3. Optical Imaging Modalities	36
1.3.1. Fundus Photography	37
1.3.2. Adaptive Optics (AO)	37
1.3.3. Scanning laser ophthalmoscopy (SLO)	38
1.3.4. Photoacoustic microscopy (PAM)	39
1.3.5. Ultrasound Biomicroscopy (UBM).....	40
1.3.6. Magnetic resonance imaging (MRI)	40
1.3.7. Optical Coherence Tomography (OCT).....	41
1.3.8. Optical coherence tomography angiography (OCTA)	45
1.4. Deep Neural Networks:.....	46
1.4.1. Deep learning applications.....	48
1.5. Conclusion.....	51

1.1. Introduction

The eye, which is one of the most important and critical part of a human's body, can be sensitive to different problems that affect visual quality. The leading cause of vision loss is primarily related to age and diabetic pathologies that affect the retina. Therefore, due to the value of vision as the main cause of independent healthy life, a worldwide warning has been declared about retinal diseases.

One of these pathologies is diabetes that affects the retina if it is left untreated. Diabetes is a considerable disease that happened when a human's blood sugar is excessively high. Eventually, extra sugar in the blood can lead to crucial health problems [1], such as heart disease, stroke, blindness, nerve damage, and kidney failure. Moreover, people with diabetes appear to be vulnerable to becoming hardly sick with the COVID-19 virus [2]; When diabetic people reveal infected, it can be tougher to heal due to vacillations in blood glucose levels and the appearance of diabetes complexities; Firstly, because the immune system performance made it stubborn to fight the virus and may lead to a more extended recovery period. Secondly, the virus may flourish in an environment of elevated blood glucose. Globally, 1 in 10 persons have diabetes where about 463 million adults over the age of 20 in 2019 were living with diabetes and more than 4 million are deceased; by 2045 this will increase to 700 million [2] while new diabetes cases have increased over the last decade in people younger than 20 years [3].

This thesis focuses on diabetic retinal-related diseases, specifically DR and DME using Spectral Domain OCT (SD-OCT) high-resolution scans that capture the retinal depth. OCT is considered a new screening modality used in the automatic classification of retinal pathologies.

In this chapter, we described the eye's structure, diseases related to age and diabetic causing the retinal problems, types of retinal pathologies, and the methods applied to observe the retina. Besides, we reviewed the powerful deep learning applications in other research fields where its success is considered the main reason to choose it as our work methodology for this thesis.

1.2. Overview of eyeball and retinal-related pathologies

1.2.1. The Eyeball

The eye is an organ located within a protective bony cavity in the human head, where the diameter is up to 2.5 cm. The eye is linked to its socket by six muscles connected to the sclera [4]. These extra-ocular muscles are responsible for moving the eye in different directions. The eye picks up the light reflected by objects, which allows

a human to see where a sensory receptor interacts with the light coming or reflected from the visual objects, and the retina in the eye converts this light into an electrical signal that is transmitted through the nerves and then towards the visual area in the brain. The anatomy of the eye is shown in Fig. 1 with different layers.

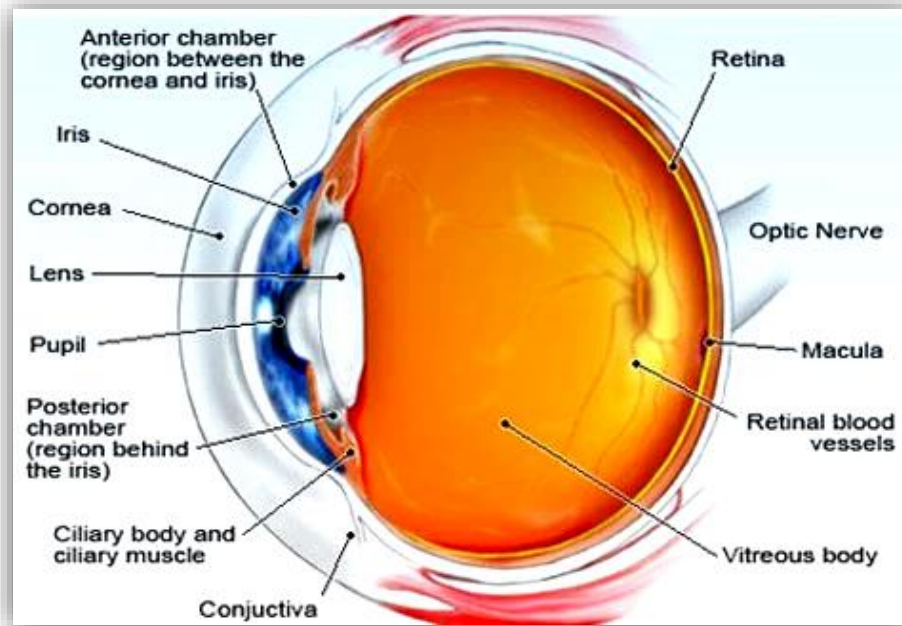


Fig. 1: Eye's anatomy [4]

1.2.2. Eye structure

The eye consists of a number of different parts working together step-forwardly to make a clear vision, where the vision task for every part is listed as follows:

- *Cornea*: a lens that captures incoming reflected light. It focuses the entry of light into the eye.
- *Anterior Chamber*: space filled with a transparent watery fluid inside the eye between the iris and the cornea's surface. It maintains the intraocular pressure, expands the globe of the eye, and provides amino acids and glucose for the ocular tissues.
- *Pupil*: a hole placed in the center of the iris of the eye that permits controlled beams of light to hit the retina.
- *Iris*: a colored division of the eye that aids to determine the entering amount of light by moving the eye muscles depending on the pupil instructions.
- *Lens*: transparent formation in the eye that directs light beams onto the retina.

- *Retina*: nerve layer that fills the end of the eye, senses, and transforms the lights into electrical impulses that transfer through the optic nerve to the brain.
- *Macula*: a small section in the retina that includes sensitive receptors and enables a clearly detailed view.
- *Optic nerve*: links the eye toward the brain and transfers the electrical impulses formed by the retina to the visual cortex of the brain.
- *Vitreous*: a clear, jelly-like substance that fills the middle of the eye.

1.2.3. Retinal layers

A retina is a thin multi-layered membrane holding light-sensitive receptors that line the internal aspect of the back wall of the eyeball. It is composed of epithelial, glial, and neural cells, which are organized into 10 distinctive layers in which a specialized group of receptors, photoreceptors, can be found [5]. These photoreceptors are localized around an area near the macula, which is the practical center of the retina. The fovea is placed in the center of the macula. The macula is responsible for high-resolution and color vision which is provided by different types of photoreceptors.

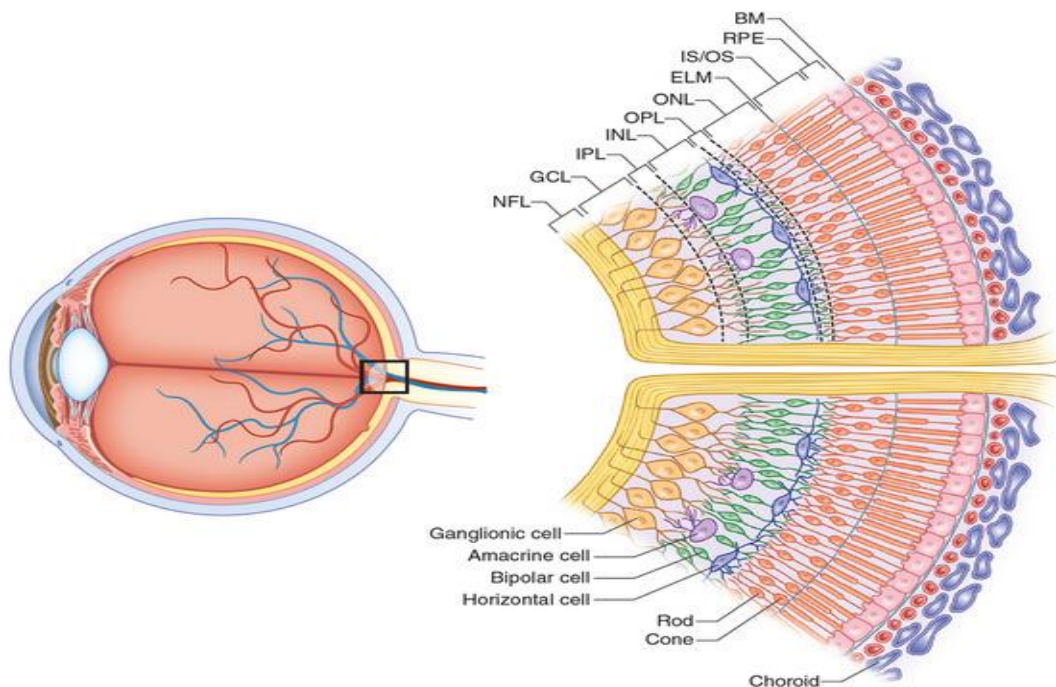


Fig. 2: Anatomy of the retina [6]

Fig. 2 presents the parts of the retina. Photoreceptors are a specialized type of neuro-epithelial cells that collect light and transform it into an electrical signal. Photoreceptors are arranged tightly together, permitting a large mass of light to be captured across a small area on the retina. Photoreceptors are organized into two groups: Rod cells are highly sensitive to light and operate in night vision, whereas

Cone cells are responsible for color vision and capable of identifying a wide spectrum of light photons.

The retina measures 0.56 mm and consists of 10 layers presented from the vitreous to the choroid as follows: (1) the internal limiting membrane (ILM); (2) the nerve fiber layer (NFL); (3) the ganglion cell layer (GCL); (4) the inner plexiform layer (IPL); (5) the inner nuclear layer (INL); (6) the outer plexiform layer (OPL); (7) the outer nuclear layer (ONL); (8) the external limiting membrane (ELM); (9) the photoreceptor layer (rod and cone inner and outer segments (IS/OS)); (10) the retinal pigmented epithelium (RPE). Fig. 3 presents an organization of retina's layers in an OCT scan.

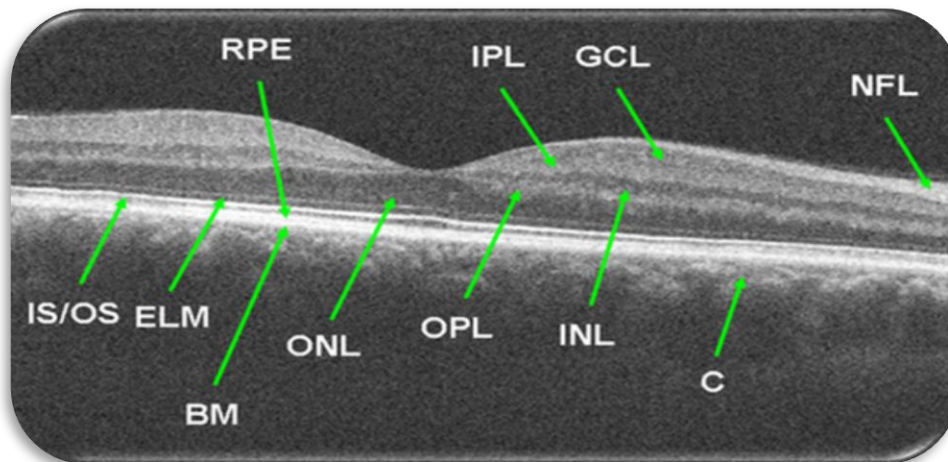


Fig. 3: A representation of the 10 layers in the retina's OCT image [6] where BM= Bruch's membrane and c= choroid

1.2.4. Retinal pathologies

This part describes the deformations that may occur in the retina caused by different deterioration such as aging and diabetic-related issues. Eye deformities do not have any preliminary warning system since any pain or vision obstacles may be felt or perceived by the person unless the problem improves. Therefore, an urgent detection approach is addressed to early discover the foremost silent retinal-related eye diseases. Thus, the main scope in this thesis is to examine the diabetic-related diseases especially DR and DME.

Internationally, the number of people with DR will develop from about 126 million in 2011 to more than 190 million by 2030 which indicates that more than a third of diabetic citizens are infected with DR and the number keeps on rising [3]. Meanwhile, the incidence over the last decade is 20% in younger diabetic patients versus 40% in older diabetic patients [21].

Diabetic eye disease is one of the eye problems that can harm people with diabetes. These ailments involve diabetic retinopathy DR, diabetic macular edema DME, cataracts, and glaucoma. Over time, diabetes can cause harm to people's eyes that may lead to vision loss progressively [1]. Several actions can be taken to prevent, early detect diabetic eye diseases or manage and control diabetes in case of infection. Usually, there are neither symptoms signs of diabetic eye disease nor blindness when the illness first occurs. An entire widened eye exam assists the doctor to detect and treat eye problems at the beginning before serious vision problems can happened.

Diabetes strikes the eyes when glucose in the blood is excessively increased, which doesn't necessarily lead to blindness in the short term, except for blurriness vision sometimes which affects people for several days or weeks, since care plan programs or medicines have been changed. This temporary blurred vision is caused by upraised fluid levels or swelling in the tissues, and vision will be normal and stable as soon as the sugar level reaches closer to average.

Over time, high glucose in the blood can destroy the blood vessels in the eyes. When the blood glucose begins firstly to rise, especially during the pre-diabetes period, the damage can begin without even diagnosed as infected. Damaged vessels commence at the beginning to leak fluid that can cause swelling which makes blood vessels start to expand and provoke a bleed in the eye, lead to scarring, or begin seriously high pressure inside the eye.

The frequently retinal diseases that can endanger the vision are:

1.2.4.1. Age-related macular degeneration (AMD)

Age-related macular degeneration (AMD) is an acute eye disease that affects the macula and can blur the central clear vision of older people above 50. AMD is a condition commonly spread in advanced western countries; it progresses slowly in some people, and for other people it proceeds faster which can lead to central blindness in one eye or both eyes as shown in Fig. 4.



Fig. 4 : Partial blindness in the central vision caused by AMD [7]

AMD is distinguished by degeneration or collapse of the macula. In AMD, blood vessels begin to swell up from the choroid located below the retina into the retina. This loss in central clear vision induced by AMD makes the person suffer to see face details, drive easily, or perform any home improvements. For AMD, there are two categories: dry AMD and wet AMD. In wet AMD, the growing blood vessels behind the retina start to arise, leaking fluid, leading to fast central vision loss. In dry AMD, the macula sensitive cells for lights start slowly breaking down leading to central vision decrease over time. Fig. 5 shows an example of fluid leakage in the retina.

Risk factors for developing macular degeneration include aging, gender (which strikes women more than men), obesity and inactivity, heredity, cardiovascular disease, and smoking.



Fig. 5: Leakage representation caused by AMD [4]

1.2.4.2. Cataracts

A cataract is a painless disease in the eye's lens providing a cloudy-like shape in the crystalline, which creates blurry vision as seen in Fig. 6. It grows slowly with age which causing some changes to the cornea. The cataracts are caused commonly by diabetes, trauma, some medications like steroids, and extreme UV light exposure. Any routine eye exam can observe the cataract while the treatments involve glasses, magnifying lenses, or risky surgery to remove and replace the lens by an artificial one [4]. Patients with cataract afford badly light sensitivity with dimming color; they suffer to see during the night with blurred vision in one or both eyes [8].

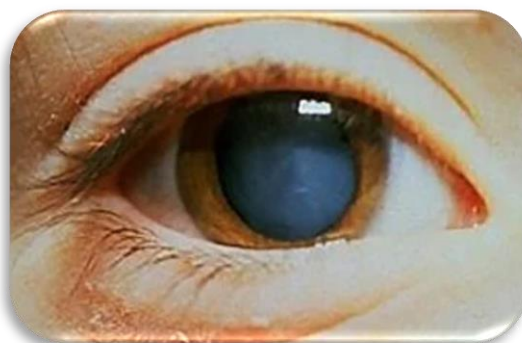


Fig. 6: An eye infected by cataracts [4]

1.2.4.3. CytoMegalovirus Retinitis (CMV)

Cytomegalovirus retinitis (CMV retinitis) is a dangerous viral eye disease of the retina founded most often in persons with weakened immune systems [9]. CMV retinitis signs can start with a blurred vision caused by a slow beginning of floaters over a couple of days, or an obstructed spot in the central vision, leading to side vision loss in one eye but often develop to the other eye. CMV retinitis can damage the retina and destroy the optic nerve if not treated properly or enhancing the immune system. People with CMV retinitis will often generate a detached retina. Fig. 7 presents a sample of CMV in a fundus image.

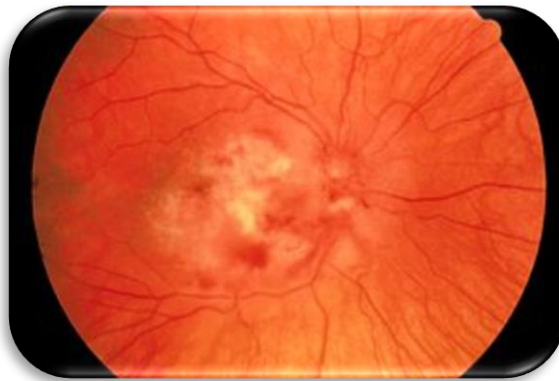


Fig. 7: Fundus image with CMV retinitis [10]

1.2.4.4. Retinal Detachment

Retinal detachment is observed by the existence of fluid under the retina that occurs when the retina detaches from its underlying tissue layers [4]. This painless lift-away issue usually happens when fluid passes through a retinal tear. The accumulation of fluid behind the retina is the reason that separates the retina from the back wall of the eye as seen in Fig. 8. The notable symptoms include the perception of flashing lights, floaters, or a curtain drawn over your visual field.

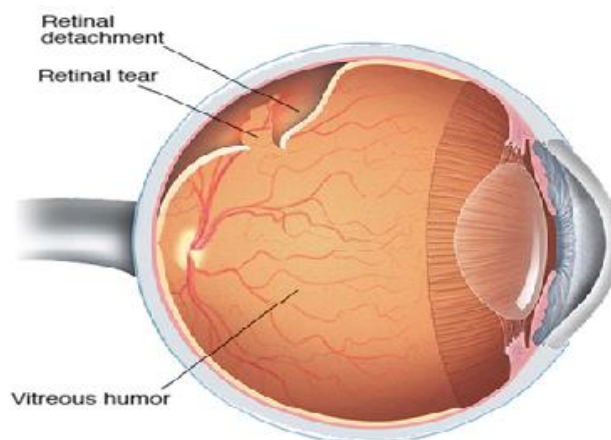


Fig. 8: Retinal Detachment representation [11]

1.2.4.5. Glaucoma

Glaucoma is a combination of eye ailments that damages the optic nerve; it originates due to raised intraocular fluid pressure inside the eye [4] causing headache, dizziness and blurry vision [12]. As seen in Fig. 10, the elevated pressure attacks the optic nerve and may lead to blindness.

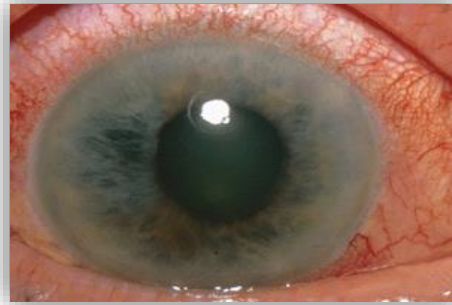


Fig. 9: An Eye representation with Glaucoma [13]

Glaucoma is commonly painless and occurs without any symptoms in the early phases, thus it is hard to detect. Glaucoma is classified either as open-angle (a most common painless form that progresses slowly over a long time) or angle-closure glaucoma (a painful form with sudden redness of the eye as seen in Fig. 9). By the time the damage is permanent and cannot be cured which can cause irreversible blindness [13].

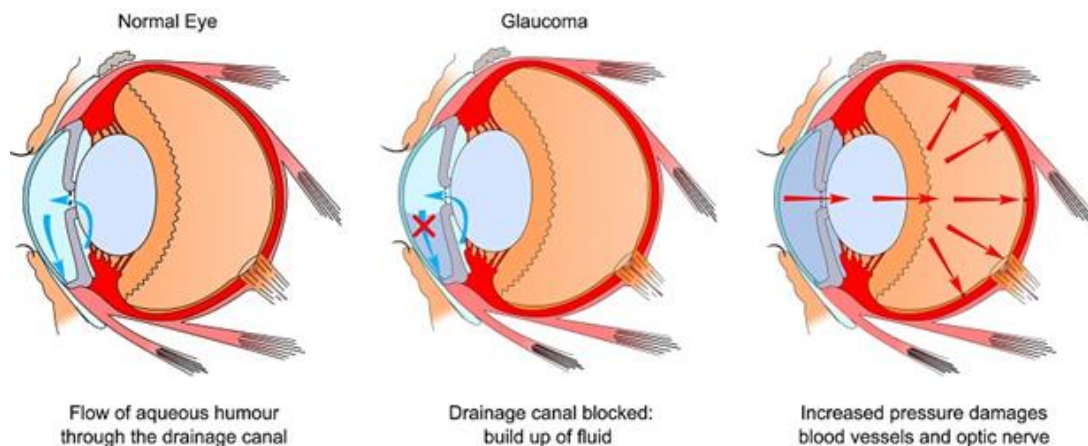


Fig. 10: Developpement of the Glucoma [14]

1.2.4.6. Cardiovascular Disease

The cardiovascular condition reveals itself in the retina in a number of ways. Hypertension and atherosclerosis cause variations in the ratio among retinal vessels' diameter and veins. Any ratio diminishing, such as arteries' thinning and veins' widening, is connected with elevated stroke risk and myocardial infarction [15].

Direct retinal ischemia can be caused by Hypertension that makes the infarcts of the retina apparent as cotton wool spots (CWS) and the infarcts of the choroid apparent as retinal yellowish spots as seen in Fig. 11.

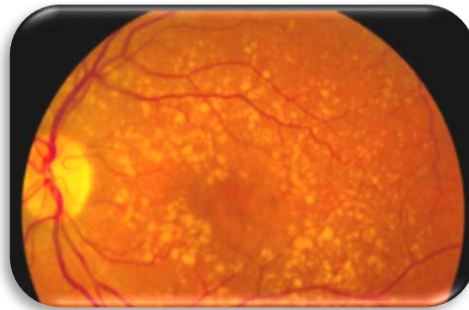


Fig. 11: Retinal degeneration with yellow spots is a sign of heart problem [16]

1.2.4.7. Diabetic Retinopathy (DR)

DR is one of the leading causes of blindness among adults around the world, and the growth of DR is attached to diabetes mellitus disease duration. DR is a disease that can affect and destroy blood vessels and harm the retina which is the inside part at the end of each eye. A person can see when the retina captures light and transforms it into signals to be decoded by a brain.

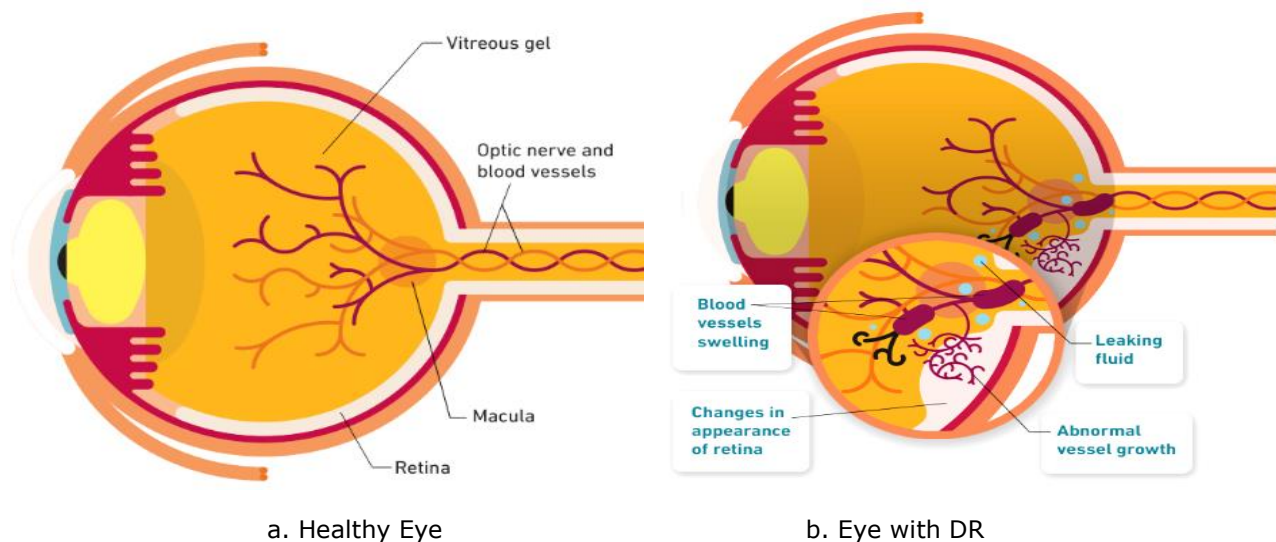


Fig. 12: A normal eye vs. DR eye showing abnormal leaking and vessels growth [17]

There are two stages of DR [1]. The first is called non-proliferative DR and the second is called proliferative DR. The non-proliferative stage occurs near the beginning of diabetic retinopathy where blood vessels can be thinner and make a bulge that leads to leaking into the retina. The stage of proliferative happens when DR becomes worse, some blood vessels are blocked causing new blood vessels to expand or increase in the retina leading to dangerous vision difficulties. Fig. 12

shows a demonstration how normal eye may look in comparison to proliferative stage where the eye with DR contains bulges, blood vessels swelling leading to abnormal growth, leaking fluid and some changes in retina's appearance [17].

DR is considered as epidemic disease affecting one-third of diabetic patients. It is considering as the most frequent reason for blindness in diabetic persons. Therefore, besides personal caring programs, early exploring and treating can reduce the risk of vision loss by 95% [3]. Fig. 13 (a) depicts an overview of healthy eye seeing 2 boys clearly meanwhile Fig. 13 (b) shows diabetic patient with DR seeing the same boys hardly with hazy and dark spots blocking the sharp vision.

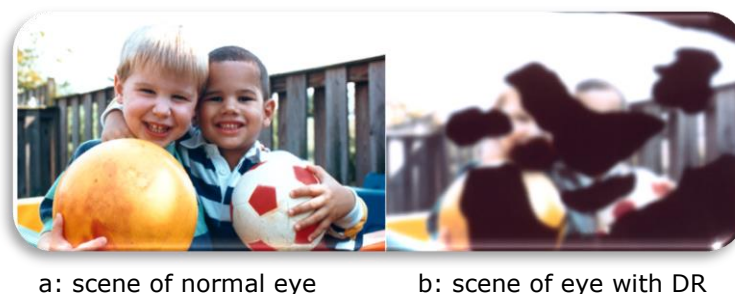


Fig. 13: A display of healthy vision vs. same one infected by DR with dark spots [18].

DR directs to progressive changes in vasculature formation (including vascular tortuosity, branching angles, and calibers) and producing malformations (microaneurysms, hemorrhages, and exudates). DR is diagnosed by visually inspecting retinal fundus images for the presence of one or more retinal lesions like microaneurysms (MAs), hemorrhages (HEs), soft exudates (SEs), and hard exudates (EXs) [19] as described in Fig. 14.

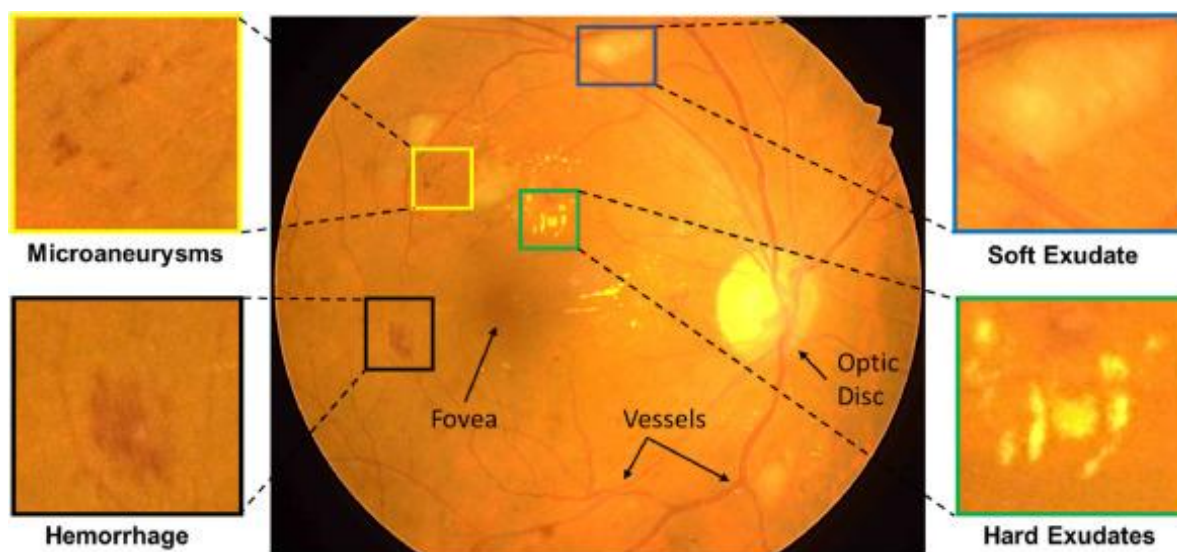


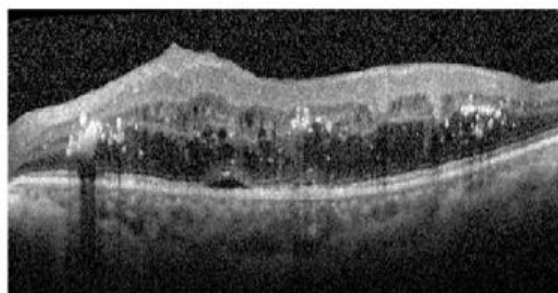
Fig. 14: Fundus retinal image representation in the center. Pointing to normal formations (fovea, vessels and optic disc) and deformations related to DR: The left section (MAs), and HEs and in right(SEs), and (EXs) [20].

1.2.4.8. Diabetic macular edema (DME)

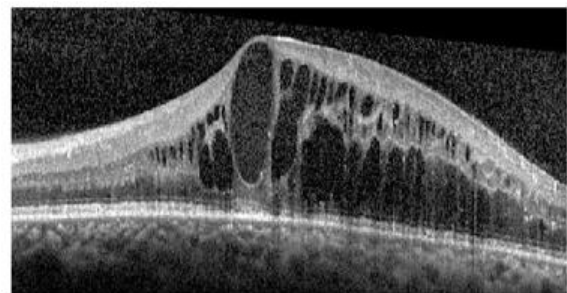
The macula is the vital partition of a person's retina allowing him to view and controls the ability of detailed sight. Any leak into the macula produces macular edema. This leak is caused by the accumulation of fluid that is called diabetic macular edema (DME). It forms a sort of swelling, and in some cases, cyst formation in the macula, caused by chronic hyperglycemia [21]. Progressively, it may damage the clear vision, leading to partial or total vision loss.

DME is commonly a consequence to people who previously have additional symptoms of DR and is secondary to retinal barrier rupture where up to one-third of diabetic patients are diagnosed with DME. The most significant molecule in the retinal barrier rupture is the vascular endothelial growth factor (VEGF). The initiation of anti-VEGF and steroid medications for treating DME has increased the understanding of pathophysiology [22]. However, the utilization of anti-VEGF drugs has shown that about one-third of patients built an immune system to intra-vitreous therapy [23].

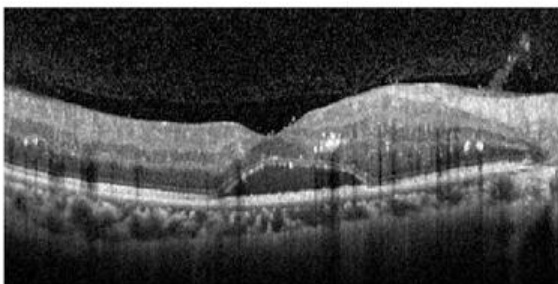
The diagnosis of macular edema is performed clinically. Traditionally, stereoscopic fundus photography was used as a standard screening system for diagnosing DME [24] but has since been replaced by OCT images. DME can be classified into focal and diffuse. Focal macular edema is characterized by the appearance of retinal thickening in local areas, derived from leakage of individual/clusters of microaneurysms.



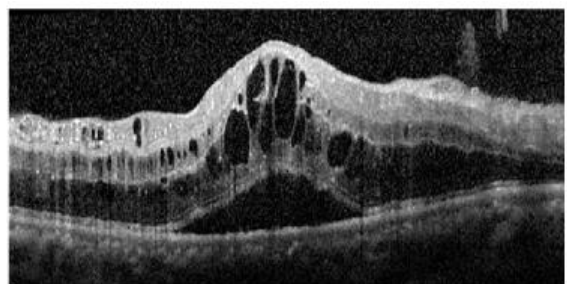
(a): diffuse retinal thickening



(b): cystoid macular edema



(d): serous retinal detachment



(e): mixed type

Fig. 15: Representative OCT images of the different types of DME [25]

Diffuse macular edema is derived from damaged capillaries, microaneurysms, and arterioles. It is distinguished by a more spread thickening of the macula. Cystoid macular edema is usually linked to diffuse macular edema, it occurs from a breakdown of the retinal barrier with a swollen fluid in the outer plexiform and inner nuclear layers. The appearance or absence of cystoid does not affect the diagnosis of DME. Fig. 15 presents different types of DME in OCT scans. If the macular edema has not improved yet, the first sign of this disease will be the hazy vision in the middle of the visual scope amongst the surrounding area of one eye. Another sign is the loss of color brightness with double vision.

Unfortunately, people affected by DME will have symptoms that vary from lightly blurry vision to complete blindness eventually. A person having DR might develop into DME when there is an unnatural accumulation of fluid in the macula left untreated caused by broken blood vessels in the retina. These blood vessels start to increase pressure in the eye and leak fluid. Fig. 16 presents an eye with abnormal vessels growth in a DR person in image (a) which is the principal cause to swollen macula in image (b). This leakage in retina leads to macular edema DME [17].

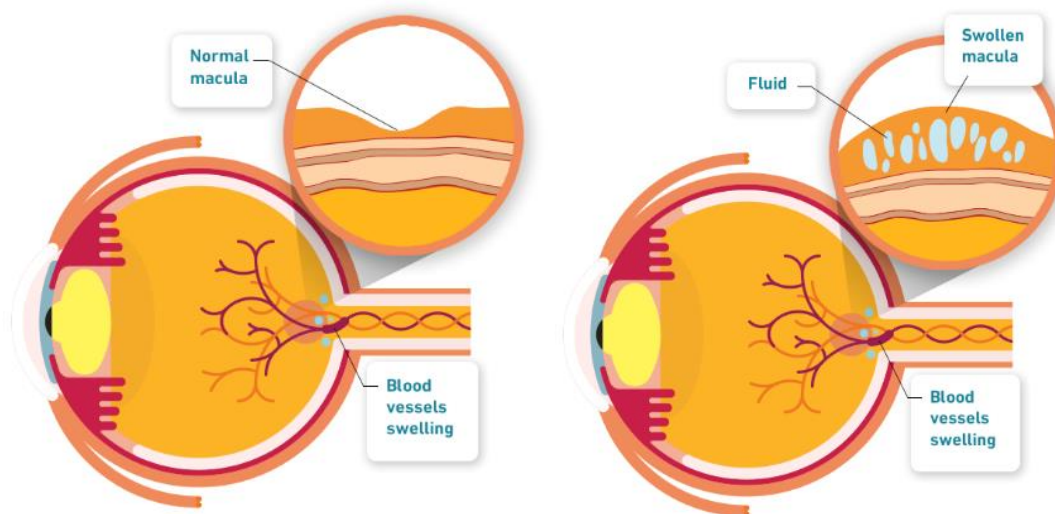


Fig. 16: DR vs. DME eye where swelling vessels in DR leads into leaking fluid in macula [17]

DME can also happen after an eye's operation, which can be related to age macular deterioration, or as a result of inflammatory disorders that impact the eye [26]. Macular edema may occur within a few weeks after any eye surgery related to retinal disease. The infection probably hits the second eye with an incidence of 50% after the first eye is infected. Contrary to macular edema caused by diabetes, this inflammation is temporary and mild, which will be healed using eye-drop treatment.

Diabetes as well as early unsupervised high blood glucose may lead to DR and DME. It is necessary to control the blood sugar otherwise, problems with vision may

occur. Fig. 17 presents a diversity of sights where each image indicates the progress of disease. DR in group 'A' may differ from thin to dark spots as seen in image 'b' and 'c'. Meanwhile group 'B' shows multiple develops of vision failure such as blurry vision like image 'd', loss of color brightness in image 'e' or blunt one in image 'f'.

Early detection of DR and DME can save lives and prevent unpleasant consequences. Thus diabetic patients should control their blood sugar initially, then undergo clinical examination to check the retina frequently using a powerful optical scan tool via Spectral-Domain Optical Coherence Tomography SD-OCT [27]. OCT is popularly used in ophthalmology for investigating the morphology of the retina for disease discovery. This thesis applies this important optical screening system to help patients with diabetic by detecting diseases in their eyes before any partial or total blindness may occur.

Group A: Eye sight with DR



a: vision of a healthy normal eye

b: small dark area

c : large dark area

Group B: Eye sight with DME



d: Blurry vision

e: fade vision

f: corrugated vision

Fig. 17: Comparison between several sight being seen by eye infected with DR (Group A) and DME (Group B) [17]

1.3. Optical Imaging Modalities

Optical imaging (especially for retina) has experienced progress in the past century [28] to provide better knowledge of the eye in wellness and illness. Important developments have occurred in hardware as well as for image analysis software. In this section, we will present different ocular imaging techniques that have been made to improve the visualization of ocular pathophysiology.

1.3.1. Fundus Photography

The easy to use fundus photography is a widely available machine that captures pictures of the retina and optic nerve from 30- to 50- degree as shown in Fig. 18. Almost all ophthalmologists own a fundus camera that is very useful at registering the appearance of the optic nerve and presence of blood accumulation, but the treatment decisions made based on the captured views are rare [29].

Although recent cameras produce colored fundus images with high resolution, fundus photography has not encountered any significant transformations since the last 50 years; improvements that are more novel, introduce developed capabilities for generating color photographs of the back and outside pole with computerized software.

The fundus dyeing process is applied using some dyes like indocyanine and fluorescein to build-up a colored fundus. Besides, a new stereo fundus imaging system represents the amount of the reflected light from several view angles for depth resolution.



Fig. 18: Color Fundus photograph of a healthy person [30]

1.3.2. Adaptive Optics (AO)

Adaptive Optics (AO) is an emerging imaging system that attempts to enhance the performance of an ocular system by decreasing the consequences of wave-front deformities [31]. AO uses active ocular components to compensate for irregularities in the visual path between the target and the camera. It includes three main elements: a wave-front sensor, an adjustment element, and a controller system [32].

The efficiency of AO is to produce cellular-level imaging of retina by adjusting ocular irregularities. In recent times AO has been integrated with some primary imaging devices, such as AO-SLO, AO-OCT. Fig. 19 depicts a presentation of 2 clinical AO system prototypes with similar contrast and resolution, where (A) Photograph of the imagine system and (B) its corresponding acquired image. (C) Photo of the scientific integrated system and (D) its corresponding acquired image.

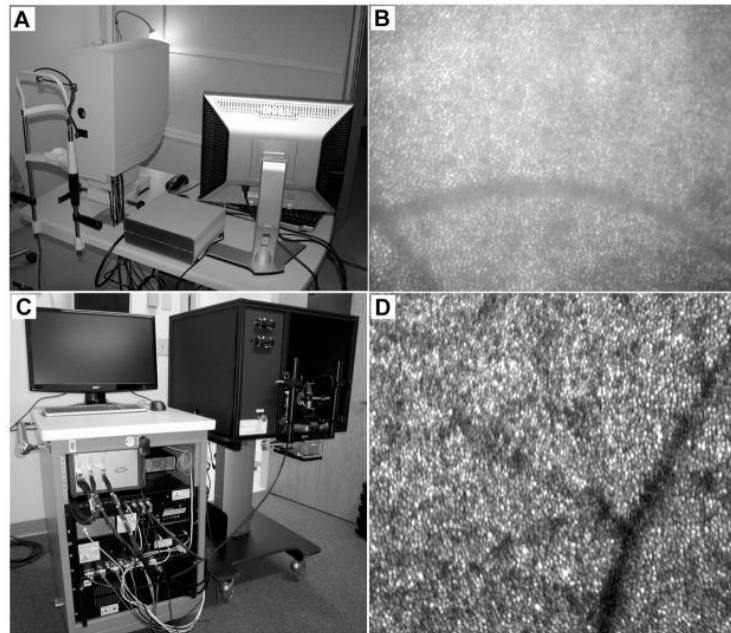


Fig. 19: An illustration of 2 AO systems with its corresponding image [33]

1.3.3. Scanning Laser Ophthalmoscopy (SLO)

Scanning Laser Ophthalmoscopy (SLO) uses a single low power, monochromatic confocal laser to acquire an image of the cornea, the retina, and the optic nerve head [34]. The obtained images show higher contrast than regular fundus images since they can decrease the factor of dispersed light.

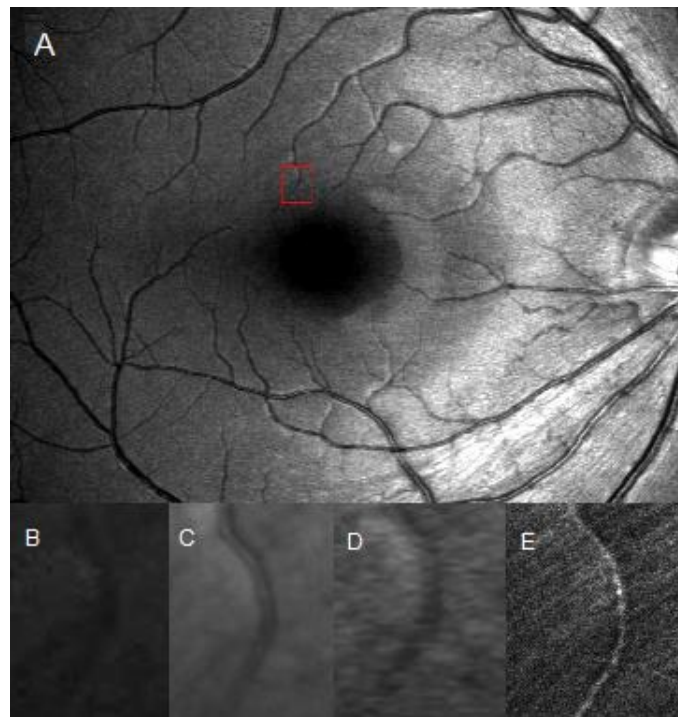


Fig. 20: Images of the normal retinal nerve fiber layer using AO-SLO [35]

A powerful enhancement in results is achieved by combining SLO to AO since AO-SLO has the capability of real-time observation of photoreceptors and vessels which is essential in diagnosing retinal ailments such as DR, and DME as presented in Fig. 20 where (A) represents a SLO image of an eye. (B): Magnified blue-channel fundus photography image of the region within the red box in (A). (C): Magnified red-free SLO image of the same red region in (A). (D): Magnified red-free AO-SLO image of the red area in (C). (E), Magnified AO-SLO image of the red area in (A). The resolution and contrast are higher in the AO-SLO images (E).

1.3.4. Photoacoustic microscopy (PAM)

Photoacoustic microscopy (PAM) is a safe, non-invasive visual excitation using ultrasonic exposure. A low pulse laser in nanosecond duration illuminates and attracts a spot tissue, thus producing ultrasonic pressure waves that have been focused on the tissue surface and record the signals to generate an image [36]. PAM can image retinal vessels besides the RPE with a better contrast ratio than any other retinal imaging modality. Quantitative ocular information can be obtained using PAM that can determine the anatomic formations of the eye. PAM is important to assess the role of oxidative damage, hypoxia, and ischemia in the pathogenesis of ocular pathologies.

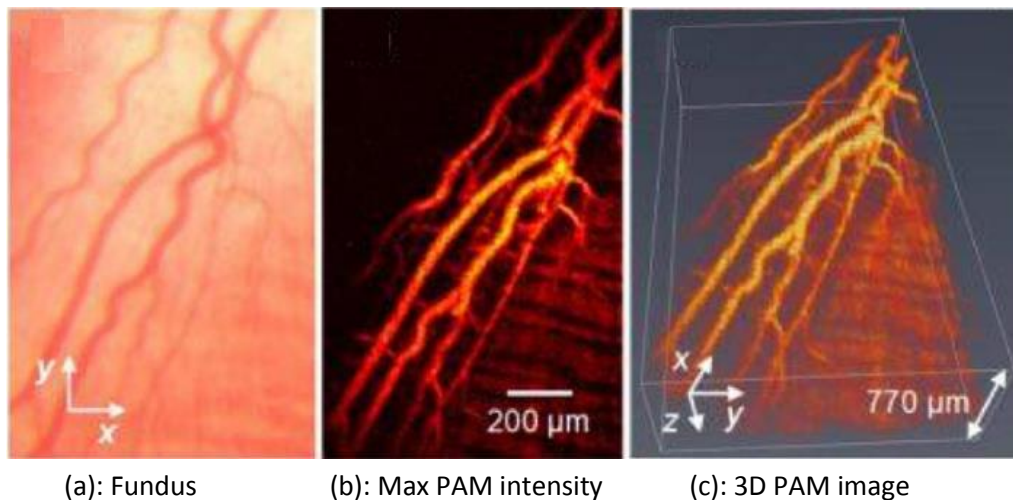


Fig. 21: PAM imaging presentation of a Rabbit [36]

When PAM is combined with other imaging techniques, (especially SD-OCT), more structural and useful information can be provided [37]. SD-OCT can acquire retinal structural information whereas PAM can provide molecular details of biological tissue. Nowadays, PAM stills in the development process without any licensed eyes clinical system existence [38], and all the imaging work has been performed in vitro. Fig. 21 presents a PAM imaging of eye retinal blood vessels for a rabbit where (a) presents a close-up look of the retinal vessels, while (b) shows a

projection of PAM waves with maximum intensity, (c) renders the PAM projection with a 3D volumetric projection. However, its improvement may considerably increase the range of retinal scanning in the future.

1.3.5. Ultrasound Biomicroscopy (UBM)

UBM provides screening of eye pathologies including the anterior and obscured area by overlying opaque pathologic optical structures. It offers diagnostically crucial information in diseases such as cysts, neoplasms, and glaucoma given that it measures the time between each high radiofrequency pulses (35–50 MHz) reflected by the ocular tissues to provide 2D scans as seen in Fig. 22. Furthermore, it presents significant biometric information concerning anterior structures, such as the cornea and its layers and the posterior chambers. Although it has been in use presently for more than a decade, including transducer arrangements, and aggregation of light with ultrasound, provide the potential for meaningful advances in 3D high-resolution diagnostic imaging structure of the eye [39].

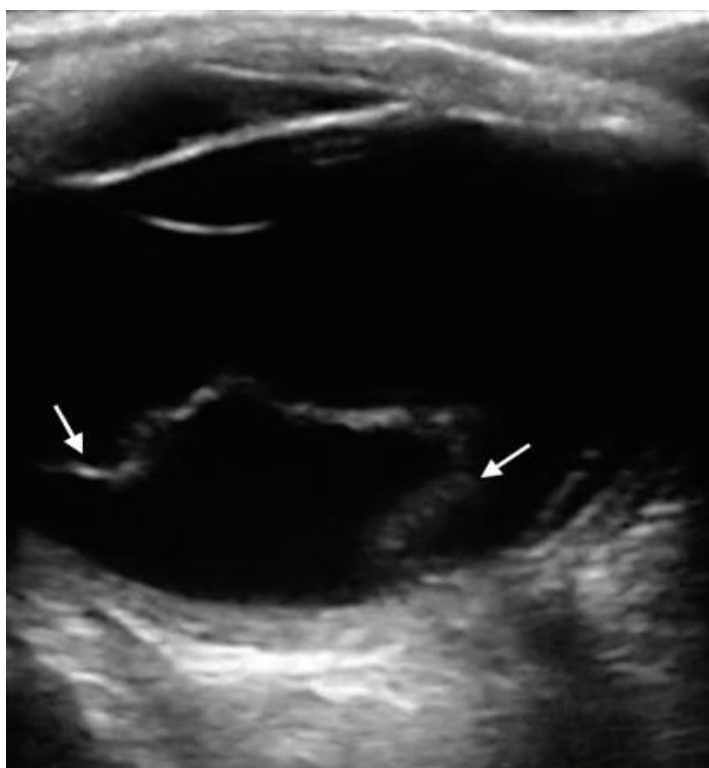


Fig. 22: An Ultrasonography presentation of retinal detachment [40].

1.3.6. Magnetic resonance imaging (MRI)

MRI is a nuclear magnetic resonance-based screening modality that provides deeply detailed high-resolution physiological photographs of the internal structure and anatomical large field of views, without depth limitations of the eye [41]. MRI

produces excellent structural details and tissue contrast for clinical analysis considering at a precise radiofrequency pulse, the nuclei particles spin to retrieve signals that involve information about the physical structure of molecules. It can capture quantitative blood-flow variations in the retina. MRI has the capability to offer unique information on how retinal blood-flow is controlled. The density of the nuclei and the internal frame affect the magnitude and the decay of the signal. Fig. 23 presents an MRI scan of both Rat and Cat retina including inner and outer layers.

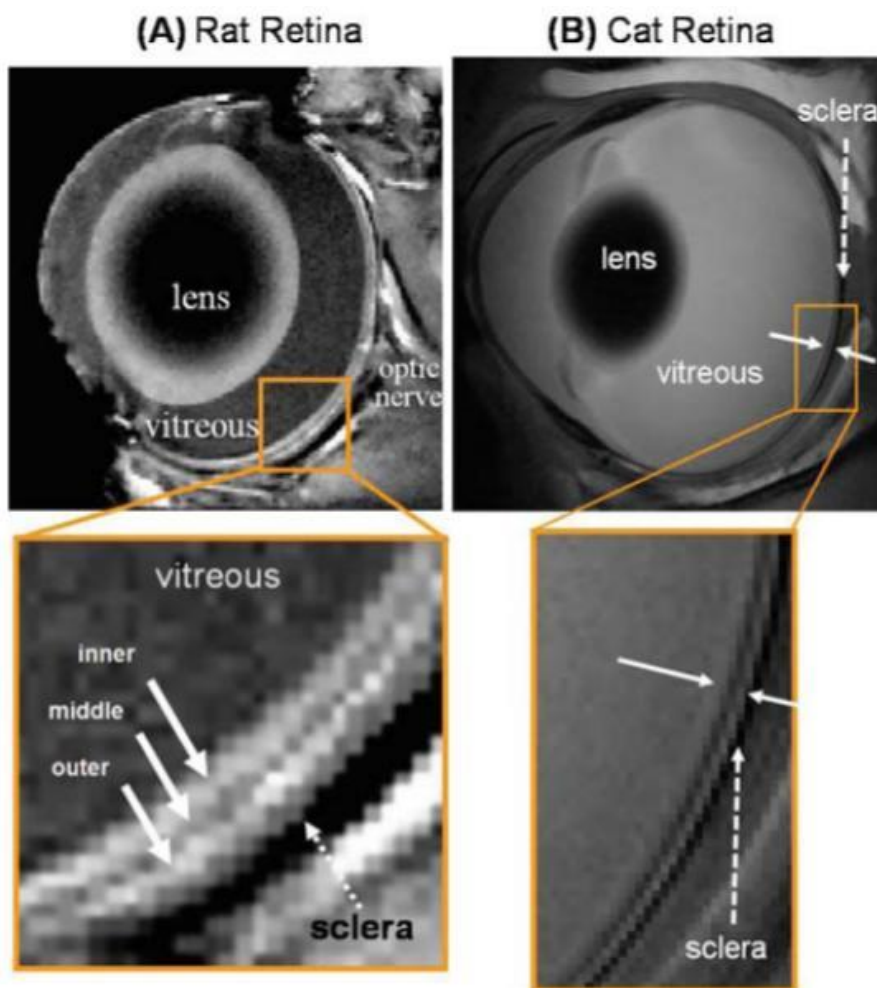


Fig. 23: A presentation of anatomical MRI of a Rat and Cat retina [41]

1.3.7. Optical Coherence Tomography (OCT)

Optical coherence tomography (OCT) is a non-invasive modality applied for cross-sectional imaging without contact. It captures the eye vessels and the subsurface structures that cannot be reached by different optical systems or operations. Using OCT imaging, ophthalmologists detect notable retinal layers that enable them to map and scale these layers that contribute to treatment [42]. It is a type of diagnostic equipment where a fast infrared laser emission is pointed at the

patient's eye, with a beam reflected at a mirror for real-time visualization. The absorbed noisy light is filtered out leading to only consistent light being caught, producing HD images of tissue structures [43]. OCT is an important imaging modality for its fast development into clinical perspectives, given the advantages that it advises clinicians. The light needed for OCT imaging is quite low to use in sensitive eye tissue and structures [44]. OCT is widely available in market just after 5 years since its inception in 1991 [45]. OCT has become the recommended modality for imaging diabetic diseases like DME and DR. It can provide volumes of retinal layers that can be segmented, allowing measurement of thickness, leading to improve diagnosis, where thinning of the tissue fiber layer indicates the onset and growth of the disease.

Compared to ultrasound medical mechanisms, OCT uses similar principles of waves where beams are oriented to the examined tissue. The rapid echo waves are reflected and scrutinized using IR light delay range to unveil the depth. This delay cannot be measured immediately, therefore an interferometer is used, where a portion of the beam is pointed to the sample and another part is directed to a reference arm. The idea of estimating low-interference is the policy, where time cohesion is a property of the light source and characterizes the continuous period of waves transmitted by the source, and measured at a specific point in space. Wave chains exiting from a light source of low temporal cohesion simply maintain a stable bond phase during a very limited period of time, which corresponds to a restricted travel range, coherence length, or coherence gate. A light source with a wide spectral frequency range consists of a set of wavelengths. The interferometer divides the emitting light from a source into two separate paths and collects the light coming from the two tracks at the output of the interferometer. The associated light strength can be measured as an electrical signal using a light detector [46].

In the initial execution of OCT [47], the reference length was changed for several scan depths. This modification is described as time-domain OCT (TD-OCT) and the principal structure is presented in Fig. 24. For every sample examined, the reference wave is scanned in-depth path and the light strength is recorded on the image detector. Thereby an amplitude scan (A-scan) for the whole depth characterization of the reflected sample is produced. A-scan utilizes a singular one-dimensional wave to estimate the density of the aimed structure. Diagnoses can be well-performed if the aiming is accurate at the targeted formation. A-scan has the capability to identify internal swelling structure and composition. Meanwhile, to generate a brightness cross-sectional scan (B-Scan), the wave is examined alongside the sample. B-scan applies a two-dimensional array of one-dimensional A-scan waves (produced at higher MHz frequencies) to determine the density and generate valuable images. B-scan is helpful due to its two-dimensional detailing.

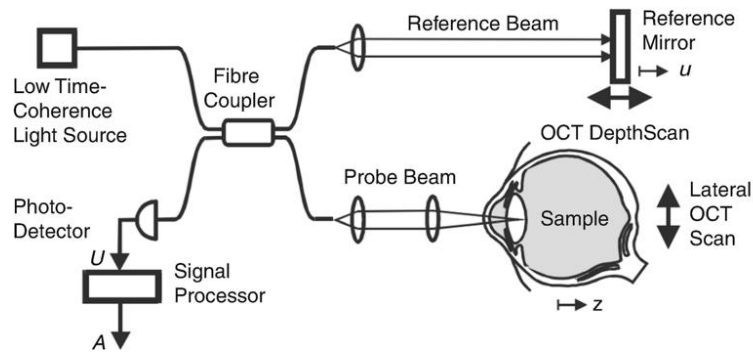


Fig. 24: Operating method of TD-OCT: light emitting from the light source is divided into the reference wave and the middle wave. The echo light received is joined again and recorded by the detector [48]

Thus, several one-dimensional scans (A-scan) are performed at different depths to create a two-dimensional image (B-scan). Those B-scans, if obtained closely and quickly, can be translated into a volumetric image (C-scan) of a retina. There are three principal benefits of OCT over older traditional methods: non-invasive, rapid scanning, and three-dimensional figures generation. Each volumetric OCT scan can take several seconds in comparison to about 20 minutes for techniques such as fluorescein tomography. However, despite the speed of OCT, the patient must be held still during scanning since the body motion and eye blinking might present artifacts measurements into imaging that is common due to the high frequency of macular disease happening in aged patients. Typically, during OCT scanning, the blinking eye generates black lines across the image, and degeneration of the signal is caused by the patient's motion [49].

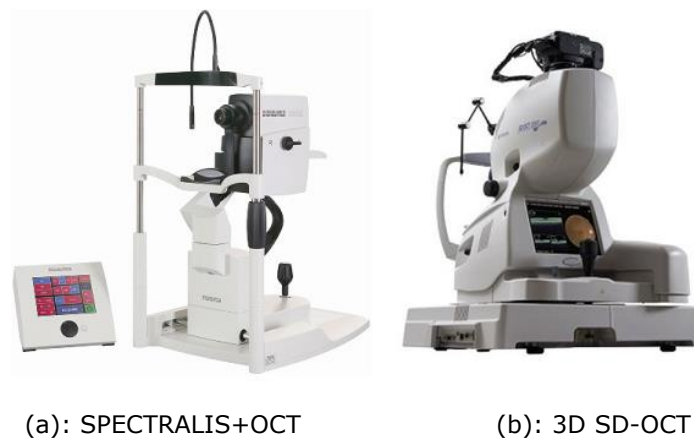


Fig. 25: Demonstration of 3D OCT-2000 SD-OCT machine and SPECTRALIS OCT device [52, 53]

Fourier domain OCT (FD-OCT or SD-OCT) is the next generation of OCT. It provides an effective implementation of the interferometer. In contradiction to TD-OCT, FD-OCT uses spectral data to produce A-scans without any mechanical scanning for the optical depth length. FD-OCT was initially introduced by Fercher et al. in 1995 [50]. The principal structure is represented in Fig. 26. SD-OCT is related

to TD-OCT, but the detector point is changed by a spectrometer. The spectrometer uses a diffractive component to divide the various emitting waves into a line-image captured by a fast line-scan camera.

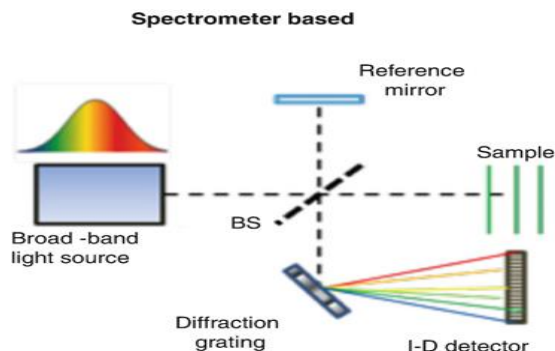


Fig. 26: Optical implementation of spectrometer based OCT (SD-OCT) which contains a spectrometer for wave division. (Diagram is taken from Drexler et al. [51])

SD-OCT instruments operate greatly as a spectrometer performs, allowing exposed results with single exposure only. These SD-OCT systems generate image resolutions higher than 12 MP within 5 microns. Fig. 25 (b) illustrates an example of an SD-OCT machine named: "3D OCT-2000 Spectral Domain OCT from Topcon Medical Systems", which is a system to integrate HD camera (12.3 MP) [52]. The appearance of SD-OCT offered the capability to surmount the restrictions of older OCT techniques. SD-OCT was able to improve image quality and capturing speed, allowing to simultaneously imaging the entire depth information. In addition, the SPECTRALIS device has been developed by Heidelberg Engineering [53] combines SD-OCT technology with a scanning laser fundus as seen in Fig. 25 (a). It was the first platform that has been proposed and it helps to find the fundus intended to be scanned with the cross-sectional OCT. This combination allowed precise motion tracking for re-scanning at the same position. Fig. 27 shows a recorded sample of the eye fundus using SPECTRALIS. The left image is the optic nerve tissue. The green line indicates the selected location of the OCT B-scan showed on the right image.

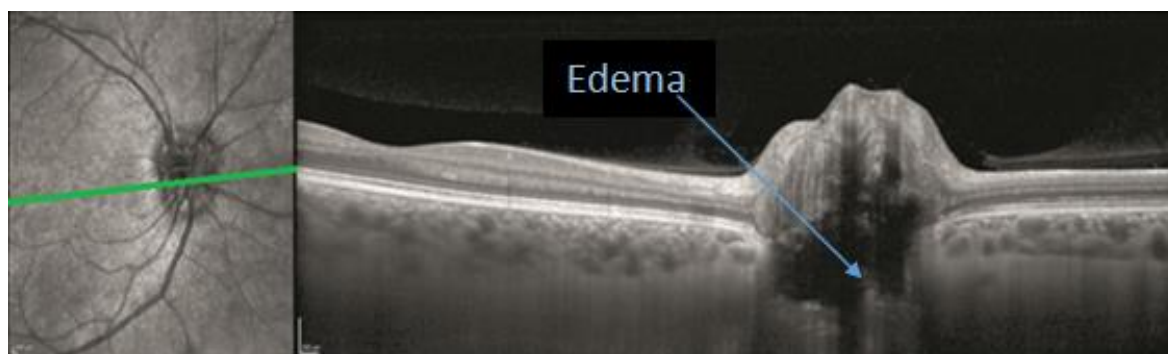


Fig. 27: The left image is a fundus optic nerve tissue captured by SPECTRALIS machine where the green line is represented on the right image as OCT cross-sectional B-scan . The Edema area is presented with blue arrow [54].

1.3.8. Optical coherence tomography angiography (OCTA)

Optical coherence tomography angiography (OCTA) is a non-invasive photograph modality aimed to reveal human retinal vascular networks [55]. OCTA utilizes low-coherence interferometry to cover variations in backscattered signals to distinguish blood-flow areas from static tissue areas. OCT needs a very high density in order to obtain the required resolution of the samples to discover the thin retinal capillaries.

During scanning, the variations of bulk tissue are excluded in order to control patient movement, guaranteeing that all detected variations are related to red blood flow [56]. OCTA becomes widely applied clinically to detect different ophthalmological disorders, such as AMD, DR, glaucoma, and artery/vein occlusions.

Fig. 28 presents samples of OCTA corresponding to a young Caucasian woman's right eye. (A) Full-thickness (ILM to BM) 3 x 3 mm OCTA. (B) Full-thickness 6 x 6 mm OCTA. (C) Consequent OCT B-scan. (D) 3 x 3 mm retinal nerve OCTA of the retinal nerve. (E) 3 x 3 mm retinal GCL OCTA. (F) 3 x 3 mm OCTA of the internal retinal depth.

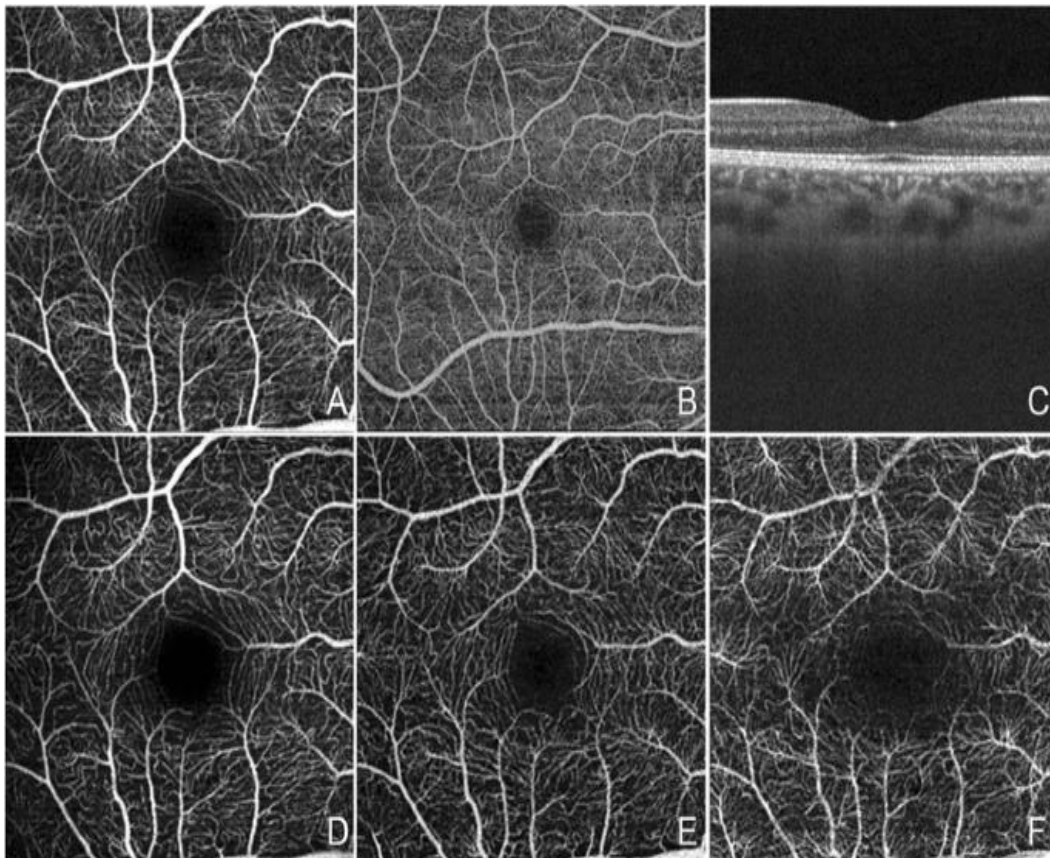


Fig. 28: OCTA Fields of View [56]

1.4. Deep Neural Networks:

Machine learning is a study domain of artificial intelligence that allows systems to learn automatically and to develop from practice without or with limited human intervention. It concentrates on the improvement of computer programs that can collect data and produce models in order to build solid decisions according to previous observations or stored data.

Machine learning systems are regularly classified as unsupervised or supervised methods.

- In supervised learning method, a model is learned on specific known data simultaneously with its corresponding labels.
- In unsupervised learning method, the labels are useless for the addressed application. Consequently, it allows for studying how systems can understand functions to determine unknown structures from unlabeled data. Semi-supervised learning is a different method aiming to utilize small-labeled datasets and large unlabeled datasets.

In opposite to traditional networks, recent work presents an important focus on deep learning where multiple fully or partially connected layers perform learning from the dataset. This dataset is passed progressively over a lengthy deep series of layers. These architectures have existed for a long time since the expression "deep" was presented by Hinton et al. [57] in 2006, the performance of multiple layers of a neural network is considering effective by pre-training a single layer and considering the other layers as an unsupervised Restricted Boltzmann Machine (RBM). In 2007, Bengio et al. [58] produced the Stacked AutoEncoder (SAE); this deep model concatenates several Auto-Encoders. Each AutoEncoder owns three layers, firstly a visible input layer, another hidden layer, and finally a regenerated layer with the same input size.

Finally, the well-known architecture is the Convolutional Neural Network (CNN) [59]. CNN contains several layers divided into 2 parts: convolution part (feature filtering) and pooling (down-sampling). The series of Conv/Pool layers is resolved regularly by a logistic regression layer in order to predict the class of the input image. 2D Raw images fed into CNN offer the benefit of shortly been processed by performing connections and balancing weights with sub-sampling.

However, it is apparent that deep architectures especially CNNs, bear complicated processing, which requires extremely robust computation machines. Fig. 29, Fig. 30 and Fig. 31 present examples of architectures for a Restricted Boltzmann Machine, an Auto-Encoder and a Convolutional Neural Network.

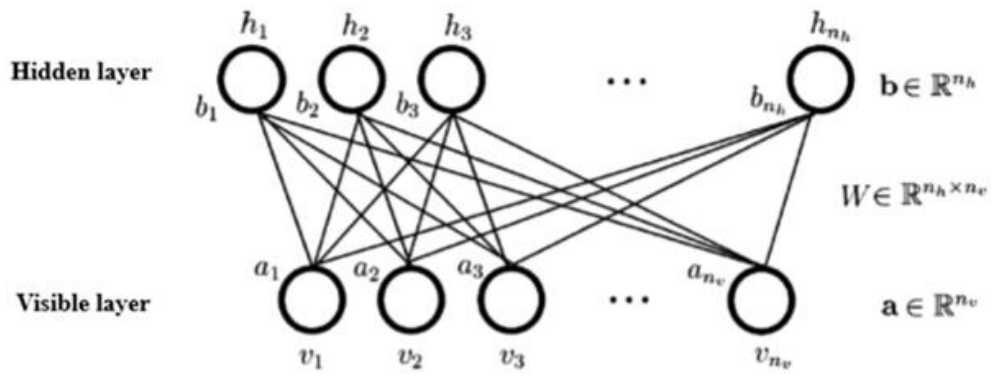


Fig. 29: An example of RBM network architecture [60].

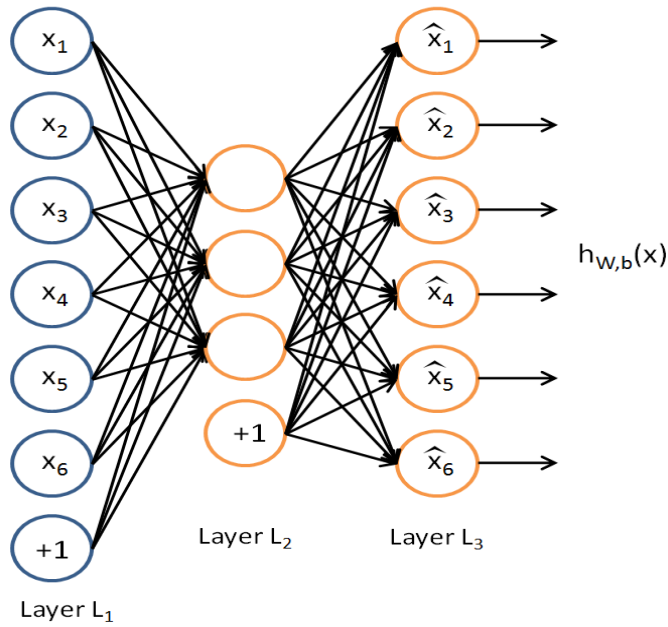


Fig. 30: An example of AE network architecture [61]

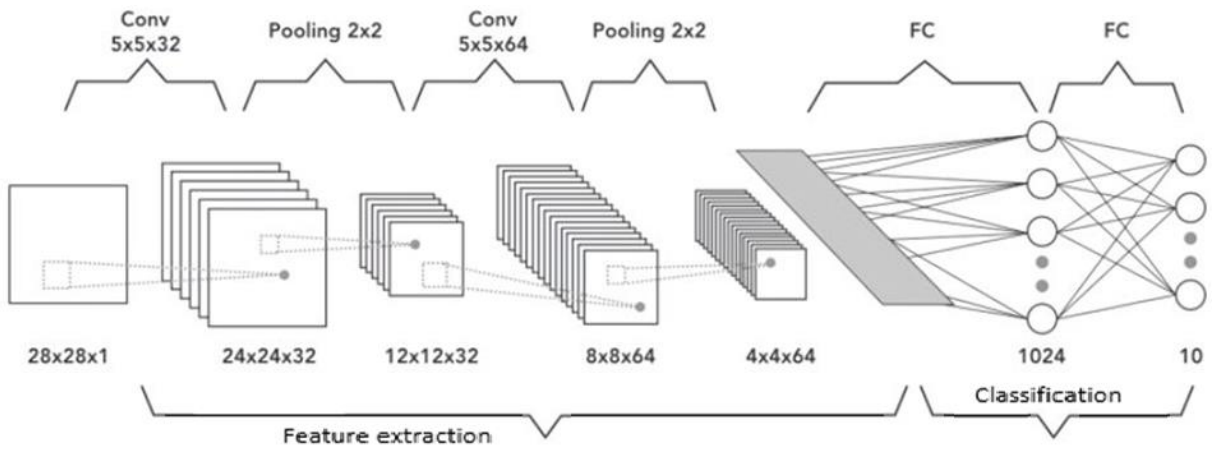


Fig. 31: An example of CNN model architecture [62].

1.4.1. Deep learning applications

Deep learning requires a large volume of data and enormous mathematical abilities, opening the skills of artificial intelligence to reach rational thought, and that lies in the program itself; much like the mind of a young child is patchy, but his flexibility is limitless.

Deep learning has been proposed to resolve problems linked to different domains of research. In optical remote sensing, for example, both AE and CNN have been implemented in two diverse ship detection and classification models. Tang et al. [63] introduced a feature extraction method based on Stacked Denoising Auto-encoder (SDA) and feature classification method using extreme learning machine (ELM). This work cannot satisfy the conditions and demand for real-life practice since it was applied using simple and tiny datasets. A related work, presented by Ying et al. [64], a CNN model has been introduced as classifier, and achieving better detection and classification accuracy compared to AE.

In market financial forecasting field, Barra et al. [65] integrate several CNNs models to create a new system that anticipates changes in the market for greater gains and fewer losses, better than previous attempts to use artificial intelligence in managing portfolios. With the new system being able to analyze the layers of existing data that have accumulated over previous data, deep learning is moving forward with market expectations, and in this way simulates savvy investors' intuition more than acting as a robotic system. The network can also modify buying and selling methods based on current and past events. Taking into account current factors increases the validity of the decision resulting from both random guessing and trading algorithms that are not able to learn in real time.

In art field, a novel CNN model has succeeded in generating realistic, reliable face graphics, from just a quick scribble of the subject proposed by Chen et al. [66]. The model was trained in portraits of the faces of celebrities and has proven ability to transform scribbles into realistic images in a revolutionary style Fig. 32 illustrates an example of scribble renovation. It is suggested to use the technology to identify suspects in criminal investigations, or to facilitate moving images in movies and electronic games.

In the biomedical domain, RBM and AE were applied to resolve problems of deformities detection and classification of Electro-cardiogram impulse (ECG) [67,68] and Electro-myogram waves (EMG) [69]. Beside solving detection problems, deep

learning is also being widely adopted for resolving recognition problems like traffic signs and symbols [70], faces [71] and fingerprints [72].

The possibility of deep learning developing special mechanisms of thinking to reach advanced stages that resemble human beings has been a subject of controversy among scholars and philosophers, many of whom promised a form of science fiction until we reached in the present era to the first phases of this new orientation of machines.

Deep learning has already begun years ago to produce creative works. Like poems, plastic arts, and music. The new approach to developing the creativity of robots began with primitive works, similar to the writings and drawings of beginners or children; but the technical development and the promotion of deep learning of machines and providing artificial intelligence with massive data, apparently an unprecedented shift in this framework, may change our stereotypical view of creativity and our bias towards the idea that it is monopolized by our species.

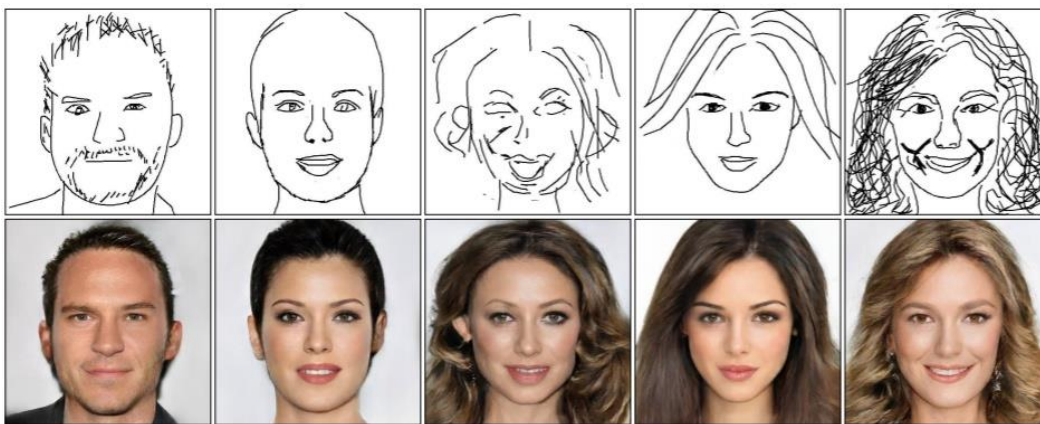


Fig. 32: Example of generated HD face images (Bottom) produced by handy drafts (top) and proposed by [65].

In this context, Bena et al. [73] have developed a system that can poetize in a professional manner, as simulating the emotions and feelings of human creators. The new system was fed with huge quantities of verses from poetry from various sources, with their classification within the categories that correspond to the type of feelings: Like sadness, anger, and joy, artificial intelligence brings new systems into its database with original verses. The new system showed the strength of automated texts to stimulate emotion similar to the emotions taken by the learned texts of human authors literary.

Moreover, in the sounds research domain, a group of researchers seeks to develop artificial intelligence that generates sound effects that simulate reality enough to fool the general public. These are news that may spoil the viewer's pleasure or cause the

aura of film magic to fade among its fans because many of the sound effects that we hear in films and television are prepared and later modified by Folie-Art artists.

Furthermore, in a recent study, a small group of participants fell into the scam. They thought the noise produced by the AI was real, according to EEE Spectrum reports [74]. Sometimes, participants chose AI voices as more realistic than real voices in the audio.

Ghose et al. [75] presented an artificial synthesis work of synchronized soundtracks for silent videos with DL where 41 out of 53 participants were deceived by the sounds generated by AI. This seems horrifying to the fate of the Folie-Art artists working on studio sound effects, but for now, the algorithm for producing simulated sounds is still far from perfect.

The researchers also created several systems for composing songs and music, including Dubai's song [76]. The world's first music track created by AI in May 2018, strongly renewed the debate about the ability of machines to think, consciousness and feeling, and to engage in one of the most specialized fields of humankind, linking the arts to emotions and human conscience.

Even if the creative aspects (Like literary works and figurative art) are among the most prominent aspects of human uniqueness, yet the AI companies strive to make progress in them, to enter the machines slowly and steadily, achieving noticeable improvements. For example, Google Deep Dream [77] learned the mechanism of object recognition by scanning millions of pixels by pixels, to learn at first how to distinguish between all colors and their grades; then to scan the boundaries of the spaces between objects, learn through time how to separate one object from another, and build an index for all objects of all the scanned images; then came up with a method for arranging and classifying objects with similar properties, and learning to reproduce random combinations of these objects.

A new AI project has shown just how cheap it can be to design celebrity faces with Deep-Fakes. The Black-hack Computer Conference showed a fake image of famous actor Tom Hanks developed by artificial intelligence using deep learning algorithms for machines. Tully [78] created a virtual version of Hanks, to see how easy it is to use artificial intelligence algorithms in disinformation campaigns. The new technology improved Hanks' image at a lower cost, although flaws appear when viewed at full resolution. Hanks' details remain familiar like the green color of his eyes and the shape of his eyebrows. To create the fake, he used a few hundred Hanks' photos and spent less than \$100 improving his facial recognition software to match his features. Faced with this advancement in face design techniques, ethical concerns emerge from their

disadvantages, as they are low-cost and easy, which may lead to an exacerbation of counterfeiting campaigns and the spread of misinformation on the Internet.

1.5. Conclusion

In this chapter, we have studied the eye structure and retinal disease pathophysiology involving a list of the optical pathologies that may hit the eyes. Moreover, screening techniques are described. After that, we introduced the different approaches of deep learning, and we pointed to the applicability of these methods in different domains of research.

To be mentioned eventually, neural networks establish a denominator of all the discussed methodologies in this thesis, the applications remain attached conceptually. Therefore, the subsequent chapters were managed with the necessity of preserving passage of the chapters sequentially. Although we assumed that, the readers are familiar with typical theories and notions of AI, CV, ML, and DL. Contrarily, it is recommended to follow the references mentioned in each chapter.

In the next chapter, we will present an overview of data pre-processing methods and classification techniques.

Chapter 2

Introduction to Data Pre-Processing and Classification in Machine Learning

Summary

2.1. Introduction	53
2.2. Data pre-processing in ML.....	53
2.2.1. Dataset gathering:	53
2.2.2. Dataset Cleaning:.....	54
2.2.3. Dataset splitting:.....	54
2.3. Classification tasks in ML.....	54
2.3.1. Unsupervised classification (Clustering):	55
2.3.2. Supervised classification:	55
2.3.2.1. Standard terms in classification algorithms.....	56
2.3.2.2. Types of classification:.....	56
2.3.2.3. Types of classification learners:	56
2.3.2.3.1. Lazy learners	56
2.3.2.3.2. Eager learners.....	56
2.3.2.4. Classification algorithms and techniques	56
2.3.2.4.1. Softmax classifier:	57
2.3.2.4.2. Artificial Neural Networks:	58
2.3.2.4.3. K-Nearest Neighbors (KNN)	59
2.3.2.4.4. Support Vector Machine (SVM)	59
2.3.2.4.5. Decision Tree	61
2.3.2.4.6. Random Forest.....	62
2.3.2.5 Classifier performance evaluation	62
2.3.2.5.1. Holdout method.....	62
2.3.2.5.2. Cross-validation.....	63
2.3.2.5.3. Confusion matrix (Precision and Recall)	63
2.3.2.5.4. Receiver Operating Characteristics (ROC curve).....	65
2.4. Conclusion.....	66

2.1. Introduction

The development of IT methodologies and algorithms has created a large amount of data in different fields and in various structures. These generated data can be stored locally in several formatting or distributed in multiple database servers, or publically published on the Internet. The importance of well-presented datasets has been discussed and shown in different studies and research fields. This stored data has been considered as a precious factor of decision making in solving detection and recognition problems.

Data preprocessing in ML/DL is a significant step that serves to improve the data quality and helps to obtain essential insights from the data. In order to produce a suitable data for creating and training DL algorithms, several methods have been covered to prepare and manipulate raw data such as normalizing, organizing, and cleaning. It is a kind of data mining system that converts raw data into meaningful formatting data. Thereby, data analysis has been considered as the major step to extract knowledge from stored information. It presents a tool for processing digital structured data for any data type such as natural text languages, images/videos, and sound, etc...

Data pre-processing is the opening action pointing to the start of the process in a DL model. In reality, data is deficient, discrepant, inexact (with errors), and usually requires specific characteristic values. To solve this, data preprocessing rebuilds this data lacking and deformation, in addition, it helps to clarify, setup, and prepare the raw data to be ready to be trained by DL algorithms.

2.2. Data pre-processing in ML

Data pre-processing in ML is a data-mining technique that changes raw data into a comprehensible, readable and clear format. In order to perform this transformation, several iterative steps have to be achieved:

2.2.1. Dataset acquisition:

An initial step to build models is to acquire the appropriate dataset collected from multiple and dissimilar sources to be combined in a conventional format to create a dataset. It contains different steps:

- Data selection, transformation, or combination: select and retrieve related data from the data sources/databases to study, where data are converted or consolidated into an appropriate and suitable form for performing tasks.

- Data integration: at this phase, various unrelated sources of data may be merged in a common source.

2.2.2. Dataset Cleaning:

Data cleaning is a stage where incomplete or inappropriate data are excluded from the set. It is crucial to recognize and precisely handle the absent values. Otherwise, incorrect and wrong results and conclusions from the data set might be faulty drawn. Typically, there are three methods to deal with missing data:

- Removing a specific column or row having a null/empty value, taking into concern that the rule requires more of two-third of the values to be missing. Nevertheless, this method is inefficient and unwise to use with inadequate dataset samples.
- Computing the median for numeric features of column or row with null value by substituting the result for the lost value that can make a variation to the dataset.
- Approximating values of linear data: Another way of estimation is the deflection of neighboring values.

2.2.3. Dataset splitting:

Dataset splitting is an essential procedure in ML models where the dataset must be divided into two isolated sets – training sets (for training the ML models with a size ratio usually varies from 70% to 80%) and test sets (for testing the trained models and predict results with a size ratio normally from 20% to 30%). The splitting method takes into consideration different variables like the shape, type, and size of the dataset. The training part of dataset is fed into classifier that is a necessary step in which intelligent methods are used to extract useful features and to recognize the surely useful patterns and features based on various criteria.

2.3. Classification tasks in ML

Classification is a task that necessitates the application of "learning from data" of ML algorithms, where the training dataset leads to obtain more reliable, observed features, used to define specific target class type. Generally, a prediction task of the target class is the subsequent step following the extracted features.

In this thesis, we are motivating to use classification and prediction; also, we can observe that there are two categories of classifications:

- Supervised classification: classes are identified in advance; they normally possess an associated substance.
- Unsupervised clustering: classes are based on the formation of objects; the substance correlated with classes is not easy to define. It is necessary to determine the measures of aggregation.

2.3.1. Unsupervised classification:

The concept of clustering is to attempt to group the related, similar, and alike kind of objects (documents) by considering the most satisfying situation; everything in the same collection should be similar. Moreover, two distinct group parts should not be similar. Unsupervised classification is applied when there are objects that are not classified. At the end of the process, the objects must be related to one of the classes produced by the classification. There are two classes of unsupervised classifications: hierarchical (HC) and non-hierarchical (NHC).

In the HC, the generated subsets are nested together hierarchically. We recognize the descending HC that begins from the group of all the objects and divides these objects into a particular number of subgroups, next each subgroup being split into another number of subgroups, and so on. Furthermore, the ascending HC begins with an individual group that is divided into subgroups, and so on. To decide which classes we are going to join, we use the aggregation criterion where it compares classes to choose the most similar classes based on certain measures, like the nearest neighbor, the average distance, etc...

In the NHC, objects are not hierarchically structured. We notice the term "partition" when each object is only part of a subgroup. Otherwise, "recovery" is the term where each object may refer probably to certain groups.

2.3.2. Supervised classification:

Supervised classification is the method used regularly for the analysis of quantifiable image data. A training dataset has been used to get fitter shapes that could be applied to detect the destination class type. Once these conditions are defined, the following task is to predict the class.

Different algorithms are convenient to do the task. Some can manage predefined classes that overlap each other (called soft classification methods); others generate firm borders between classes (called hard classification methods).

2.3.2.1. Standard terms in classification algorithms

- Classifier: An algorithm that outlines the input data to a particular class.
- Classification model: A classification model attempts to form some result from the input provided from training. It will predict the class type for the new data.
- Feature: A feature is a specific measurable characteristic of an input data being observed.

2.3.2.2. Types of classification:

- Binary Classification: a task with two potential results (True/False, Yes/No).
- Multi-class classification: a task with more than two class classification results. Each sample unit is mapped to only one target class label.
- Multi-label classification: a task where each sample unit is assigned to more than one class target labels.

2.3.2.3. Types of classification learners:

2.3.2.3.1. Lazy learners

It is a type of learner based on comparing a new test sample with the stored training data. The classification is handled based on the most similar data in the stored one. Lazy learning methods are usually slower to evaluate. Different classification algorithms may be found like k-nearest neighbor (KNN).

2.3.2.3.2. Eager learners

Eager learner's methods build a classification model based on the provided training data before accepting data for classification. It must be capable to perform to a single hypothesis that satisfies the whole sample space. Due to the model architecture, eager learners necessitate a long time to train and less time to predict. Different classification algorithms may be found like Decision Tree and Artificial Neural Networks (ANN).

2.3.2.4. Classification algorithms and techniques

Based on the application purpose and nature of the possessed dataset, different classification algorithms are now available in ML with no conclusion that one is above the others. Among the classification techniques, in this thesis, we will be present the used classifier listed by some state-of-the-art methods.

2.3.2.4.1. Softmax classifier:

Softmax classifier is a class of multi-class Logistic Regression based on Softmax function. Contrary to other functions such as RELU and Sigmoid, Softmax classifier is a type of activation function used for classification usually; it uses the cross-entropy loss and provides "probabilities" for each class. The Softmax regression is a class of logistic regression presented in Fig. 33 that normalizes input into a vector of values that result in a probability distribution where the sum of the total is up to one. The Softmax regression model calculates first a score for each class, and then estimates the probability of each class by applying Softmax function to the scores. In ML, Softmax is employed through a neural network layer added with the same number of nodes as the output layer. The Softmax function is presented in equation (1) as below:

$$P(y = j | z^{(i)}) = \phi_{Softmax}(z^{(i)}) = \frac{e^{z_k^{(i)}}}{\sum_{j=0}^k e^{z_k^{(i)}}} \quad (1)$$

Where z is an input defined in equation (2) as:

$$z = w_0x_0 + w_1x_1 + \dots + w_mx_m = \sum_{l=0}^m w_lx_l = w^T x \quad (2)$$

The softmax function calculates the probability that the training sample $x^{(i)}$ belongs to class j knowing the weight and the net input $z^{(i)}$ which is a vector containing the score of each class for the instance $x^{(i)}$. Note that, w = weight vector, x = feature vector of a single training sample, and w_0 = bias unit.

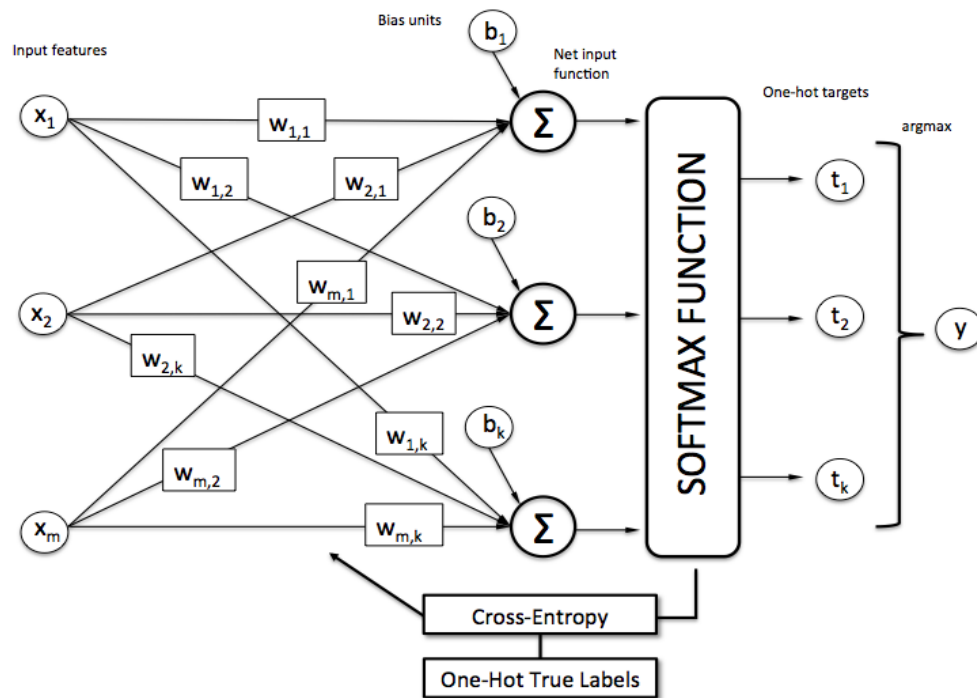


Fig. 33: : Softmax Regression [79]

Therefore, Softmax regression estimates the probability that the instance $x^{(i)}$ belongs to class j given the scores of each class for that instance. Moreover, Softmax regression classifier predicts the class with the highest estimated probability. To achieve this prediction model, a cost function is defined to be minimized using an optimization algorithm, which leads to estimates a high probability for the target class and hence a low probability for the others. Cross-Entropy is a type of cost function that penalizes the model when it estimates a low probability for a target class. Cross-Entropy is the most known hinge loss of Softmax regression in discriminative models that is regularly applied to calculate losses that can be assumed when training a dataset. This loss computes the difference between the predicted output to the actual output using a cost function J and a loss function H as presented in equation (3).

$$J(W) = \frac{1}{n} \sum_{i=1}^n H(T_i, O_i) \quad (3)$$

$J(W)$ is the average of all cross-entropies over the n training samples where T stands for the true class labels, O is the actual output and i is the number of layers. The Cross-Entropy function is defined in equation (4) as:

$$H(T_i, O_i) = - \sum T_i \cdot \log(O_i) \quad (4)$$

2.3.2.4.2. Artificial Neural Networks:

ANN is a collection of related input and outputs known as connectionist systems. The network in the learning stage learns by modifying the weights of the processing inputs examples to be capable to predict the right class label. Several network architectures are now possible like convolutional, feed-forward, etc. The suitable architecture hangs on the purpose of the model. For image processing, feed-forward models provide fairly reliable results but convolutional networks have superior performance.

The number of multiple hidden layers in the architecture depends on the intricacy of the task that is running to be outlined by the model. More hidden layers lead to complex multi-connected models like deep networks and take a lot of time to train and modify weights. Fig. 34 shows a demonstration of two hidden layers of ANN architecture where all neurons are connected. Due to the unknown characteristic significance behind the learned weights, ANN has inferior model interpretability in comparison to model like Decision Tree. ANN performed better with continuous and consecutive inputs/outputs values based on real applications in the world, and capable to classify inexperienced exemplars with high sensitivity to noise. ANN is a type of eager learners considering it trains earlier a model to practice it for later prediction.

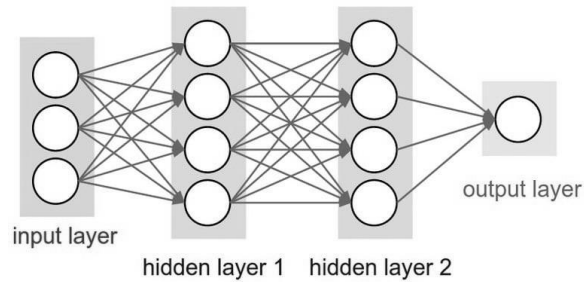


Fig. 34: A flowchart overview of ANN architecture [80]

2.3.2.4.3. K-Nearest Neighbors (KNN)

KNN classifier is a lazy type non-learning algorithm based on the majority vote of the case's neighbors to classify a new case among all available stored cases. The case is related to the class type which is most apparent amidst its k nearest neighbor calculated by a distance function, like Manhattan and Euclidean (used for continuous variables), and Hamming (used for categorical variables), where k is an integer number handy hardly specified. It weights the participation of every of the k neighbors depending on their distance utilizing the following query providing superior weight to the nearest neighbors.

$$w \equiv \frac{1}{d(x_q, x_i)} \quad (5)$$

KNN is a model that simply classifies items based on similar feature but it is computationally consuming in time and resources. Moreover, a pre-processing step as denoising is necessary before proceeding for KNN and variables must be normalized else higher scale variables may bias it.

2.3.2.4.4. Support Vector Machine (SVM)

SVM is a supervised ML algorithm that is mostly used in classification challenges. It is a discriminative classifier formally defined by a dividing hyper-plane (in multidimensional spaces). Given labeled training data, the algorithm outputs an optimal hyper-plane that classifies new samples, such as a line (called decision boundary) in two-dimensional space that separates the data points in two parts wherein each class lay on either side. The perfect decision boundary is the one that has the longest distance from the nearest points of these two classes (called support vectors). The gap that the closest points define around the decision boundary is called the margin. As a result, SVM firstly finds lines or boundaries that properly classify the training dataset, then, it chooses the one that has the max distance from only the closest points (high Gamma).

Theoretically, in the Maximum Margin Classifier (MMC), the idea is to make the margin as large as it can be, therefore, the threshold boundary must lay in the halfway between these closets points. Thus, this classifier cannot achieve zero tolerances with perfect separation due to outlier observation. Since MMC is extremely sensitive to outliers in the training dataset, it is better to Tradeoff by allowing misclassification of the outlier observation to solve the sensitivity issue. This misclassification permit is an example of Bias/Variance Tradeoff that plagues all of ML. The Regularization "C" parameter tells the optimizer how much you want to evade misclassifying each training example.

In other words, before we allowed misclassification, we picked a threshold that was very sensitive to the training dataset (low bias), and it performs poorly when we got a new data sample (high variance). In contrast, when we picked a threshold that was less sensitive to the training data and allowed misclassification (higher bias); it performed better when we got new data (low variance). When we allow misclassification, the distance between the observations and the threshold is called Soft Margin (SM).

In order to achieve the optimal SM, we use Cross-Validation to determine how many misclassifications and observations to permit inside of the SM to get the best classification. Therefore, when we use SM to determine the location of a threshold, we are using an SM Classifier (called Support Vector Classifier "SVC") to classify observations.

In real-world applications, finding an optimal class for millions of tons of overlapped training datasets takes a long time, and neither MMC nor SVC don't perform well with this type of data due to a lot of misclassifications. Therefore, Support Vector Machines (SVM) can handle this issue by (1) starting with data in relatively low dimension; (2) move the data into a higher dimension (data transformation); (3) find an SVC that separates the higher dimensional data into two groups. In order to choose the best mathematical data transformation, SVM use Kernel Functions to find systematically SVC in higher dimensions.

In data transformation, we can use Polynomial Kernel with a parameter "d", which stands for the degree of the Polynomial; when $d=x$, the Polynomial Kernel computes the relationships between each pair of observations in x-Dimension, and these relationships are used to find SVC. In summary, Polynomial Kernel systematically increases dimensions by setting the degree of the Polynomial "d", where a good value of d is found using Cross-Validation. Another very commonly used Kernel is the Radial Kernel (called Radial Basis Function Kernel "RBF") where

it finds SVC in infinite dimension. The nearest neighbors have a lot of impact on organizing the new observation by using their classification.

Therefore, Kernel functions only calculate the relationships between every pair of points as if they are in the higher dimension; they do not actually do the transformation. This trick, calculating the higher dimensional relationships without actually transforming the data, is called the Kernel Trick that reduces the amount of computation required for SVM by avoiding the math that transforms the data from low to high dimensions and it makes relationships in the infinite dimensions used by the Radial Kernel possible.

2.3.2.4.5. Decision Tree

Decision Tree is a decision-maker algorithm graphically represented like a tree model. It breaks the data into two or more homogeneous parts (leaf) depending on the most important predictors in the input variables. It generates a series of rules that can be applied to classify the data. To determine a differentiator, the algorithm analyzes all features and produces a binary separation on them, where the categorical data is divided by category, and for continuous data, a cut-off threshold is applied. It will next pick the one with the lowest cost with the highest accuracy and iterates recursively until divides the data into all sides successfully or by reaching the deepest depth.

DT needs lightly data preparation that is used for numerical and categorical data classification. Furthermore, it can create complicated trees that do not generalize properly, as well as small variations of data may lead to feeble decision trees. An example of a decision tree is shown in Fig. 35.

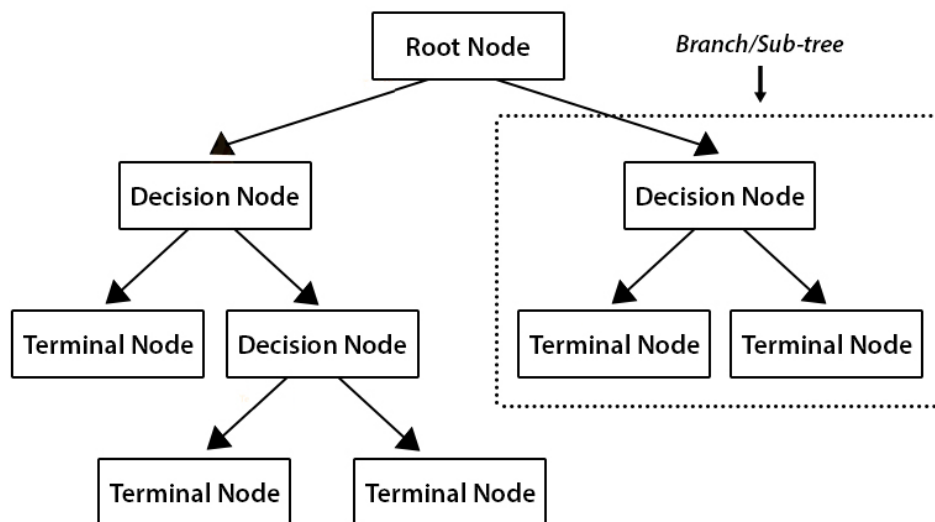


Fig. 35: Decision Tree classification algorithm architecture [81]

2.3.2.4.6. Random Forest

Random forest is a supervised algorithm usually used for classification. It offers several trees and classifies items depending on the votes of the majority of the trees. This is a classifier that works as an estimator to matches a number of decision trees on several samples of data and calculates average that increases the accuracy of the model and limits over-fitting.

RF can manage high dimension extensive dataset with missing data while keeping the accuracy. An example of a RF is presented in Fig. 36.

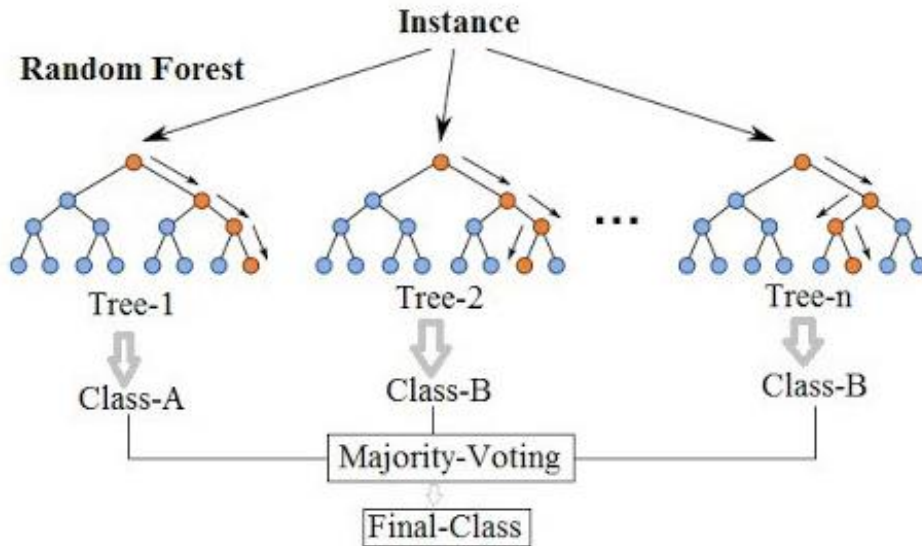


Fig. 36: Random Forest algorithm outline [82]

2.3.2.5 Classifier performance evaluation

The evaluation of the classifier performance is the most fundamental procedure to prove its applicability after training a model. Since supervised and unsupervised classifiers are learned on a calculable training dataset, the trained classifier has to be evaluated experimentally on a different examination test data, where the performance on the multi-set is a representative for the achievement on unseen data. It checks the classifier's generalization capability. Therefore, there is a need for a measure function estimating the classifier performance empirically.

2.3.2.5.1. Holdout method

The holdout is the most frequent method where the given dataset is split into two sets, to test and train a model with a ratio that is set to 20% and 80% or 30% and 70% respectively. The train partition will be applied to learn from the dataset and train the model, while the unseen test partition will be utilized to test the performance of prediction.

2.3.2.5.2. Cross-Validation

Cross-Validation is a classifier validation method where the dataset is randomly split into k mutually exclusive sub-groups, each almost with identical size, and one is held for experimentation while others are utilized for training. This process is repeated during the entire k -folds. K -fold Cross-Validation can be handled to prove that the model is not over-fitting ensuring that there is no overlap between the validation samples. An example of a 10-fold is presented in Fig. 37.

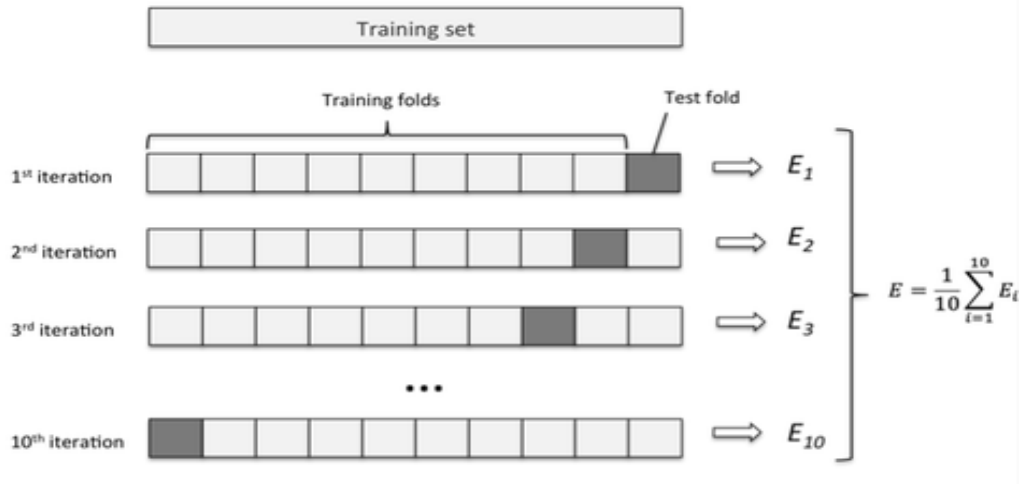


Fig. 37: A representation of 10-fold Cross-Validation [83]

2.3.2.5.3. Confusion matrix (Precision and Recall)

Confusion Matrix is a type of error measurement matrix that measures the performance of a supervised classifier. The confusion matrix presents visually in a layout of the contingency table matching the predicted class (formed by column) with the actual class (formed by row) of the objects forming the sample. We have two types of information: the number of times the model was mistaken and the type of error when grading.

Table 1: Confusion Matrix for binary classification

		Predicted Class	
		X (+)	Y (-)
Actual Class	X (Yes)	$V_{X,X}$ (TP)	$V_{X,Y}$ (FN)
	Y (No)	$V_{Y,X}$ (FP)	$V_{Y,Y}$ (TN)

Table 1 presents a confusion matrix example for a model of two classes X and Y where the performance is assessed through the values of the table by comparing the assignment situation and the origin situation in order to estimate the error rate. In this table, $V_{X,Y}$ represents the number of cases of class X assigned to class Y and $V_{Y,X}$ represents the number of cases of class Y assigned to class X, while $V_{X,X}$ and $V_{Y,Y}$ represent the correct number of classification.

A confusion matrix is built from the four values generated as a result of classification. A binary classifier predicts all test samples as either negative or positive. This prediction provides four results – true positive (TP) for the correct prediction as positive, true negative (TN) for correct prediction as negative, false positive (FP) for incorrect prediction as positive and false-negative (FN) incorrect prediction as negative.

From this confusion matrix, we can identify six common types of measures:

- **Error Rate**

Error rate (ERR) is measured as the sum number of wrong predictions (FN + FP) divided by the entire number (P + N) of the dataset where ERR=0 is the optimal result.

$$ERR = \frac{FP+FN}{TP+TN+FP+FN} = \frac{FP+FN}{P+N} \quad (6)$$

- **Accuracy**

Accuracy (ACC) is computed as the sum number of all correct predictions (TP + TN) divided by the entire number of the dataset (P + N), where ACC=1 is the optimal result, moreover ACC + ERR= 1.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TP+TN}{P+N} \quad (7)$$

- **Sensitivity (Recall or TP rate)**

Sensitivity (SN) or recall (REC) is measured as the number of correct positive predictions (TP) divided by the entire number of positives (P) where SN=1 is the optimal sensitivity.

$$SN = REC = TPR = \frac{TP}{TP+FN} = \frac{TP}{P} \quad (8)$$

- **Specificity (TN rate)**

Specificity (SP) or (TNR) is determined as the number of correct negative predictions (TN) divided by the entire number of negatives (N) where SP=1 is the optimal specificity.

$$SP = TNR = \frac{TN}{TN+FP} = \frac{TN}{N} \quad (9)$$

- **Precision (Positive predictive)**

Precision (PREC) or (PP value) is determined as the number of correct positive predictions (TP) divided by the entire number of positive predictions (TP + FP) where PREC=1 is the optimal PPV.

$$PREC = PPV = \frac{TP}{TP+FP} \quad (10)$$

- **False positive rate (FP rate)**

False positive rate (FPR) is determined as the number of wrong positive predictions (FP) divided by the entire number of negatives (N) where FPR = 0 is the optimal value. Moreover FPR+TNR=1.

$$FPR = \frac{FP}{TN+FP} = 1 - TNR = 1 - SP \quad (11)$$

- **F1-Score**

F1-Score is a classifier measurement for the harmonic mean - ranging from 0 to 1 - between precision and recall. It is a measure of test accuracy where a greater precision rate accompanying with less recall rate provides a remarkably accurate result; nevertheless, a large missing number of samples are complex to classify. Higher F1-Score leads to better model performance.

$$F_1 = 2 \cdot \frac{PREC \cdot REC}{PREC + REC} = 2 \cdot \frac{1}{\frac{1}{PREC} + \frac{1}{REC}} \quad (12)$$

2.3.2.5.4. Receiver Operating Characteristics (ROC curve)

ROC curve is a graphical chart that represents the distinguishing efficacy of a classifier method, as its discernment threshold is diverse. It is commonly used for measuring the performance and the accuracy of classification models, which shows the conciliation between the true positive (TP-Sensitivity) and the false

positive (FP-Specificity). A perfect accuracy model should have an area of 1.0. Fig. 38 shows a demonstration of ROC Curve chart.

The area under the ROC Curve (AUC) calculates all the areas beneath the whole ROC curve, it presents a measure of performance overall potential classification thresholds, and measures the quality of the predictions regardless chosen classification threshold. AUC varies value from zero to one where AUC=1 signifies a 100% correct predictions.

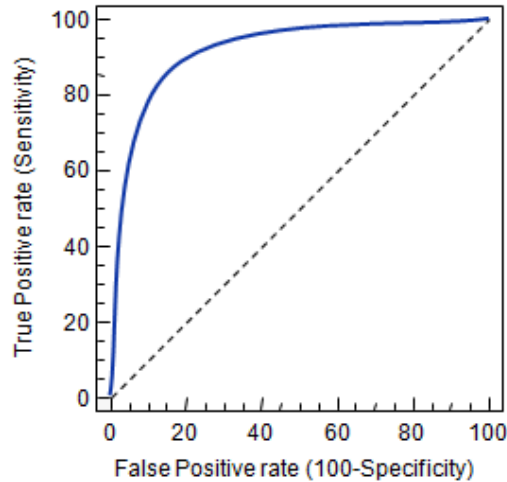


Fig. 38: A representation of ROC Curve plot [84]

2.4. Conclusion

In this chapter, we presented the different definitions and notations of data processing as a pre-step for the classification process in machine learning. This pre-step is considered as an important phase that provides improvement of the qualified data that assists to reach overview of the dataset. We presented also that the pre-processing phase is considered as a main step that helps to extract features, besides it aims to organize and prepare the dataset to become appropriate to be fed into ML learning process.

Moreover, the different types, learners, and models of classification have been defined, besides the description of the concepts of the algorithms and techniques of classification have been presented also. Finally, a presentation of different evaluation methods has been included. The following chapter presents the related work for retinal disease segmentation and classification.

Chapter 3

State of the Arts

Summary

3.1. Introduction	67
3.2. Literature review.....	67
3.2.1. Conventional ML analysis	67
3.2.2. Deep learning and Neural networks models	70
3.2.2.1. Segmentation approach	70
3.2.2.2. Classification approach	73
3.3. Perspective and motivation	79
3.4. Conclusion.....	81

3.1. Introduction

In this thesis, we are attracted to image processing for classification using CNN based methods of DL. More specifically, we are working on the detection of two diseases, which are DME and DR. In this chapter, we present some methods compared to our work by covering their limitations. These similar works include image segmentation, classification, and feature extractions. Finally, we conclude this chapter by defining the circumstances and motivations, which hold us to present this work.

3.2. Literature review

3.2.1. Conventional ML analysis

The most initial techniques to identify and distinguish retinal pathologies from images involved various image processing methods accompanied by feature extraction plus classification. In this section, we evaluate some existing approaches for detecting diabetic pathologies based on image analysis.

In order to approach the problem of DME detection, Sankar et al. [85] proposed a model based on a semi-supervised learning model using the Gaussian Mixture Model (GMM) over OCT volumes, where the method begins with a pre-processing stage involving resizing, flattening, filtering. Then extracts features considering the intensity and Local Binary Pattern (LBP) after being reduced using PCA as seen in

Fig. 39. The proposed method allows the classification and identification of the abnormalities as outliers within the volume. Experiments are performed over two datasets with 32 OCT volumes for the first dataset; the results achieved a specificity and sensitivity of 93% and 80%, respectively. For the second dataset that includes 30 OCT volumes, the results achieved a specificity and sensitivity of 80% and 100%, respectively.

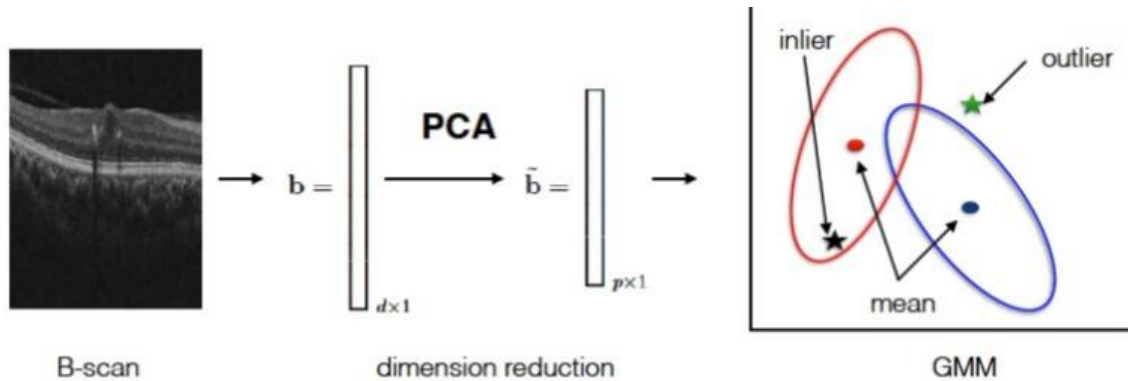


Fig. 39: Flowchart of the Gaussian mixtures model creation [85]

Srinivasan et al. [86] presented a classification system to differentiate normal, DME, and AMD SD-OCT volumes. These volumes were improved by firstly decreasing the noise using the sparsity-based block-matching and 3D-filtering (BM3D) denoising method where BM3D is a collaborative filtering process that group similar blocks from the image. A block is grouped if its dissimilarity with a reference fragment falls below a specified threshold- block matching. All blocks in a group are then stacked together to form 3D cylinder-like shapes. Filtering is done on every block group. Linear transform is applied followed by Wiener filtering, and then transform is inverted to reproduce all filtered blocks. Secondly flattening the curving of the retina and cropping the images. Fig. 40 shows an overview of the proposed algorithm. Next, histogram of oriented gradients (HoG) is used to extract features information for each slice of a volume, and finally, support vector machine (SVM) is applied as a trained linear method using a dataset of 45 patients evenly divided within three groups and produced a classification percentage of 100, 100 and 86.7% for normal, DME and AMD cases, respectively. This dataset needs to be refined since it has different sizes of scans in each volume. Besides, it does not offer a large variety of DME lesions.

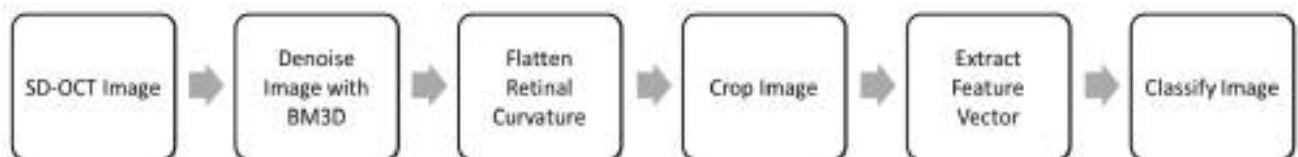


Fig. 40: Overview of the algorithm for classifying SD-OCT volumes [86]

Alsaih et al. [87] proposed a classification analysis of OCT data for detecting DME. First, a step of preprocessing including (i) denoising, (ii) flattening, and (iii) cropping has been used. Next, a concatenation of the extraction of the histogram of oriented gradients (HOG) and local binary pattern (LBP) features have been presented. A multi-resolution approach has been used as well as principal component analysis (PCA) and bag of words (BoW). Fig. 41 illustrates an overview of the presented algorithm. The results led to sensitivity (SE) and specificity (SP) of 87.5% and 87.5% respectively using PCA and SVM.

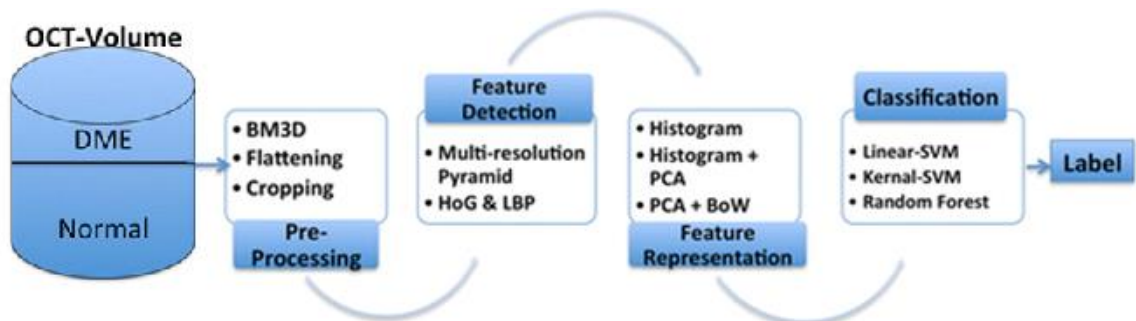


Fig. 41: Outline of the algorithm for preprocessing and classifying [87]

Another method has been proposed by Lemaitre et al. [88]; the framework combines different preprocessing steps in association with Local Binary Patterns (LBP) features and various mapping approaches. Using linear and nonlinear classifiers, results showed sensitivity and specificity of 81.2%, and 93.7% respectively. The framework has been tested on a balanced small dataset of 32 patients and the result is not considering reliable compared to novel work based on neural network. The outcomes of preprocessing steps are incongruous with several classifiers and feature configurations, and did not present a better performance result.

Liu et al. [89] suggested a model for detecting macular diseases in SD-OCT images using LBP. Each scan was adjusted by flattening and a multi-scale pyramid with three levels was built. Moreover, boundaries were distinguished by applying Canny detector. Next, a histogram of LBP was obtained for each level of the pyramid. The resulted histograms were concatenated into global descriptor and fed into a diminished dimension using PCA. Subsequently, a combination of classifiers has been applied using an SVM with an RBF kernel. Using a small dataset of 326 OCT scans, the methodology showed a considerable performance of DME or AMD, among an AUC of 0.93.

All methods described earlier presents a distinguished sensitivity and specificity score since it is trained and tested on the same dataset using a single

representation, where the decision is inaccurate when tested on a different dataset. This substandard score does not help the doctors to surround the anomalies precisely. Table 2 presents some methods for the traditional classification task.

Table 2: State of the art summary for retinal disorders detection methods using conventional ML analysis

Ref#	Disease	Dataset Size	Feature	Classifier	Evaluation	Results
[85]	DME	32	Pixel-intensities	GMM	SE, SP	80%, 93%
	DME	30	Pixel-intensities	GMM	SE, SP	100%,80%
[86]	DME, AMD vs Normal	45	HOG	SVM	ACC per Disease	100%,86.7%, 100%
[87]	DME	32	LBP, HOG	SVM, RF SVM-RBF	SE, SP	87.5%,87.5%
[88]	DME	32	LBP	KNN, LR, RF, GB, SVM	SE, SP	81.2%,93.7%
[89]	AMD, DME	326	Edge, LBP	SVM, RBF	AUC	0.93

3.2.2. Deep learning and Neural networks models

Nowadays, many papers have addressed the problem of DME classification using deep learning models since CNN [90] has shown exceptional performance when applied to medical images examination. CNN showed advantages in analysis over traditional machine learning, besides the immense ability to automatic feature extraction. Therefore, based on deep learning, several works have been reported a superior performance in classification and segmentation tasks.

3.2.2.1. Segmentation approach

Segmentation approaches attracted the interest of researchers because it supplies the physicians with more illustrated outlook allowing them to check the situation of the retina's layers more clearly, where the most remarkable way to recognize a patient with DME is the significant thickness of macular in the retina. Moreover, the recognition of possible cases is performed by the identification of the accumulation of fluids in the sub-retinal area.

Lee et al. [91] developed a fully automated segmentation method for OCT-based on CNN model with a total of 18 Convolutional layers. The SegNet CNN model segmented the images with a 0.911 Cross-Validated Dice Coefficient, compared with segmentation by experts and detected intra-retinal fluid (IRF) on OCT in a manner indistinguishable from clinicians. The features collected from the

segmented layer are employed to determine layer deformations that allow identifying the diseases. Despite using a very deep CNN architecture and trained over 200000 iterations, this method is slow and doesn't show a promising result as well as it requires additional human impact since the validation of the suggested model was examined using a tiny sized sample. Therefore, the notion of automated or semi-automated segmentation using deep learning has not shown popularity in medical image processing field yet since not all results are encouraging compared to other concepts.

Liu et al. [92] proposed a novel automatic model for the segmentation of OCT images using deep residual neural networks (ResNet) as a feature extractor to extract deep features, and structured random forests classifier for layer segmentation. The proposed method started by data preprocessing of a total of 110 OCT images from 10 subjects (normalization and denoising), followed by feature extraction, which involves network training. Next, a combination of these features with the extracted hand-designed features is presented in Fig. 42 (a). Finally training the structured random forest and predicting the boundaries of each layer for retinal layer segmentation as seen in Fig. 42 (b). The results show that the proposed method achieves an F1-score of 0.885. The size of labeled images in this method is small thereby data augmented methods can enhance the segmentation performance, DME is the only pathology addressed by this work, other pathologies may have similar deformation features compared to DME layers. This may lead to segmentation problems and conflict.

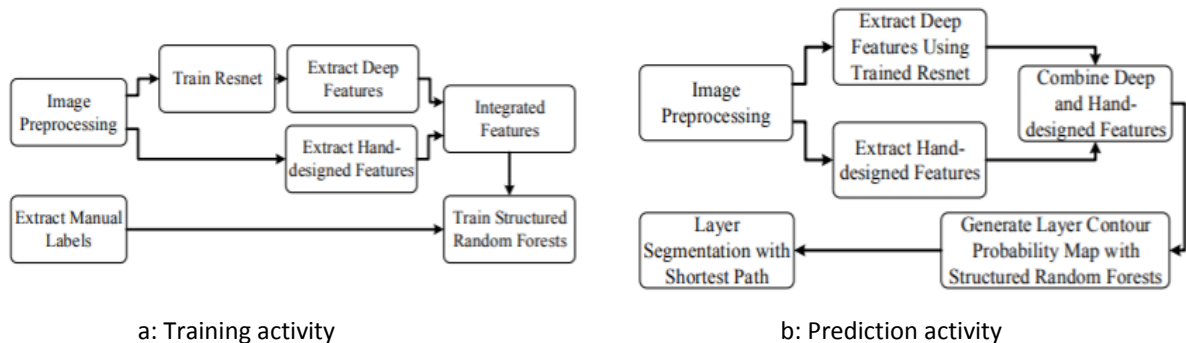


Fig. 42: Flowchart for OCT image [92].

Lu et al. [93] proposed a new structured model for the segmentation of multi-level fluid class in the OCT images of the retina. The three classes are: IRF intraretinal fluid (IRF), subretinal fluid (SRF), Pigment epithelial detachment (PED). A fully CNN was trained to recognize, identify, and label the fluid pixels based on the intensity of retinal images and segmentation of the retinal layers by a graph-cut algorithm. Random forest classification was conducted on the segmented regions to detect and discard the deceitfully labeled regions. The proposed model

achieved as a result of the segmentation performance a mean Dice of 0.766 and for the detection performance a mean AUC of 1.0. Fig. 43 presents a flowchart of the multiclass fluid segmentation. The new framework comprises a fully CNN and combines prior information via fluid features, a relative distance map, as well as post-processes by level sets and random forest-based classifiers to present the final three-class (IRF, SRF, PED) fluid segmentation. The level-set method is able to detect the regions with low intensity; it can distinguish the whole fluid region from the background

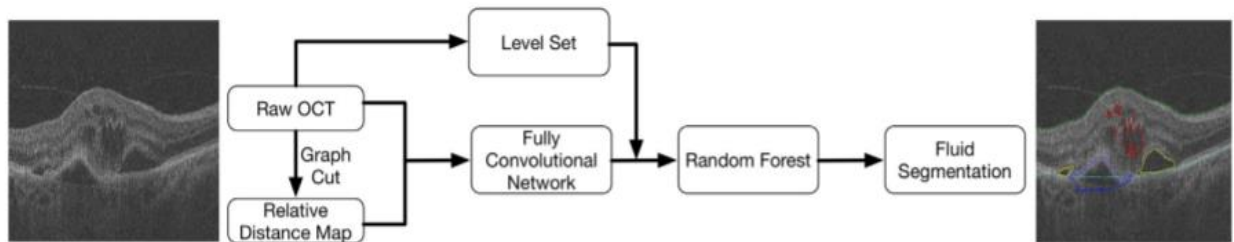


Fig. 43: Flowchart of multiclass fluid segmentation [93]

In additional, automated segmentation of cystoid formations in OCT images is proposed by Karthik et al. [94] where the new method aimed to find and surround cysts in 3D OCT slices using a CNN model intended to learn a mapping method. Finally, the segmentation is achieved via the clustering of the located cysts' positions. After training, the proposed method achieves a mean dice coefficient of 0.71 on a public dataset. Another experiment was examined by cross-validation using a private dataset and achieved a mean dice coefficient value of 0.79.

It is evident that segmentation approaches remained ineffective as a worldly identification system of retinal pathology, taking into consideration that comparing various retinal layers' information captured by separate machines still yet ineffective. Table 3 presents a summary of some segmentation method.

Table 3: State of the art summary for segmentation methods using neural network

Ref#	Pathology	Dataset Size	Pre-processing	CNN Model	Classifier	Evaluation	Results
[91]	DME, AMD	1289	Augmentation	SegNet	Sigmoid	Dice coefficient	0.911
[92]	DME	110	Norm, Denois	ResNet	RF	F1-score	0.885
[93]	DME, DRUSEN	70	-	U-Net	RF	Dice coefficient	0.766
[94]	DME, AEI	30	-	Novel	Sigmoid	Dice coefficient	0.71

3.2.2.2. Classification approach

Different deep learning-based classification works have been presented using either transfer learning, pre-trained network, or end-to-end model. Since algorithms require a sufficient amount of images and an elevated number of epochs for convergence, and since the employment of pre-trained models is helpful for faster convergence with less data.

Karri et al. [95] presented a model based on GoogLeNet [96] network, where the method aimed to classify OCT with models pre-trained on the non-medical ImageNet database which contains more than 14 million images grouped into more than 20000 categories. The model is applied to classify three cases: normal, AMD, and DME. The proposed model involves fine-tuning GoogLeNet to identify OCT images with pathology. In addition, it includes a preprocessing stage for filtering and flattening OCT image. By using cross-validation for the proposed model, the means of decision pooling across all validations for normal, AMD, and DME are 0.99, 0.89, and 0.86, respectively. After 10 experiments to achieve the optimal model, the prediction test showed a ratio of 96% of model accuracy via SVM classifier. This method predicts each image class but it should predict the subject class using the stack of images from an expected subject.

Fang et al. [97] introduced an iterative fusion CNN (IFCNN) methodology for the OCT image classification. The proposed network embraces an IF approach, which merges features from the preceding and current CNN layer, and therefore utilizes the several features of CNN layers to obtain a precise classification of OCT. The method includes two categories: the basic part (used to extract features of multiple layers) and the fusion part. In order to evaluate the classification's performance, a 10 folds cross-validation has been used for the suggested IFCNN model. Fig. 44 shows a demonstration of the proposed classification method. This model, which is applied to classify several classes (DME, Drusen, CNV, and Normal), revealed an overall accuracy of 87.3+2.2%. Compared with numerous famous classification methods, the results have revealed that the suggested method exceeds other related methods. Thus, despite some improvement in the performance shown by this IF method, if the amount of feature maps increases, the performance will degenerate to some amount while still considerably producing further computational cost.

Rasti et al. [98] proposed a diagnosis model to identify two types of macular diseases (AMD and DME) versus normal retina using an ensemble of multi-scale CNN to classify the SD-OCT dataset. To evaluate this diagnosis system, the

performance and the time complexity were examined based on two separate datasets. First, a preprocessing pipeline channel is used as shown in Fig. 45.

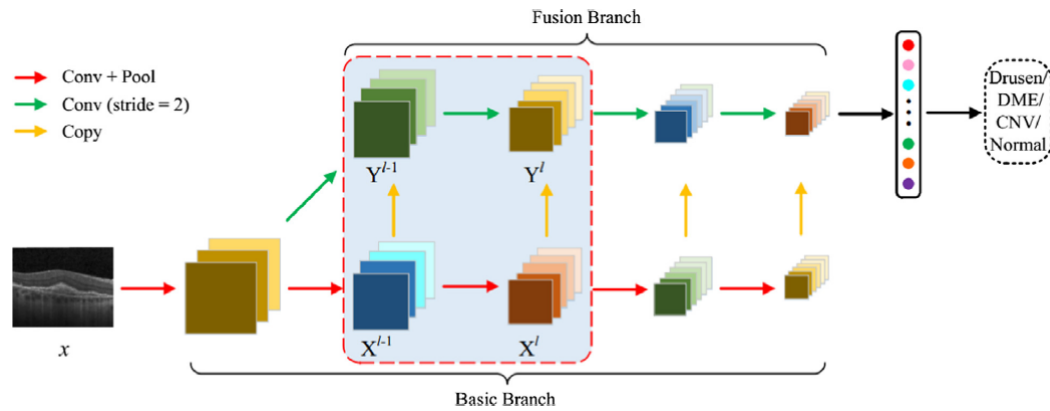


Fig. 44: Details algorithm for preprocessing and classifying [97]

Then a combination of CNN experts is applied as a powerful classifier. This mixed model is introduced from the notion of the "divide and conquer" strategy in ML literature. It mixes the outputs of different expert classifiers by training a gating network allowing each expert network evaluates the following pattern of the feature space being distributed by a gating network as seen in Fig. 46. The offered AUC model was 0.9985 and an overall average precision rate was 98.86%. To improve the model accuracy, a larger database with a larger mass of different cases should be significantly augmented, since the model handled two datasets (containing 193 retinal OCT volumes including AMD, DME, and normal subjects).

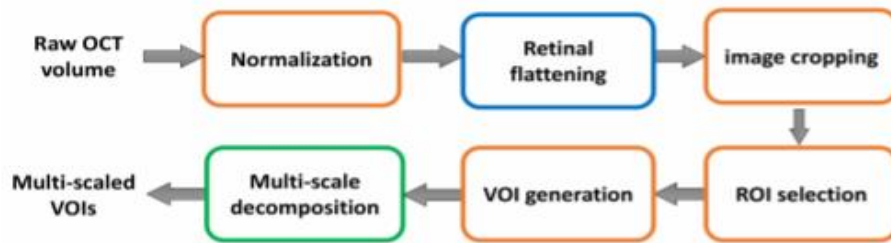


Fig. 45: The preprocessing steps [98]

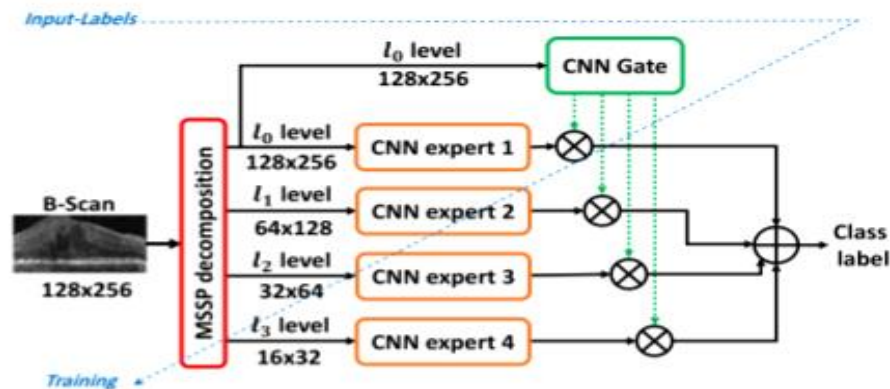


Fig. 46: Experts and CNN gate network are fed by precise scales of the input model [98]

Mehta et al. [99] proposed a novel neural network algorithm that expanded OCT images with meta-data such as age, gender, and visual data for multi-label classification since patients show various pathologies in the same time. The model was examined to classify four different diseases as follows epiretinal membrane, DME, AMD, and neovascular AMD (NVAMD). Two training strategies have been examined: an ImageNet pre-trained network was applied for transfer learning, either the randomly initialized weights have been trained by the network. Since OCT scans are grayscale images and ImageNet pre-trained network has been trained over colored images. Therefore, the learned features may not concentrate on the targeted disease features. The system achieves an overall accuracy of 86%.

Chan et al. [100] presented a classification method with feature reduction (Bow, PCA) for distinguishing patients with DME. The features of the retina have been extracted from the SD-OCT images using pre-trained Convolutional Neural Network (CNN). By using 8-fold cross-validation examinations, the accuracy results showed a percentage of 96.88 % using RF classifier. The model has experimented over 32 subjects (16 DME, 16 normal).

Ji et al. [101] presented an automatic transfer learning-based model to identify AMD and DME from SD-OCT images. The model uses the Inception V3 pre-trained neural network as features extractor by eliminating the last several layers. In order to learn the shifts in the feature space, the extracted features from the altered V3 have been fed to a CNN. The overall system accuracy was above 98% using two distinct retinal OCT images datasets.

Hwang et al. [102] presented an assisting tool for the diagnosis and treatment of AMD. To make this tool, three different CNN pre-trained models (VGG16, ResNet50, and InceptionV3) have been trained over OCT images. The results show that the used models performed an accuracy of 91.20%, 95.87%, and 96.93% for VGG16, ResNet50, and InceptionV3 respectively.

Nugroho [103] presented a comparative study between the traditional feature extractors and deep neural network features. The dataset is small and distributed in four classes (CNV, DME, DRUSEN, and normal). The training and testing images were regenerated into features vector using Histogram of Oriented Gradient (HOG) [104], Local Binary Pattern (LBP) [105], pre-trained CNN: Residual Network (ResNet50) [106], pre-trained CNN: Densely-Connected Network (DenseNet-169) [107]. The result showed an accuracy of 88% and 89% for DenseNet169 and ResNet50 (trained using ImageNet dataset) compared to 50% and 42% for HOG and LBP respectively.

Awais et al. [108] proposed a model for classifying DME pathology in OCT image volumes. The model aimed to extract features from these volumes by using a pre-trained VGGNet [109] architecture and classify these features using different classifiers. The feature vector provided by the VGG-16 network provided a representation with a size scaling from 4096 to 1000 depending on which level of the fully connected layer (FCC) the classifier is connected to as shown in Fig. 47. In order to evaluate the model, KNN is first used with $k=1$ and $k=3$ then 100 trees of random forest have been tested. The evaluation is completed using the "Leave one patient out cross-validation (LOPO-CV)" method. The majority rule is applied to classify the entire volume after evaluating each B-scan. Different experiments have been done over OCT volume. These experiments focused on data-preprocessing on the image being fed into the CNN model. Data manipulation focused on denoising and/or cropping raw images in separate experiments where the highest accuracy obtained at volume level was 93%. The model used a pre-trained model over ImageNet without fine-tuning and utilized a small raw dataset without any dimension reduction. These updates can lead to better results confirmed by the proposed work in this thesis.

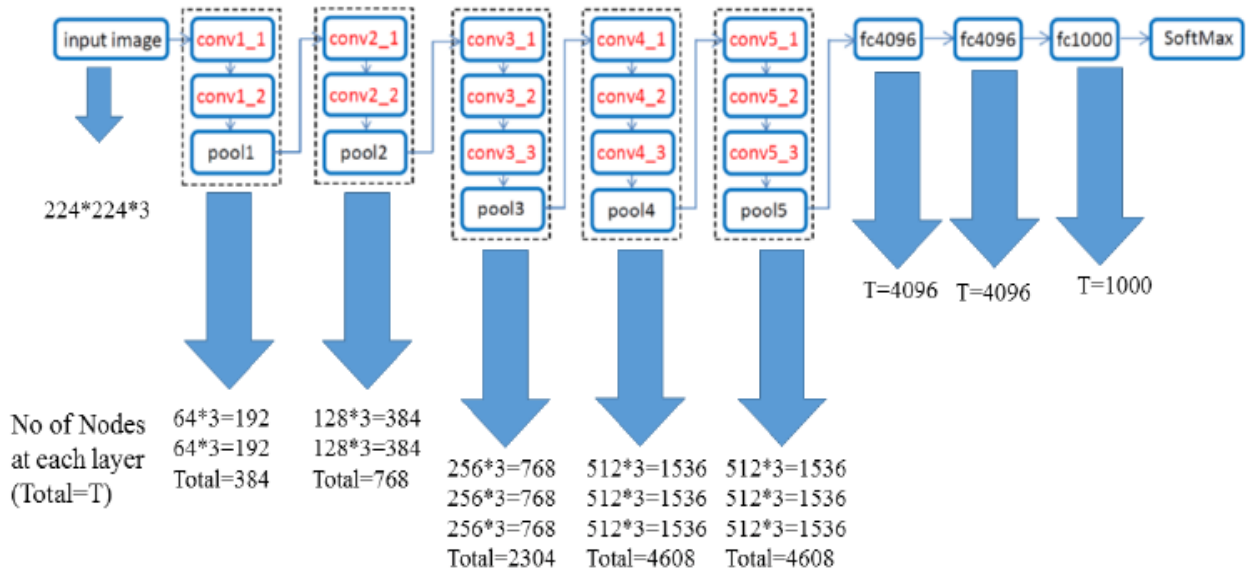


Fig. 47: VGG16 Layer block model used by [108]

Similarly, Feng et al. [110] proposed a transfer learned CNN model based on the pre-trained VGGNET network (With 16 Blocks) for automatic classifying AMD and DME in OCT images. A total of 207,130 retinal OCT images were picked from 5319 patients containing several types of macular pathologies for the experiment. After labeling and flattening the images, only 1000 images (divided equally for each pathology) from 22% of patients were selected for validation while the rest images from other 88% of patients were used as a training dataset. The method

utilized the transfer-learning approach to fine-tune the VGG-16 network and estimated its performance on the validation dataset. The results showed a prediction accuracy of 98.6%, with a sensitivity of 97.8%, a specificity of 99.4% retinal diseases. This model compared to [108], which are both based on VGGNet, showed better results in detecting not only DME but also AMD and used a larger dataset to train and a small part of data to validate. However, compared to the work proposed in this thesis. The proposed work showed better accuracy results in detecting two types of pathologies (DME and DR) with less network complexity and converge faster since a pre-processing step was leading to better focusing on the region on interest (ROI).

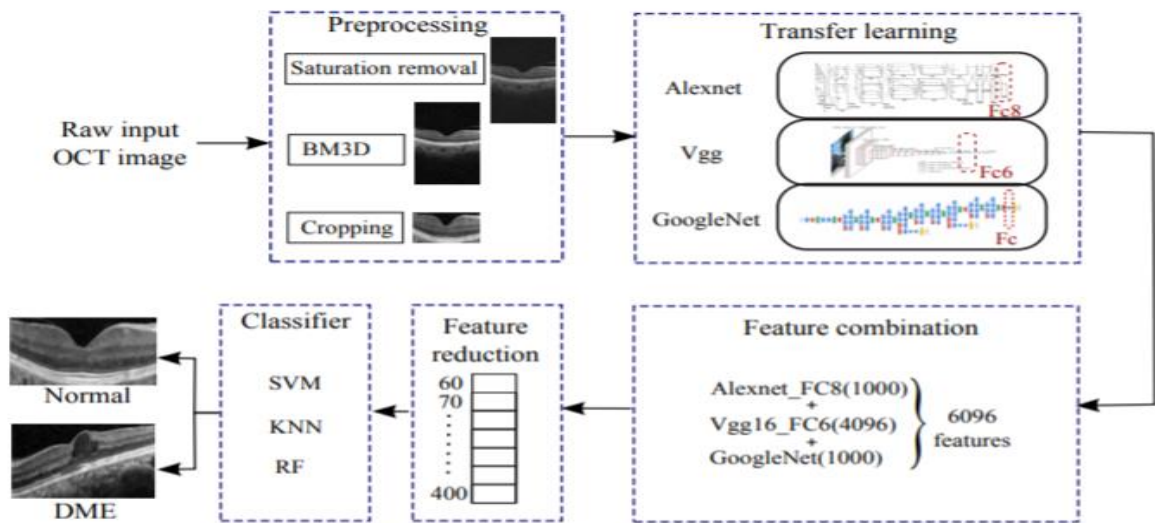


Fig. 48: Overview of combined framework of feature extraction using AlexNet, VggNet and GoogleNet proposed by [111]

Moreover, Chan et al. [111] presented a fusion work between different pre-trained CNN architectures to combine features: AlexNet [112], VGGNet, and GoogleNet, with performing a feature reduction using PCA. The results were evaluated using a "Leave-Two-Patients-Out Cross-Validation" LTPO-CV at the volume level. The model started by pre-processing steps on the SD-OCT volumes: Image denoising was applied first to enhance image classification results. Next, the image was filtered using BM3D. Then, images were cropped, leaving only the layers with local intensities that distinguished normal and DME features. Finally, the images were resized based on the terms of each pre-trained model. The model architecture is demonstrated in Fig. 48. 15 models of classifiers were trained such as Linear, Quadratic, Fine/Medium Gaussian, Support Vector Machines (SVM), cosine, and K-Nearest Neighbour classifiers (kNN). The accuracy was measured for each classifier to compare the achievement of the classification, and the highest accuracy was SVM with a ratio of 93.75%. The number of OCT volumes was ineffectively grouped (4 for DME and 79 for normal). The disadvantage of the

suggested method is that the weights are already pre-trained. Consequently, there is a restriction in enhancing the weights to improve classification performance. Fine-tuning should be introduced to optimize the weights leading to better improvement in performance.

Perdomo et al. [113] presented an end-to-end CNN model using an OCT dataset to extract features and to classify the two classes (DME and Normal). The model is demonstrated in Fig. 49. The model started by cropping and resizing the volumes. Using K-fold Cross-Validation (leave-one-patient-out), where one fold holds all volumes in the training set except one volume was set to test the model. The classification of a single volume was completed using a quorum rule, wherever 50% or more scans for a patient volume was classified as infected, the whole volume was identified as infected leading to a DME infected person. The model obtained accuracy, sensitivity, and specificity of 93.75%.

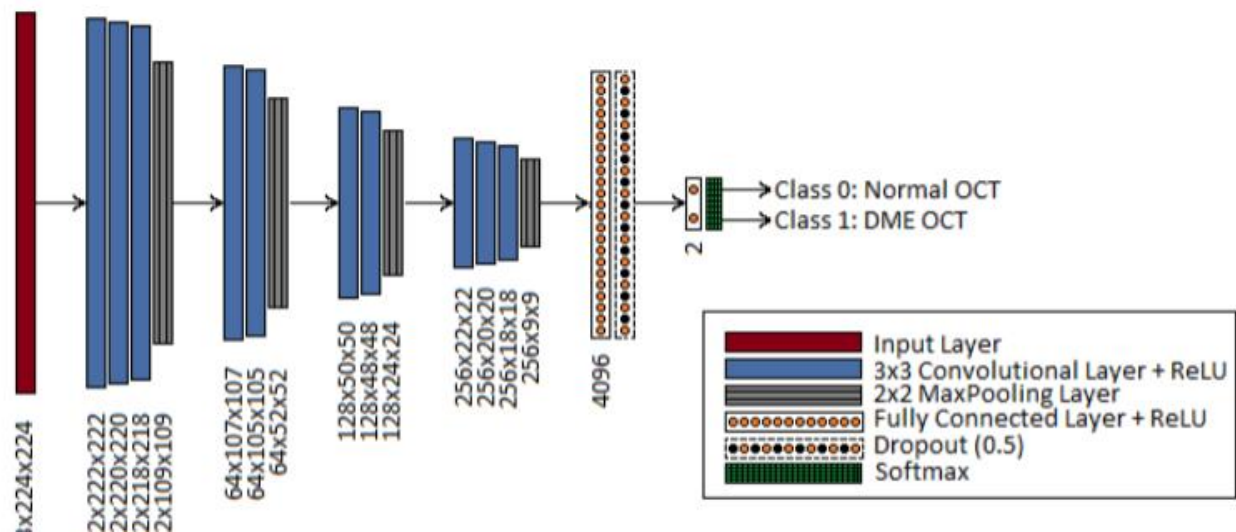


Fig. 49: Architecture of the CNN model proposed by [113]

Kamran et al. [114] proposed Optic-Net a novel convolutional neural network incorporated by two novel architectural to identify retinal diseases, this architecture proposed a new residual unit subsuming Atrous Separable Convolution, a novel building block and a mechanism to prevent gradient degradation. The results show an accuracy of 99.8% and 100% for two separately public available OCT datasets respectively. The first dataset contains 84484 images divided into 83484 train-set images and 1000 test-set images. The second dataset consists of 3231 image samples divided into 2916 train-set images and 315 test-set images. For both the dataset, 10-fold Cross-Validation was used to find the best models.

As presented earlier, classification approaches provide an effective method for retinal pathology identification compared to the segmentation approach taking into consideration the slight information needed from the retinal layers in feature extraction using CNN. Table 4 presents a summary of some classification methods.

Table 4: State of the art summary for classification methods using neural network

Ref#	Pathology	Dataset Size	Pre-processing	CNN Model	Classifier	Evaluate	Results
[95]	AMD, DME vs Norm	45	Filter/Flat	GoogleNet	SVM	Acc	96%
						Dice	0.89, 0.86, 0.99
[97]	DME, Drus, CNV	84484	-	IFCNN	SVM/RF/Bayesian	ACC	87.3+ 2.2%
[98]	AMD, DME	193	Norm/ROI/Flat/Crop	Multi-Scale CNN	Softmax	AUC	0.9985
						Precision	98.86%
[99]	AMD, DME, NVAMD	125978	Crop/Aug/Metadata	ImageNet	Softmax	ACC	86%
[100]	AMD, DME	32	Denois/Crop	AlexNet	SVM	SE, SP	93.75%, 100%
[101]	AMD, DME	45+ 1680	Denois/Filter Norm/Crop	Inception v3	Softmax	ACC	98%
[108]	DME	32	-	VGG-16	Softmax	SE, SP	100%, 81.25%
[110]	AMD, DME	207130	Flat	VGG-16	Softmax	ACC	98.6%
						SE, SP	97.8%, 99.4%
[111]	DME	83	Denois/Crop	AlexNet/VGG-16/GoogleNet	SVM/RF/KNN	ACC	93.74%
[113]	DME	32	Crop	Novel OCT-Net	KNN/RF	ACC	93.75%

3.3. Perspective and motivation

In this thesis, we decided to manage two retinal disorders (DR and DME) in first work using a new CNN model. Next, another work focused on DME using two approaches (CNN and GAN) due to their significant applicability in computer vision (CV); taking into consideration the researches rarity of synthesizing OCT images using GAN.

Moreover, the detection of diabetic pathology is a time-consuming task for human experts since it contains thousands of images. However, our perspective is to build an automated system to simply detect pathologies in real-time using only SD-OCT screening modality since it becomes deeply attractive for its significant performance for diagnosis despite its cost that is more expensive than the fundus photographing technique. Therefore, this thesis will present a simple computerized detection task to accomplish with training and learning.

There are many obstacles faced to implement such strategies as:

- OCT dimension: using a single channel vs. three channels scans. OCT is a grayscale 1-channel image where some deep pre-trained models have learned to detect features from colored images that depend on color intensity from extracted features.
- Brightness circumstances: In any detection operation, image brightness is an outstanding factor and it is a tricky problem to solve. Hence, the need for image preprocessing such as normalization and denoising is indispensable.
- Interloped dataset: the dataset used in consisted of crossed diseases where we can find several diseases in a single patient.
- Cross-sectional retina position: A fovea can be located in different positions in an image where it can sometimes be rotated or in curvature shape within retina's layers; in this case, the model must be capable to detect the features whatever its position.
- Image size: The size of OCT scans is different from one dataset to another or from one eye to another, therefore the responsibility of implementing an algorithm that unifies datasets without having consequences of its features.
- Components size: There is also the size of the components of the OCT features such as the drusen, the fluids, the leakage, or swallow varying from dataset scans to another which requires a vast awareness when performing the detection.

Concerning the feature detection and CNN applications, the challenges encountered are:

- Augmented dataset quality problem.
- Several liquid and swallow formations.
- The computational time-consuming neural networks.
- The hyper-parameters interpolation complexity.

Moreover, due to some limitations of related works shown in the preceding section, our goal in this thesis is to propose simple, reliable, and strong methods that deal with several OCT datasets represented by different institutes.

The related works reveal that many pre-trained CNN models are based on very deep CNN models, which are trained over colored images to recognize non-medical images. Thus, our model will apply a new simple CNN architecture to be trained and tested over mid-size OCT images in order to focus on extracting the most relevant needed features. Other related works aimed to use a features extraction model using a deep pre-trained CNN architecture to detect unrelated features from without processed the dataset rather than creating a new CNN structure that can utilize pre-processing raw data to detect related features.

Generally, most of these works classify the OCT images into three classes, while we intended to classify four interloped classes. Besides, they suppose that a huge dataset empowers the generalizability of the learned deep models; this assumption was not necessarily accurate, since the best accuracy did not reach 90% using the largest used dataset. Thus, we need to construct a model with data-driven filters on an adequate number of data to avoid over-fitting using an augmented dataset.

3.4. Conclusion

In this chapter, we have reviewed several retinal states-of-the-art methods. Firstly, we discussed classical state of the art using different feature extraction methods and different classifiers, and then we have reviewed the OCT segmentation researches using OCT for detecting pathologies. Finally, we have reviewed novel classification models based on CNN. Successively, we showed the motivations and the perspective of our work. In the following chapter, the proposed model will be introduced.

Chapter 4

Diabetic Retinal Tomographical Image Classification using Convolutional Neural Network

Summary

4.1. Introduction	82
4.2. Image processing materials and methods	83
4.2.1. Dataset acquisition:	83
4.2.2. Dataset Pre-processing method	84
4.2.2.1. Dimension reduction:	84
4.2.2.2. Image denoising	84
4.2.2.3. Image cropping	84
4.3. The proposed method.....	86
4.3.1. CNN architecture	86
4.3.2. Setting-Up Hyper parameters:	89
4.4. Experimental performance evaluation	90
4.5. Discussion	94
4.6. Conclusion.....	95

4.1. Introduction

Diabetic retinopathy (DR) and Diabetic Macular Edema (DME) are usually the popular reason for untreatable vision loss among adults with diabetes. However, early detection and treatment can reduce the risk of partial or total blindness, since the discovery of eye problems at the beginning will pull the alarm to prevent untoward effects of the developing diseases. An automatic method to detect diseases and oversteps the biased personnel evaluation of ophthalmologists is needed. Therefore, the problem of automatic classification for Spectral-Domain Optical Coherence Tomography (SD-OCT) images has been addressed in this chapter for the identification of patients with DR, DME versus normal subjects.

Nowadays, the examination of retinal pathologies is depending on the objective clinical analysis of OCT images by skilled ophthalmologists. To accelerate the

diagnostic process and enable the automatic classification of diseases, deep learning is an encouraging field of research that is applied over OCT for classification. Deep learning is a class of ML algorithms that extracts and organizes the feature representation of the data without the intervention of expert feature extractors.

Based on Convolutional Neural Network (CNN), a proposed convolutional model is presented and fully trained from scratch for the classification of OCT retinal imaging, which is widely used in ophthalmology for scanning the morphology of the retina for disease detection.

A new end-to-end CNN model will be presented in this chapter and has been fully trained using the Chinese University of Hong Kong (CUHK) dataset with 3 blocks of convolutional layers and max-pooling layers, which considered simpler to other end-to-end model and converge faster with less learned feature and higher accuracy. The last layers consisted of two fully connected layers of 256 units beside four units for the classification of DR, DMR, DR/DME, and normal cases. The classification has been achieved using a softmax classifier. The evaluation has been performed by randomly shuffling and separating the training data along with test data.

4.2. Image processing materials and methods

4.2.1. Dataset acquisition:

The dataset was acquired by the Chinese University of Hong Kong (CUHK) [115]. It consists of 189 SD-OCT volumes (4 DME, 67 DR, 39 DR/DME, and 79 normal cases). Each volume contains 128 B-scan with a resolution of 512 x 1024 pixels. All scans are interpreted and assessed by trained graders and identified based on the personal evaluation. Fig. 50 shows a demonstration of OCT images including DME, DR, DR/DME and normal cases.

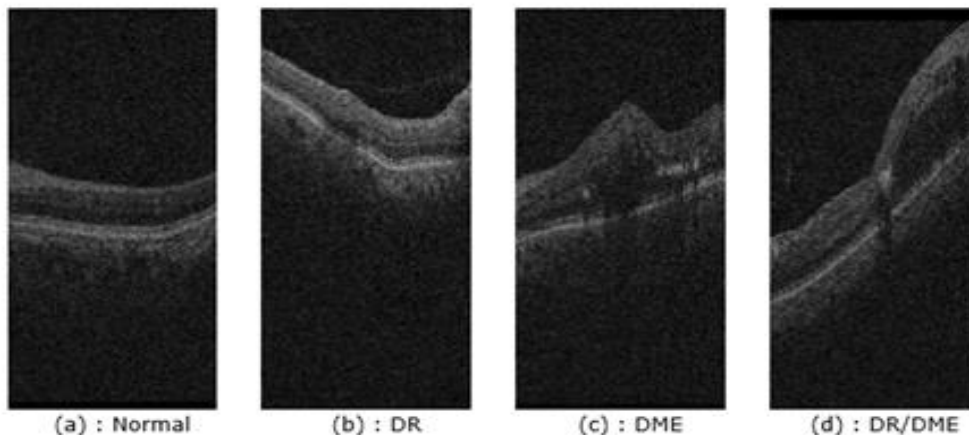


Fig. 50: An example of OCT images of diabetic pathologies compared to normal case

4.2.2. Dataset Pre-processing method

In this chapter, the model proposed requires preprocessing the original images regarding the retinal layers that may be turned, distorted, and moved vertically from the acquisition process. Thus, a manipulation process of the OCT volumes is needed. Furthermore, based on CNN computational requirement, uniform size of 256×256 dimensions for all images in the dataset is required. To enhance deep learning model result, a pre-processing step is applied where the OCT volumes have been routed in three stages: (a) normalizing and resizing, (b) denoising (c) and cropping.

4.2.2.1. Dimension reduction:

The images have been labeled and normalized via dividing all image pixel intensity by the max. After, the images have been resized from the original 512×1024 pixels into the half 256×512 .

4.2.2.2. Image denoising

OCT images endure scattered noises similar to other image patterns like Ultrasound [116]. The OCT images are improved by reducing these noises, where each slice of B-scan has been denoised using the NLM (Non-Local Means) [117]. Fig. 51 shows an example of a slice denoising for a normal patient with an enhancement in results. NLM has been strongly applied to ultrasonic images to decrease noise [118]. The filtering mechanism of NLM maintains good structures compared to other methods such as BM3D [119]. Denoised images can accelerate the model since it is end-to-end learning.

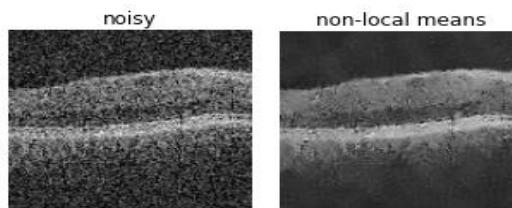


Fig. 51: Example of NLM image denoising for a normal patient

4.2.2.3. Image cropping

According to CNN computational prerequisite, homogeneous size of 256×256 dimensions has to be reached, and since the images ensuing from denoised step has a dimension of 256×512 . Thus, an elimination process of 256 pixels from each B-scan rows was applied to reach 256×256 pixels. The method consists on analyzing all images and exterminates from each image the rows containing unwanted black pixels, aiming to focus on the retina structure region. The cropping algorithms automatically computes, for every row in every image, the count of black

pixels and remove the rows with the highest count starting from top and/or bottom until achieving the target as shown in Fig. 52 (Group A); where the red rectangles are the area to eliminate with the count of 256 rows, which let us focus on wanted area with 256×256 pixels for all datasets as input for the Convnet as seen in Fig. 52 (Group B).

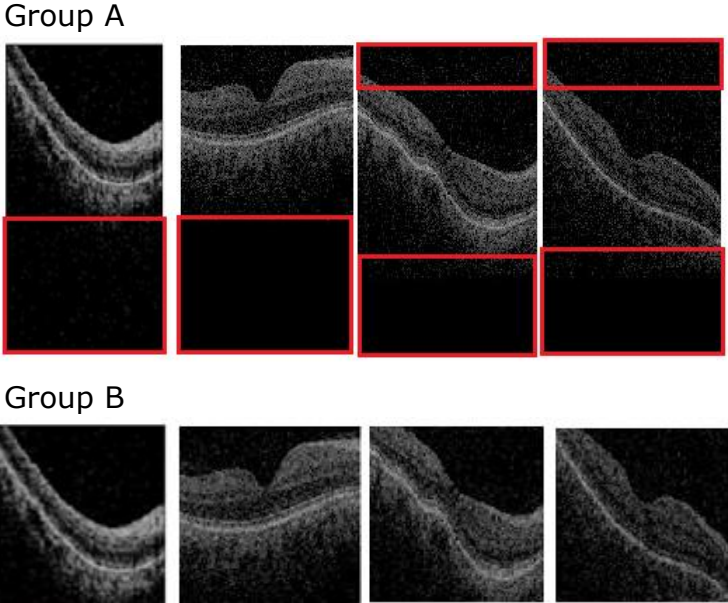


Fig. 52: Cropping example where red rectangles indicates the area to eliminate 256 rows each (Group A) leading to a 256×256 pixels images (Group B).

Since the acquired scans may contain tilted and/or distorted images resulting from the acquisition procedure that leads to a curved retina. The cropping algorithm mentioned earlier can simply remove tiny informative data to adjust the resizing process. This missing data did not have a huge effect on the result since the region of interest (as swallow and leaks) in the retina exists in the middle of the image and it will not be compromised.

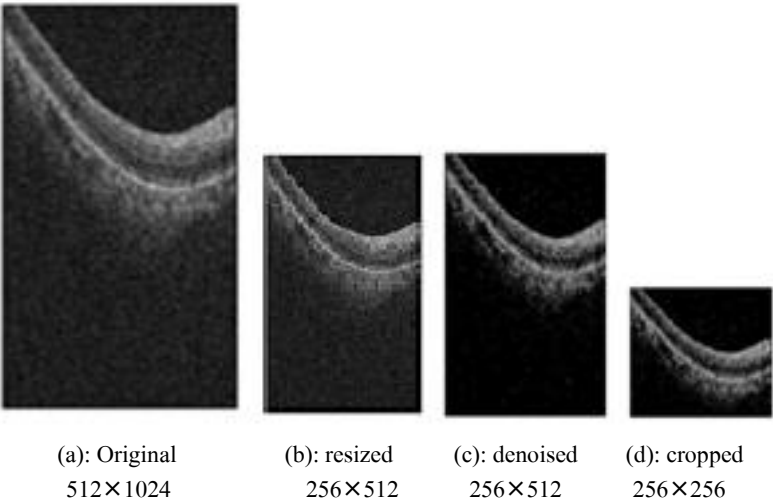


Fig. 53: From original to cropped image, Example of OCT image.

It is worth to notice that the utilization of an adequate flattening strategy is kept to use in the next chapter in order to align the layers and correcting the retina curvature to concentrate on RPE. Fig. 53 shows example of images for each step of data processing.

4.3. The proposed method

4.3.1. CNN architecture

The proposed model is discussed in this part where the study will include used datasets in the system setup after image pre-processing (including denoising and resizing). Fig. 54 presents a flow chart description of the proposed model.

CNN holds several hidden layers that are operating convolution and sub-sampling techniques to extract different levels of features of the input data for single and double dimensions. CNN showed a powerful performance in many fields, especially in the medical domain and computer image processing. Fig. 55 presents the steps to develop a convolutional architecture for SD-OCT classification.

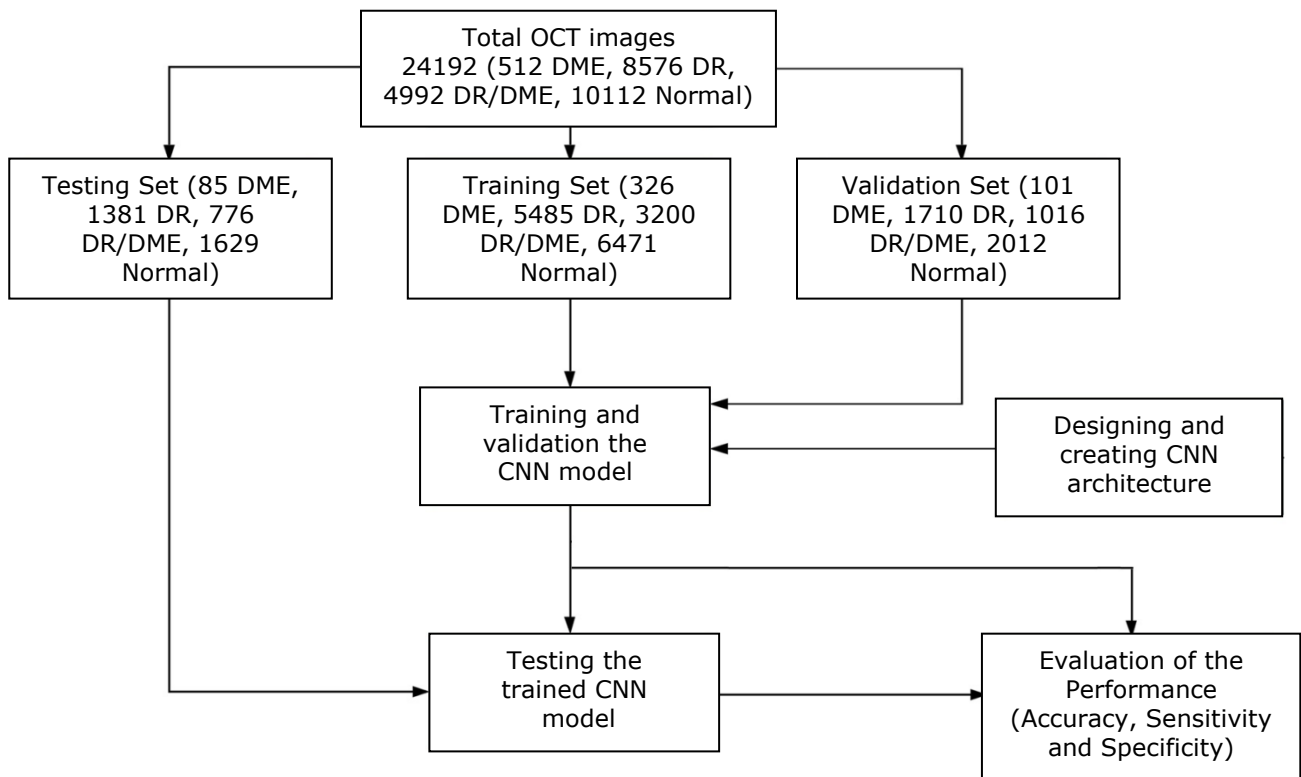


Fig. 54: Flow chart diagram

Our proposed CNN model contains five arranged and ordered types of layers (input, multi-layers convolution with RELU, several fully-connected layers, classification, and output); The proposed model is used to classify the input images into 4 classes (DR, DME, DR/DME, and Normal) and it consists of 16 layers as shown in Fig. 56.

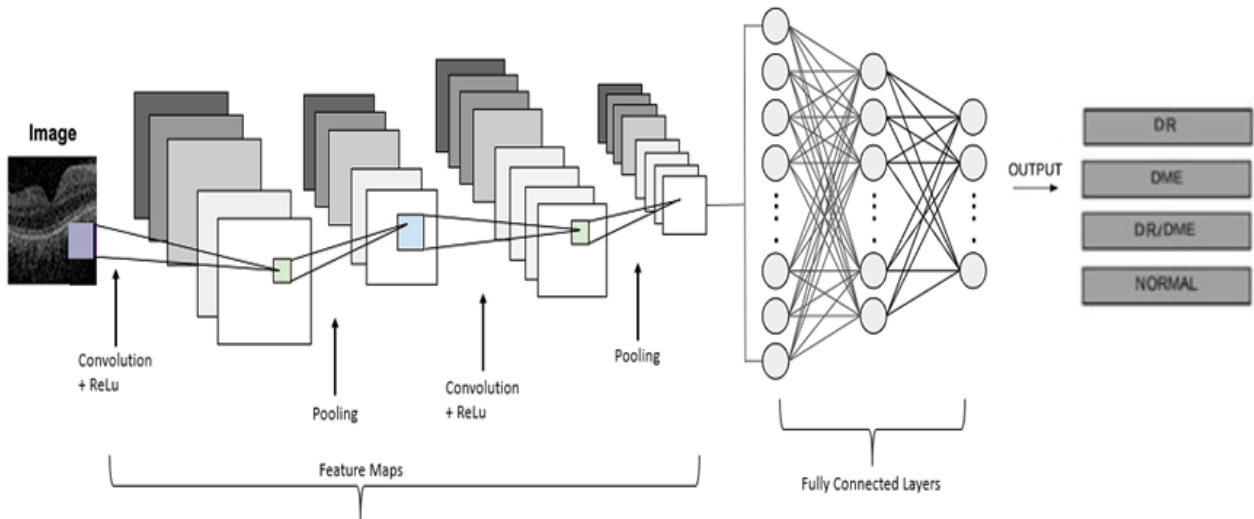


Fig. 55: Steps to create a deep learning-based CNN model for DR diagnosis using SD-OCT

This proposed architecture is composed of five convolutional layers with 3×3 kernel's size accompanied by Relu activation, three max-pooling with 2×2 kernel's size without zero-padding, and an incremental number of filters from 16 to 256 to learn and extract different feature representations. The input to the new model is fixed to 256x256 grayscale images for the CUHK dataset, and the stride is set to one. Each maxpooling is set to each group of convet layers with a single stride.

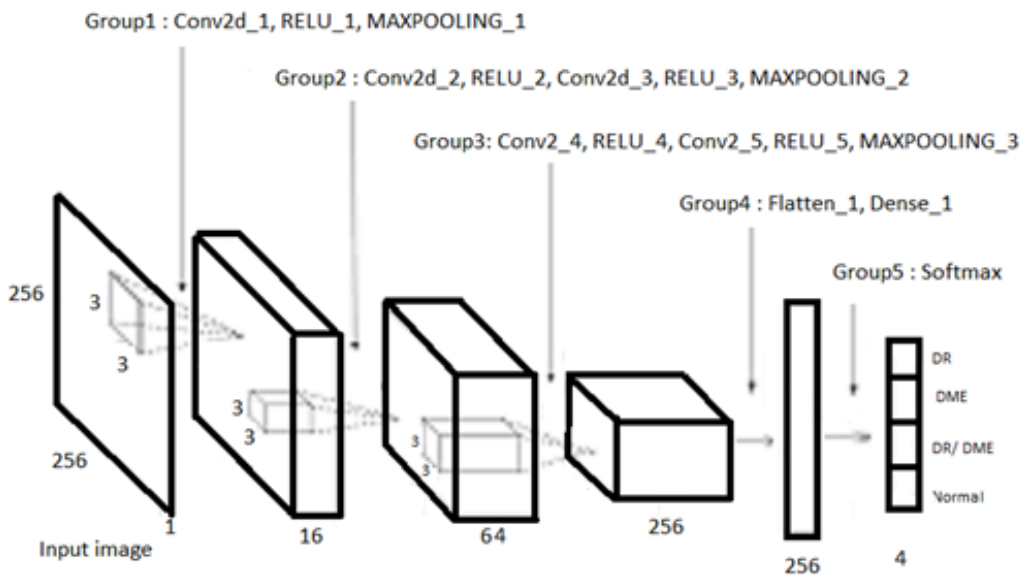


Fig. 56: Architecture of CNN model depth for CUHK dataset: the convnet contains 16 layers. The output is classified over 4 class labels.

The stack of Convolutional layers is followed by two Fully-Connected (FC) layers: the first has 256 channels; the second performs 4-way classifications. Moreover, between FC layers, a dropout [120] layer has been set to 0.2. To finish, the Softmax classifier layer has been utilized as the model prediction. A presentation of the model and its parameters are shown as the diagram in Fig. 57.

Table 5 presents detailed parameter values about layers in the proposed CNN model.

Table 5: Layers' values for the proposed architecture

#	Layer	Values
1	Input layer	256×256×1
2	Convnet 1	Number of filters : 16
		Kernel Size: 3×3
		Activation: RELU
3	Batch norm 1	Number of channels: 16
4	Maxpooling 1	Kernel size: 2×2
		Stride : 1×1
5	Convnet 2	Number of filters : 32
		Kernel Size: 3×3
		Activation: RELU
6	Convnet 3	Number of filters : 64
		Kernel Size: 3×3
		Activation: RELU
7	Batch norm 2	Number of channels: 64
8	Maxpooling 2	Kernel size: 2×2
		Stride : 1×1
9	Convnet 4	Number of filters : 128
		Kernel Size: 3×3
		Activation: RELU
10	Convnet 5	Number of filters : 128
		Kernel Size: 3×3
		Activation: RELU
11	Batch norm 3	Number of channels: 128

This CNN model reduces the number of the layers compared to other pre-trained learning methods such as transfer learning technique, which has been reduced from 25 layers in AlexNet, 48 layers in Inception-V3, 101 layers in ResNet-101, etc. to only 16 layers; these pre-trained networks are usually not trained using medical data. Decreasing the number of layers will reduce the system resources and time needed for training and detecting new images class.



Fig. 57: Architecture of the CNN model. The input layer accepts B-scans with 256x256 pixels from CUHK dataset. The parameters (feature maps and size) are reported beneath every layer.

4.3.2. Setting-Up Hyper-Parameters:

The model was trained from scratch over CUHK dataset. CNNs are difficult to configure since different parameters have been required to be handled. Grid search [121] has been used as a method to choose the best hyper-parameters values for the model.

Different batch sizes as well as different epoch numbers were tested where results showed that batch size of 20 and 100 epochs as maximum size achieved the best result using 'Categorical Cross-Entropy' as a loss function. Different solvers were tested also between a suite of optimization algorithms (SGD [122], 'RMSprop [123]', 'Adagrad [124]', 'Adadelta [125]', 'Adam [126]', 'Adamax [127]') and the results suggested 'Adam' to be the best optimizer.

"Uniform" weight initialization achieved the best result comparing to other network weight initialization techniques [128] ('uniform', 'lecun_uniform', 'zero', 'normal', 'glorot_uniform', 'glorot_normal', 'he_uniform'). Different activation functions have been evaluated from a suite of functions [129] ('softmax', 'softplus', 'softsign', 'relu', 'tanh', 'sigmoid', 'hard_sigmoid', 'linear') and the 'relu' activation function achieved the best results. Finally, the dropout rate of 0.2 reached the best accuracy. Furthermore, the subsequent hyper-parameters provide the most reliable results; initial learning rate = 0.001, L2 regularization is $1.000e^{-05}$.

4.4. Experimental performance evaluation

To assess the performance of the proposed model in classifying the diabetic eye retina pathologies, a comparison between the classification outputs with the ground-truth label of the image was applied. These images were labeled manually by specialists to evaluate classification results. The efficiency of the model was computed by estimating the accuracy, sensitivity, and specificity, where these values determine how accurately the retina is diagnosed. To calculate these values, four statistical rates (TP, FP, TN, FN) should be computed first using the confusion matrix (as described in chapter 3, section 3.3.2.5.3).

In addition, the ROC curve was presented to plot the performance of the proposed architecture (as described in chapter 2, section 2.3.2.5.4). The best performance result of ROC curve is when the curve will be closer to the left upper corner. Moreover, the AUC of ROC was also computed where the higher value leads to accurate prediction.

Furthermore, to evaluate the performance of the proposed model, a 'Categorical Cross-Entropy' loss function is compiled in the last layer of the network. This evaluation method tells how well the proposed method models the provided dataset. If predictions differ too much from actual results, the loss function would come up with a very large number. Progressively, with the help of the effective 'Adam' optimizer, the loss function learns to diminish the error in prediction. To minimize the errors, the current error is propagated back to the previous layer to modify the weights and bias.

The CUHK dataset contains images of 4 classes (DME, DR, DR/DME, Normal) and has been resized, denoised and cropped into 256 x 256 pixels resolution as input for to feed the CNN model. The dataset has been divided into training (64%), validation (20%) and testing (16%) set. A total of 189 patients composed of 128 OCT scans each formulated the dataset, corresponding to 24192 samples separated into 15482 for training, 4839 for validation and 3871 for testing. Table 6 shows detailed information about the used images.

Table 6: The distribution of images used in the system

Case	Number of training images	Number of validation images	Number of testing images	Total number of images
DME	326	101	85	512
DR	5485	1710	1381	8576
DR/DME	3200	1016	776	4992
Normal	6471	2012	1629	10112

A workstation with Intel Xeon E5-2670 2.3 GHz, and NVIDIA Tesla K40c GPU [130] has been used for training and testing the proposed model. Early stop [131] technique was used in the training phase to monitor the evaluation of loss function; the training will stop when the quantity monitored has stopped decreasing. Early stopping and dropout have been applied to reduce over-fitting.

We can observe that training and validation are satisfactorily performed with an elevated accuracy percentage, where the result of the CNN model after 100 epochs showed a training accuracy rate of 99.58% using CUHK OCT dataset with a loss value about 0.021 as shown in Fig. 58; and a validation accuracy of 98.63%; besides a sensitivity of 99.59% and 98.64% for training and validation respectively, a specificity of 99.91% and 99.75% for training and validation respectively as presented in Table 7.

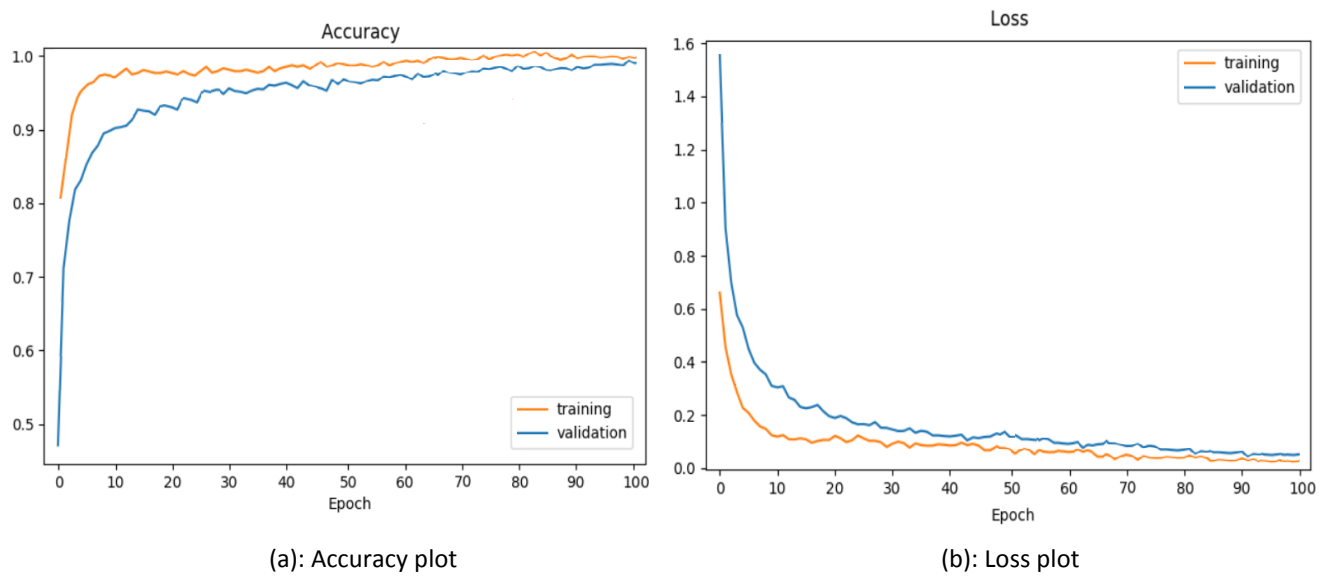


Fig. 58: Accuracy and Loss rate

Table 8 and Table 9 present the confusion matrix for training and validation respectively. The training and the validation loss have been reduced by time and the test loss ends slightly over the training loss. The curve ends in considerably noteworthy without over-fitting

Table 7: Evaluation results of the testing dataset

	Training	Validation	Testing
Accuracy	99.58%	98.63%	99.02%
Sensitivity	99.59%	98.64%	99.03%
Specificity	99.91%	99.75%	99.79%

Table 8: Performance result of proposed model during training

Confusion Matrix		Predicted Class				
Actual Class	Class	DME	DR	DR/DME	Normal	Total
	DME	326 100%	0	0	0	326
	DR	0	5453 99.42%	10	22	5485
	DR/DME	0	3	3190 99.67%	7	3200
	Normal	0	36	14	6421 99.23%	6471
	Total	326	5492	3214	6450	15482

To evaluate the proposed model in detecting the classes of retinal eye pathologies, this system was tested using the testing unseen part from the dataset, where the model achieved an accuracy of 99.02%, a sensitivity of 99.03%, and a specificity of 99.79% as seen in Table 7. Using the test set, a confusion matrix is presented in Table 10.

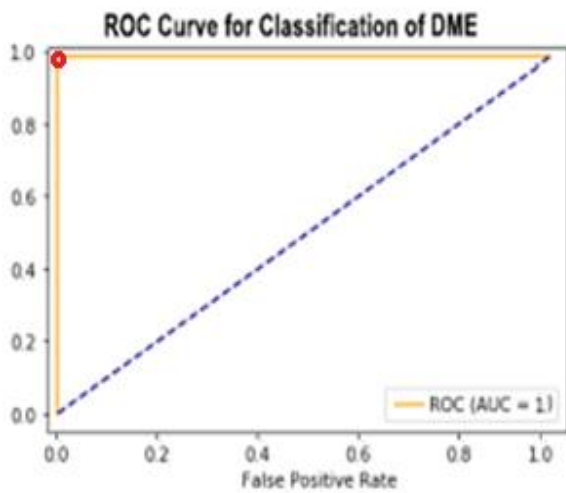
Table 9 : Performance result of proposed model during validation

Confusion Matrix		Predicted Class				
Actual Class	Class	DME	DR	DR/DME	Normal	Total
	DME	100 99.01%	0	1	0	101
	DR	0	1686 98.57%	8	16	1710
	DR/DME	0	2	1003 98.72%	11	1016
	Normal	0	26	10	1976 98.22%	2012
	Total	100	1714	1022	2003	4839

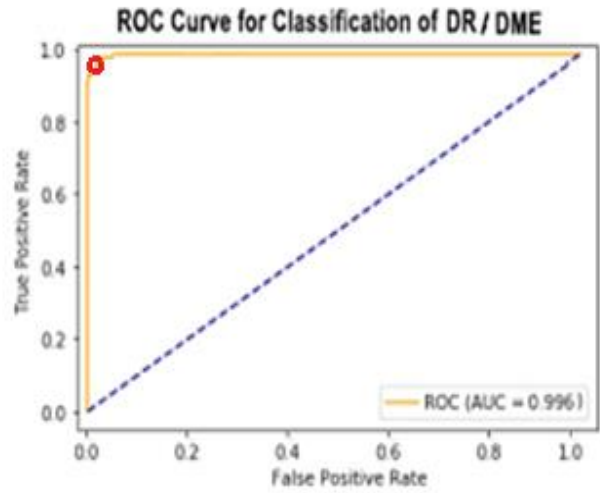
The ROC curves for each class are plotted in Fig. 59 (a) DME ROC, (b) DR/DME ROC, (c) DR ROC, (d) Normal ROC. The AUCs of the various classes were above 0.98 and have the following values 1, 0.996, 0.992, and 0.984 for DME, DR/DME, DR, and Normal respectively. Table 11 demonstrates the evaluation of the performance results of testing set showing the ratios of the recall, precision, and F1-score for all classes whereas noticing that all ratios bypassed 98%.

Table 10: Confusion matrix showing the testing rate

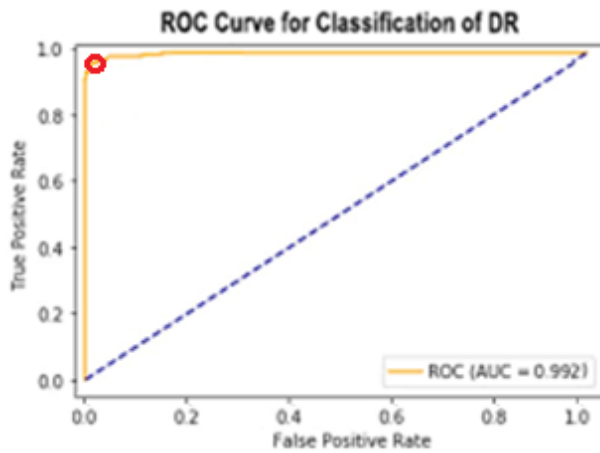
Confusion Matrix		Predicted Class				
Actual Class	Class	DME	DR	DR/DME	Normal	Total
	DME	85 100%	0	0	0	85
	DR	0	1362 98.62%	5	14	1381
	DR/DME	0	2	769 99.1%	5	776
	Normal	0	19	7	1603 98.39%	1629
	Total	85	1383	781	1622	3871



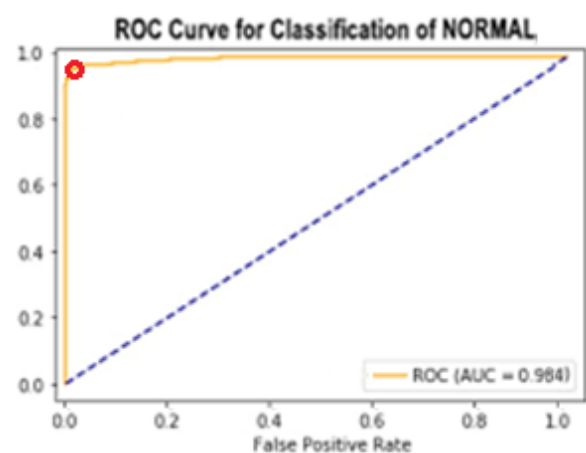
(a)



(b)



(c)



(d)

Fig. 59: Curve plot showing the performance of proposed model during testing

Table 11: Performance ratios of testing set

	Recall	Precision	F1-Score	Support
DME	100%	100%	100%	85
DR	98.62%	98.48%	98.55%	1381
DR/DME	99.10%	98.46%	98.78%	776
Normal	98.40%	98.83%	98.62%	1629
AVERAGE	99.03%	98.94%	98.99%	3871

4.5. Discussion

When comparing the results of the proposed model with some other related works based on CNN architectures (mentioned in the state of art), we can note that our work focused on the main pathologies in the retinal eye diabetic diseases, using crossed data between the DR and DME patients and forming together a new class for patient having both diseases that helps to extract new feature.

The majority of related works use different OCT databases or an aggregation of several datasets to form a large one, and it is noted that they use diverse classes, usually more than three classes that can affect the performance of the applied classification methods. However, [95] proposed a model based on transfer learning techniques using CNN from GoogleNet architecture. This method used 10,000 images to classify three classes with an accuracy of 96%. [97] used an iterative fusion CNN for classifying four classes of OCT images using 84,484 resulting in an accuracy of 87%. [86] used SVM with HOG features from 3247 images to classify three classes with an accuracy of 95.5%. [99] proposed a transfer learning method based on ResNet-V2 CNN architecture to classify four classes from 113,397 OCT images leading to 86% accuracy. [101] used also a transfer learning CNN method based on Inception-V3 to classify three classes from 1680 images with an accuracy of 98%. [113] presented an end-to-end CNN model using 4096 OCT images to classify two classes with an accuracy of 93%. [102] used a transfer learning CNN method based on (InceptionV3, VGG16, and ResNet50) to classify four classes from 35,900 images with an accuracy of 90%. [103] also used a transfer learning CNN method based Dense-Net and ResNet50 to classify four classes from 84,484 images with an accuracy of 88%. The proposed work presented an end-to-end model to classify four classes (three types of volumes) from 24192 images with an accuracy of 99.3%.

From the model listed above, it is remarkable that many pre-trained CNN models are based on very deep CNN models such as GoogleNet and VGG networks, which are trained over three channel images to recognize non-medical images. It is worth noting that some of the related works aimed to use pre-processing methods and implemented transfer learning models with pre-trained CNN rather than creating a new CNN structure that can utilize raw data without pre-processing. In addition, some works utilized ML techniques with extracted features from the segmented ROI of OCT as a classification method. However, most of these methods have achieved high detection rates, more than 87%.

Finally, most of these works classify the OCT images into three classes, while a few models classify four classes. Moreover, a general supposition in ML and DL is that a huge number of data is needed to improve the generalizability of the learned model; this assumption compared to the-states-of-art methods mentioned above was not necessarily accurate, since the biggest dataset used was 113,397 images but the accuracy result is almost insufficient and did not reach 90%. Meanwhile, the proposed work applied a model based on the middle size dataset of OCT images that have never been used in such class types with 24,192 images to classify DR and DME. The construction of a higher-order model with data-driven filters on an inadequate number of data drives to over-fitting, which had a bad impact on testing performance.

4.6. Conclusion

In this chapter, a representation of a simple new end-to-end-CNN classification model has been proposed with the ability to detect patterns for diabetic macular edema using OCT images. Contrary to manual retinal diseases diagnosis based on personal clinical examination, this method showed the capability to automatically predict the most two diabetic retinal diseases.

This CNN model is formed for the analysis of retinal diseases using raw images and it was trained and tested using images from a university of Hong Kong containing four different volumes (two diseases, normal cases, and a mixed volume of these 2 diseases). Besides, the proposed model was trained and tested using a relatively big dataset of 24192 images. Furthermore, by tuning the hyper-parameters of the proposed model CNN structure, the more accuracy and lower time cost were achieved.

The result showed that the proposed method has been highly effective for predicting image with diseases where the experimental results show that the proposed method achieves a performance and an accuracy test rate of 99.02%. The results, using split validation, showed improvement with accuracy and sensitivity, comparing to other CNN similar end-to-end or some pre-trained model, which is exceptionally

promising in the field of early detection of diabetic diseases such as DME and DR to aid ophthalmologists in biomedical technologies.

The proposed method showed an advantage with classifying not only DME patients (as reported by most CNN-OCT paper focusing on DME), but also patients with DR and/or DME, with high efficacy to detect both diseases in patients as recall showed a percentage above 98% (see Table 11).

Finally, the phase of pre-processing the raw data showed a better result, therefore, in the next chapter, the work would focus on using flattened and augmented images that produce a larger dataset leading to reduce the error rate while focusing on the most risky diabetic disease.

Chapter 5

Convolutional Neural Network Classification Model for Diabetic Macular Edema using Spectral Domain Tomographical Scans

Summary

5.1. Introduction	97
5.2. Image processing methods	98
5.2.1. Dataset acquisition:	98
5.2.2. Dataset Pre-processing technique:.....	99
5.2.3. Synthetic data augmentation:	101
5.2.3.1. Generator:	104
5.2.3.2. Discriminator:	105
5.2.3.3. Adversarial Model:	106
5.2.3.4. GAN architecture:	106
5.2.3.5. Synthetic image evaluation	107
5.3. The proposed method.....	110
5.3.1. Fine-tune the pre-trained model.....	110
5.3.2. Re-train the model	110
5.4. Performance Analysis	113
5.5. Discussion	117
5.6. Conclusion.....	119

5.1. Introduction

The macula is the most essential part of retina, providing for each person the capability to see and adjusts the sharp vision by the observation of detailed scenes. Therefore, any leakage within the macula, caused by the accumulation of fluid or swelling, will lead to diabetic macular edema (DME). Progressively, it may destroy the blood vessels, leading to complete blindness. People with DR are commonly exposed to DME. Therefore, prior to any obvious signs of vision problems resulting from the development of the retinal diseases for diabetic people, it is necessary to detect any

cysts in the retina as soon as possible using an automatic diagnosis system by inspecting any layer deformations.

In present, the detection of retinal diseases is based on the manual analysis of OCT B-scans through skillful doctors. Thus, it is urgent to build a system capable to defeat human skills by expediting the process of detection using an automated and computerized retinal pathologies classification. Computer vision methods and especially deep learning algorithms are the most powerful tool of digital image processing and recognition. DL introduces a new era in the research fields of medical image classification using SD-OCT scans. Based on the pre-trained CNN architecture presented earlier in chapter 4 using OCT scans, a fine-tuned CNN model has been proposed for identifying DME given retinal SD-OCT images, which aimed, with an augmented training data, to automatically recognize DME as it considers the major cause of irreversible blindness.

SERI dataset is used in this chapter to feed the fine-tuned model that is relatively similar to CUHK dataset presented in the previous chapter. In addition, the experiments have been tested over two datasets provided by different institutions.

5.2. Image processing methods

5.2.1. Dataset acquisition:

The experiments have been accomplished using three datasets, provided by three different institutions. The first "SERI" has been used to fine-tuned the pre-trained model proposed in chapter 4 while the two other datasets (DUKE and NOOR) are used to test the model prediction result.

5.2.1.1. SERI Dataset

The dataset was acquired by the Singapore Eye Research Institute (SERI), using CIRRUS TM (Carl Zeiss Meditec, Inc, Dublin, CA) SD-OCT device [132].The dataset includes 32 SD-OCT volumes (16 DME and 16 normal cases). Each volume contains 128 B-scan with a resolution of 512×1024 pixels. All volumes have been read and assessed by trained graders, and identified as normal or DME cases [133] based on evaluation of retinal thickening, hard exudates, intra-retinal cystoids space formation, and sub-retinal fluid.

5.2.1.2. DUKE Dataset

The dataset was acquired utilizing Spectralis SD-OCT (Heidelberg Engineering Inc. in Germany) scans at Duke University, Harvard University, and the University

of Michigan. The dataset includes 45 SD-OCT volumes [134]: (15 normal, 15 AMD (excluded from the study in this chapter), and 15 DME patients. All volumes contain several B-scans that vary from 31 to 97 with a resolution of 496x768 pixels.

5.2.1.3. Noor Eye Dataset

This dataset was acquired by the Hospital of Noor Eye in Tehran [135]. It consists of 148 SD-OCT volumes (50 normal, 48 dry AMD "excluded from this study", and 50 DME). All volumes contain a number of B-scans that vary from 20 to 61 with resolution of 496x512 pixels.

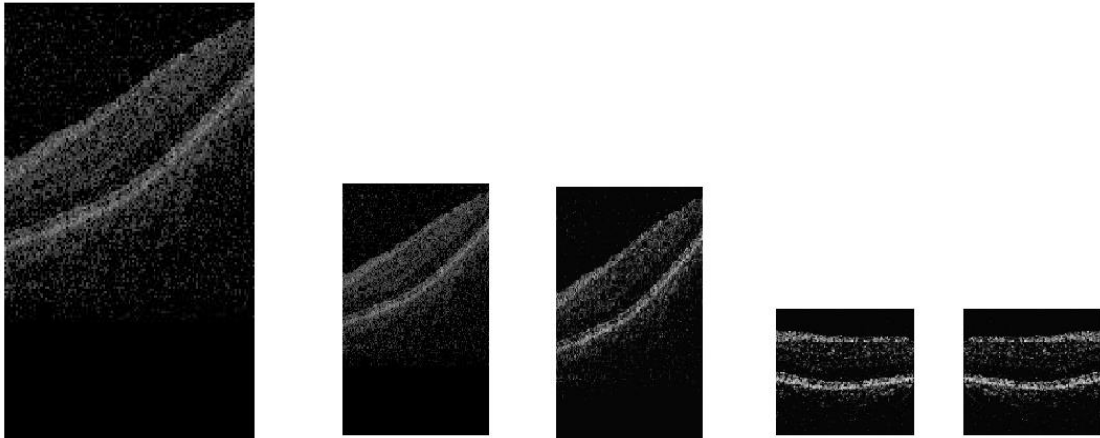
5.2.2. Dataset Pre-processing technique:

Since the proposed model in this chapter is based on the trained model presented previously in chapter 4, therefore, the steps of data manipulation and preparation are necessary to fit the requirement of the model. The acquired images are captured in different eye position, thus, the raw images showing the retinal layers may be tilted, misrepresented, and/or vertically slid composed from the acquisition. The pre-processing phase is also applied over the OCT scans and managed in four steps: (a) normalizing and resizing, (b) denoising, (c) cropping, (d) flattening (e) and mirrored where Fig. 60 shows an example of both classes.

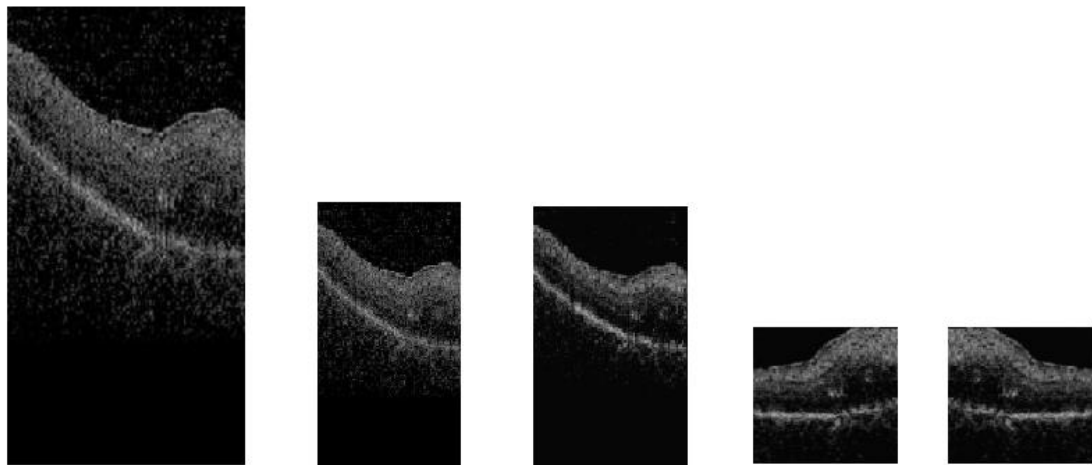
The labeled images have been normalized and resized from the original 512x1024 pixels into 256x512. Noisy OCT images are enhanced by denoising each slice of B-scan using the NLM (Non-Local Means). NLM preserves better structures compared to other methods such as BM3D as presented previously. Consequently, denoised images can speed up the fine-tuned model. The following process was to flatten the retinal curve shape to diminish the variations between images within patients. OCT images have a curvature caused by the acquisition of captured images that alters between subjects and within each volume.

To flatten the image curvature, first, a detection of the retinal pigment epithelium layer (RPE) was applied since RPE is considered as the most reflective layer. The outermost of the two highest local maxima in each column of the denoised image has been assigned as the estimated RPE location [87]. Next, the convex hull around the RPE has been calculated and the lower border of the convex hull as an estimate of the lower boundary of the retina has been used. A median filter has been used to remove outliers. To create the flattened image, a second-order polynomial has been fitted to the estimated retinal lower boundary points and shifted each column up or down so that these points lie on a horizontal line.

Group A: Normal OCT



Group B: DME OCT



(a) Original
512×1024

(b) resized
256×512

(c) denoised
256×512

(d) flattened & cropped
256×256

(e): mirrored
256×256

Fig. 60: Example of an OCT image pre-processing phase, from original to flattened, cropped and mirrored image for a healthy person (Group A) and a diabetic patient with DME (Group B).

Moreover, based on CNN requirement, SERI is the only dataset used to be trained and augmented, and it must be standardized and reached the dimension size of 256×256 as well as other datasets presented later. Thus, an exclusion process of 256 pixels from each B-scan rows was applied to reach 256×256 pixels using the same technique presented in previous chapter where the method consists of eliminating from each image the rows outer the region of interest. Fig. 61 represents a simulation and an example of flattening and cropping algorithm.

Finally, since all images have been flattened and cropped leading to a standard shape for all the dataset that focus on the ROI, a mirrored technique to augment the data has been implemented over SERI images. These mirroring techniques are more powerful than traditional methods such as rotating where the same

transformation formatting is preserved leading to double the images in order to enlarge them using GAN network as presented later.

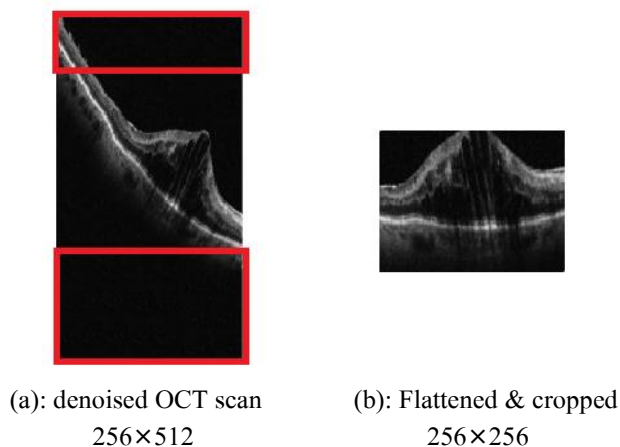


Fig. 61: (a) Cropping simulation where red rectangles indicate the area to eliminate with a size of 256×256 pixels resulting a flattened and cropped image with a size of 256×256 (b).

5.2.3. Synthetic data augmentation:

In deep learning models, the most significant problem in training a network is the absence of large training datasets that are obligated to be considered enough and representative to accurately train the model and prevent over-fitting. In this situation, the generation of synthetic data can be valuable for over-sampling the classes and create additional datasets.

In many cases, it is common for databases to have minority classes that are under-represented especially when examining medical data, where the count of healthy persons is considerably higher than diseased ones. When this occurs, the classification may have struggles recognizing these classes. To overcome this problem, it is necessary to augment the minority data by creating new input by modifying the raw images in meaningful patterns. This strategy resolves the problem of under-representation of minority classes by enhancing this representation, which helps to prevent over-fitting.

To increase the training set and enhance the classification results in the DME classification task, we firstly doubled the SERI dataset by mirroring all the images after the pre-processing steps, and then we augmented the dataset by synthesizing new samples using the GAN model. Contrary to classic augmentation methods that include a diversity of image manipulations, GAN showed a promising advance in image augmentation results since it was first introduced by Goodfellow et al. [136] in 2014. GAN is a type of generative model that can produce new content based on its training data. We followed a popular GAN model, the Deep Convolutional GAN

(DCGAN), where we are inspired by the architecture presented by Radford et al. [137] with some tweaking in parameter values presented later.

DCGAN attacks the problem of unsupervised learning by training two deep networks, called Generator and Discriminator, that compete and cooperate with each other. The discriminator is trained to minimize the last classification error between real and generated images, while the generator is trained to maximize it. Thereby, the idea of adversarial networks derives from it. The balance is attained when the generator network provides samples that match the probability distribution and the discriminator foresees as fake or real with a 50% probability each.

While training, it is necessary that both networks learn evenly and converge concurrently. Otherwise, the generative network can be stuck if the discriminative network learns better to identify fake images.

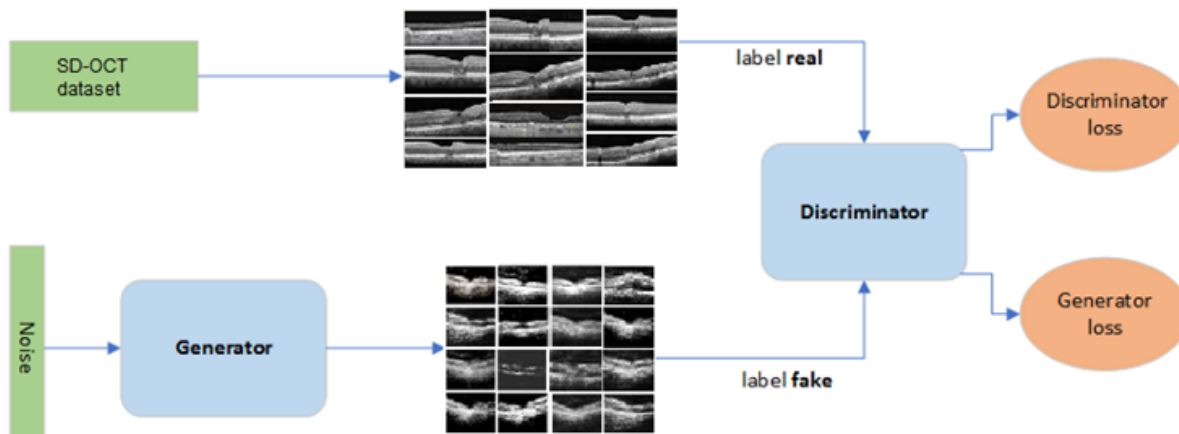


Fig. 62: A representation of GAN model

When training the discriminator, we neglect the generator loss and only utilize the loss of discriminator. Therefore, the weights of the discriminator network are modified by using back-propagation; the weights of the generator are not modified. Whilst, when training the generator, we utilize the loss of the generator; the discriminator network is suspended through generator training and exclusively the weights of the generator network are modified using back-propagation. Fig. 62 shows an overview of the DCGAN model.

The subsequent steps are performed back and forward letting the DCGAN model to deal with contrary generative difficulties:

- a) Pick a set of real OCT images from the training set.

- b) Generate many bogus images. This is done by examining noise vectors randomly and generating images using the generator.
- c) Train the discriminator for many epochs using both unreal and real images. This will modify just the discriminator's weights by identifying all the real images as one and the fake images as zero.
- d) Produce different numbers of fraudulent images.
- e) Train the entire DCGAN model for many epochs applying just unreal images. This will modify only the generator's weights by identifying all false images as 1. Fig. 63 shows the training plan of the discriminator where the weights of generators are frozen which lets the weights of discriminators to be modified only using back-propagation for real and fake images; while Fig. 64 presents the training plan of the generator where the discriminator weights are constant, letting only the weights of the generator be updated through back-propagation

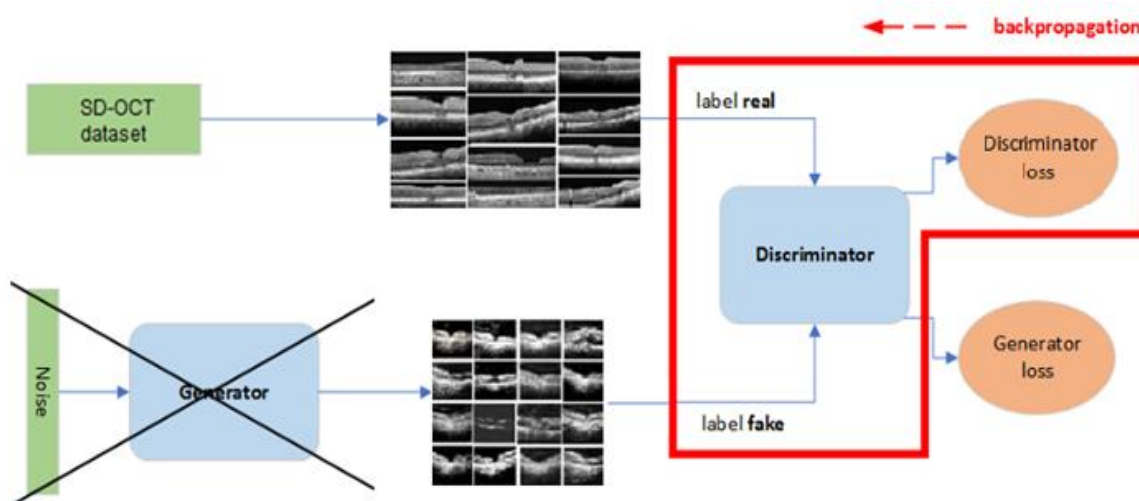


Fig. 63: Discriminator training presentation where the discriminator learns to identify fake images.

Various works on GAN extensions such as DCGANs, Cycle GANs and Progressively-Growing GANs [138] were published in 2015, 2017, and 2017, respectively. Applying Data Augmentation has become increasingly popular with works such as Neural Augmentation [139], Smart Augmentation [140], and Auto-Augment [141]. Data Augmentation GAN for image synthesis was used by Frid-Adar et al. [142] for liver lesion classification. This improved the classification performance from 78.6% of sensitivity and 88.4% of specificity using classic augmentations to 85.7% of sensitivity and of 92.4% specificity.

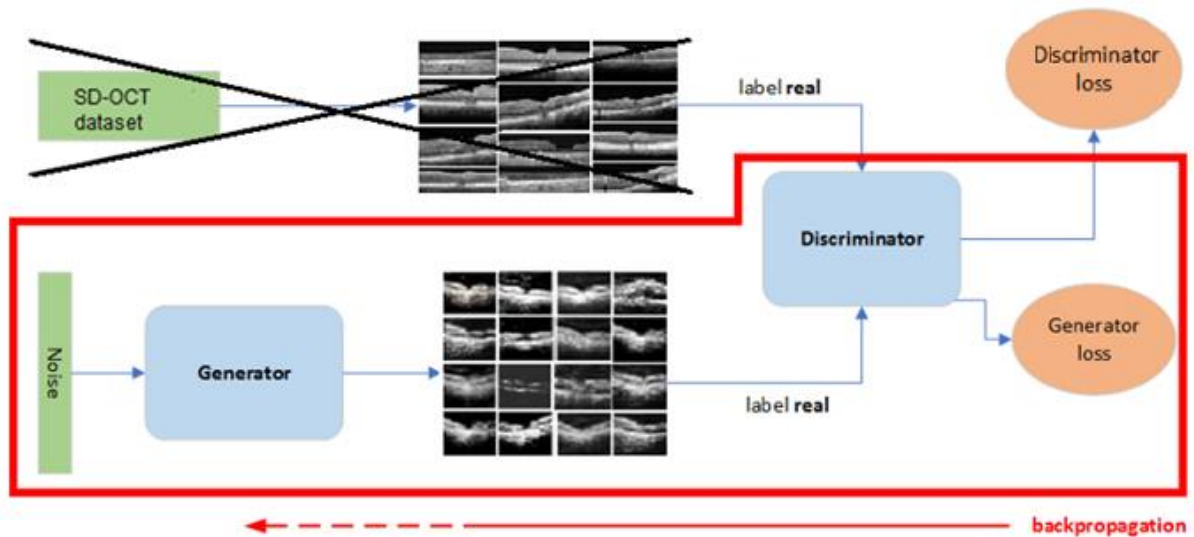


Fig. 64: Generator training presentation where the generator learns to fool the discriminator.

5.2.3.1. Generator:

The generator accepts a noise vector of 300 numbers of the latent space and generates an image with the same size as the discriminator's input size ($256 \times 256 \times 1$). The generator learns to predict a sample. The generator is estimated by the discriminator, indicating that its purpose is to generate data that are alike to the original data. By using the dense layer, the generator increases the representation of the noise vector to have sufficient values to be reshaped and fed into the first generator layer that must be identical to the dimension of the final block in the discriminator model.

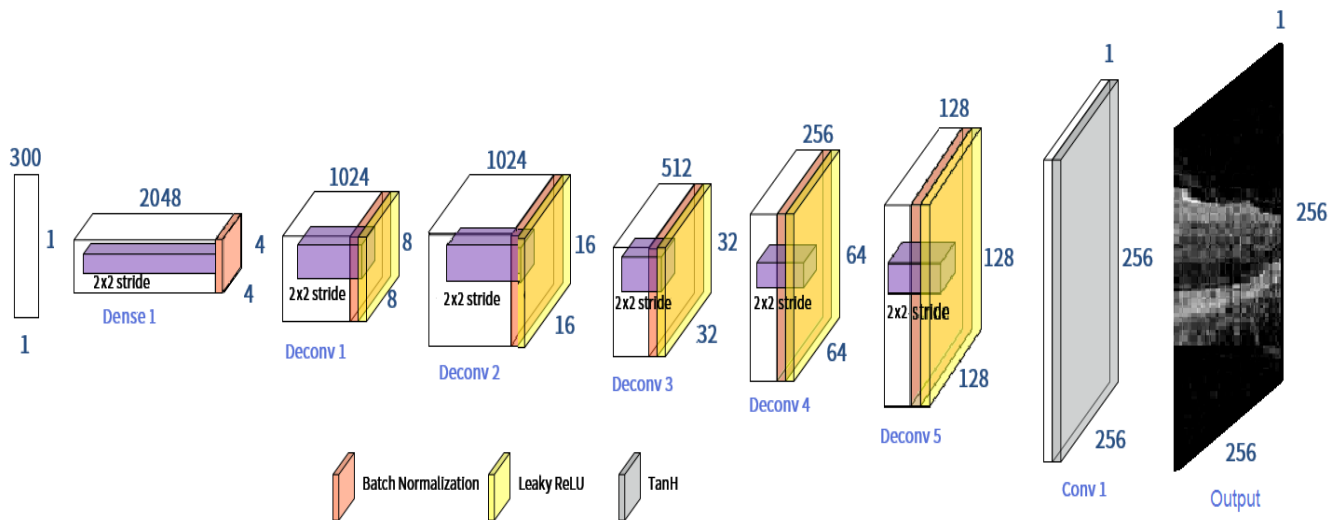


Fig. 65: Representation of the proposed generator model

DCGAN architecture as presented by [136] consists of a fully connected layer reshaped to size $4 \times 4 \times 1024$ and four strided de-convolutional layers to up-sample the image with a 5×5 kernel size. Meanwhile, we tweaked this model by involving 5 blocks where each block implements a deconvolution layer with 2×2 strides, batch normalization, and instead of RELU, an activation function using Leaky-ReLU [143] at the end of each block as presented in Fig. 65.

At first, a fully connected layer reshaped to $4 \times 4 \times 2048$ dimension, and a final convolutional layer are applied, where the output of the "tanh" at the last layer produces the counterfeited image.

5.2.3.2. Discriminator:

The discriminator is qualified for estimating the condition of the data produced by the generator. It is a typical CNN architecture that accepts input data from either the original dataset, or the generated one, and tries to foresee the source of the input sample. For the SD-OCT dataset, the input is an image with dimension size of $256 \times 256 \times 1$.

We presented 5 blocks of CNN layers as architecture for the discriminative model, where each block involves a convolutional layer with batch normalization, followed by a second convolutional layer including a striding of 2 to downscale the image and an extra batch normalization layer since batch-norm stabilizes learning in between layers. Subsequently, Leaky-ReLU is the activation function applied at the end of each CNN block. Contrary to typical CNN models, the discriminative model does not have any max-pooling layer between CNN layers.

The result proceeds, instead of using average-pooling, accompanied by a dense sigmoid layer to outcome a single probability result. The sigmoid output is a probability value to test the reality appearance of the image, where the value "0" means that the generated image is surely fake, while "1" means real, between 0 and 1, it shows a gray uncertainly area. Binary cross-entropy has been employed for the loss since the output is a sigmoid.

Adam has been used as an optimizer with a momentum of 0.5, and a dropout of 0.25 is applied to limits over-fitting at the end of every CNN block except the fully connected layer. Finally, the Learning rate is set to 0.0002.

Fig. 66 presents a demonstration of the proposed discriminator model.

5.2.3.3. Adversarial Model:

The adversarial model is simply the generator-discriminator piled and trained together at the same time. The generator model is attempting to deceive the discriminator and concurrently learning from its errors. The discriminator will try to become more expert at identifying real versus unreal data whilst the generator is working on producing outputs that are progressively near to the original.

5.2.3.4. GAN architecture:

Traditional CNN regularly has a training loss that will be decreased in value during the training process. Therefore, early stop techniques or calculating the loss on the validation set can determine the training end time, which prevents over-fitting. Meanwhile, GANs control two losses (generator / discriminator), thus, to prevent the over-powering for one over the other or the training gets stuck, a balance should be maintained in order to get learn at the same rate and force each other to be improved. Therefore, different learning rates have been implemented for both model and tuning the values based on which regulation the balance should achieve. Then, all generated images for the discriminator training are labeled as alpha and the fake ones as one - alpha instead of 0 and 1 where alpha is a stochastic variable inspected between zero and alpha_max. In order to make the generated prone result appears more realistic. We use in both

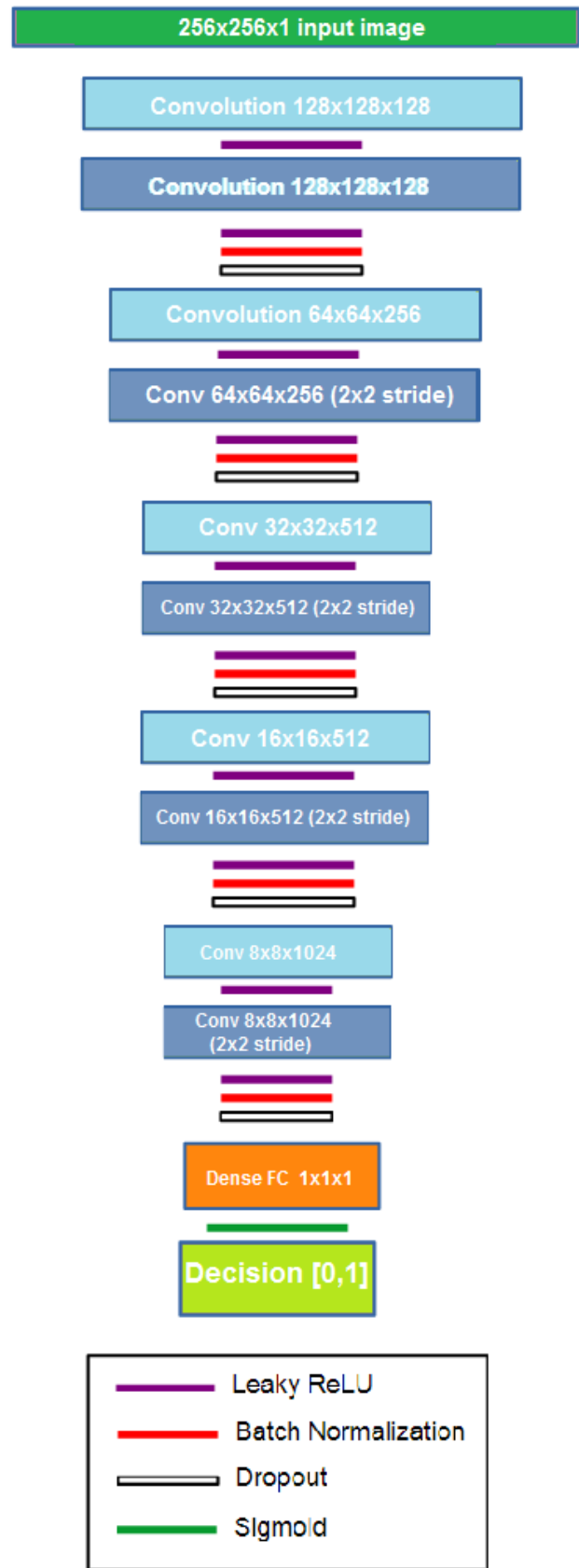


Fig. 66: Representation of the proposed discriminative model

models architecture with five blocks where the following parameters proved an advance in results: Leaky-ReLU is used as the activation function with a negative slope of 0.3 for all layers of both models, except for their final layer where "sigmoid" and "tanh" were used. Weights were initialized to a zero-centered normal distribution with a standard deviation of 0.02. Batch size = 32. Binary Cross-Entropy is applied as the loss function. Adam is adopted as the optimizer algorithm. GAN was trained for 10000 epochs using a learning rate of 0.0002.

5.2.3.5. Synthetic image evaluation

Since SERI dataset was too small for sufficient and efficient training, we combined classic augmentation for the training process. As our baseline, we added the mirrored data (see section 5.2.2) to the original pre-processing set in the direction of forming the entire dataset, in order to enlarge the data being fed into the GAN. Later, we saved the augmented results to increase the amounts of data. We applied the customized DCGAN architecture to separately train every class. After the generator was trained to recognize each class data distribution, it was able to create new samples from a noisy input vector.

Table 12 and Table 13 depict a sample example of synthesized images from both "DME and Normal" cases compared to the real ground-truth images which correspond to real pre-processed SERI data. Finally, we fed the whole set (that includes the original pre-processed images in addition to mirrored and generated ones) into a network to evaluate the classification accuracy result. A detailed explanation of the fine-tuning step is presented in the next section.

Table 12: Examples of generated DME images

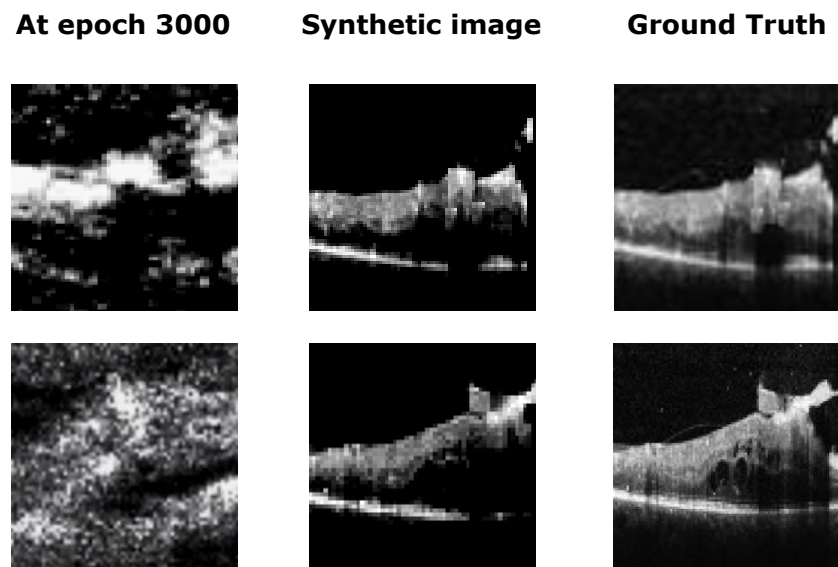
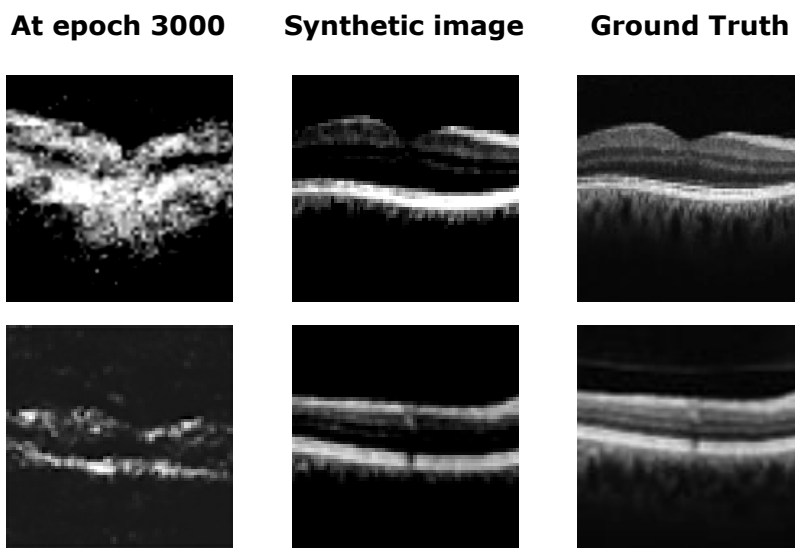


Table 13: . Examples of generated images for normal cases



As seen in Table 12 and Table 13, the first column presents the generated images at epoch 3000, where both loss functions are near to settle down as presented later. The second column presents the synthesized image at epoch 10000, where these generated images seem similar to compared real images presented in the last column. The generated are likely reliable since they show the most needed features to be observed and learned.

Based on these results, we observe a simultaneous decrease, in the beginning, in the loss of the discriminator and generator during their training at different epochs. Fig. 67 presents the loss plot of the discriminator and the generator, where the loss of the discriminator is pursued by some instability for a short epoch's period.

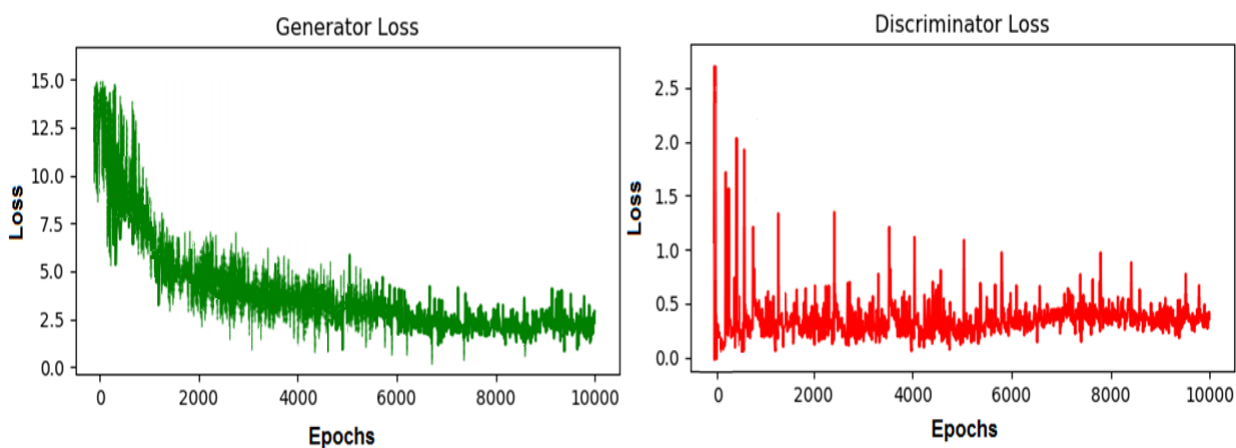


Fig. 67: Generator and Discriminator Loss

We notice that generator loss was first stable higher than the discriminator loss, indicating that the generator was creating images trying to fool the discriminator. However, the generator soon learned to make better quality and more realistic fraudulent images, and its loss dropped nearly above the discriminator's loss within some neighborhood, while the discriminator learned to catch them. We distinguish an asymptotic convergence, where the loss of the generator and discriminator began to stabilize and settled close at about 6000 epochs with the generator's loss being about 2.2, and the discriminator's loss being about 0.39. The continual vacillations in both discriminator and generator loss plot, in the beginning, are signs of the persistent adversarial struggle that resides well past the point at the end which the graph seems to have a neighborly near loss.

The proposed GAN paradigm leads to a method of increasing the cases and the needed experience with a realistically synthetic retinal image representation. From this graph, the results confirm that synthetic retinal SD-OCT images can be created using GANs. Potential employment will be applicable to the area of ophthalmology and generally in medical healthcare. In the current approach, professional trainees or doctors are trained by observing anomalies of patients with particular diseases. The essential rules of diagnosing such patients are discovered while caring. However, for most common diseases such as DME, the trainees are learned by getting unlimited exposure.

We collected considerable numbers of generated images for both classes and formed data groups of synthetic samples additively. To preserve balanced classes, we generated the equivalent number of synthetic images for both classes. To summarize, the augmented data formation process is presented in Table 14, where the total number of images for both classes is 16384 duplicated 3 times making a summation number of 49152 images.

Table 14: Representaton of SERI preparation

	Original pre-processing images (volumes × scans)	Classical augmented images (mirrored)	Synthetic images using GAN (from original and classical)	Total number of images
DME	$16 \times 128 = 2048$	2048	$2048 + 2048 = 4096$	$8192 \times 3 = 24576$
Normal	$16 \times 128 = 2048$	2048	$2048 + 2048 = 4096$	$8192 \times 3 = 24576$

5.3. The proposed method

5.3.1. Fine-tune the pre-trained model

Transfer learning consists of extracted features learned on a task and leveraging them on a different but similar task. TL methodology is based on freezing almost all layers from a pre-trained model to retain the learned features and weights, and leaving only the fully connected ones free. Then append new trainable layers on top of the frozen layers to be trained. Considering SERI (containing 16 DME and 16 normal cases) is a small dataset to train from scratch over a full deep model (which leads to over-fitting), hence, transfer learning techniques can usually be used for such little dataset problems. Since the proposed model in the prior chapter (see section 4.3.1) is previously trained over a similar OCT dataset (CUHK dataset) and because SERI was augmented in this chapter to relatively considerable dataset size, therefore, the fine-tuning technique (which is a branch of TL) is applied with the purpose of achieving significant improvements. Fine-tuning (FT) consists of unfreezing (a part or the whole) pre-trained model and re-trains it on the new dataset with a low learning rate. FT will re-train the fully connected layers and also the convolutional and the pooling layers that incrementally adapt the pre-trained features to the new dataset.

Consequently, our strategy involves fine-tuning the pre-trained CNN model presented in the previous chapter, to enhance its predictive ability based on a model previously trained on a similar dataset. This fine-tuned model could effectively detect the most harmful diabetic pathology (DME) for adults compared to classical learning. This method aims to take advantage of a pre-trained model that has been trained on similar SD-OCT images since OCT images have limited data samples. We can use the already trained architecture and leverage its performance to compose predictions and acquire improved results.

5.3.2. Re-train the model

Almost all DL models either handled the raw OCT images to train an end-to-end CNN demanding a large mass of training data and days to perform classification accuracy or used the feature-based TL models without additional adaptation and further fine-tuning resulting in reduced classification performance. As this is also the case with classical ML, TL has been presented, where the model is trained on a related and similar work for which convenient data is sufficient, and the parameters are fine-tuned for the necessary target work.

The development of a model with an inadequate number of data drives to over-fitting, which has a bad impact on the effectiveness while testing, given that, SERI data is relatively small; therefore, we are distressed to fine-tune the CNN model due to over-fitting concerns, taking into consideration that SERI data is similar to the pre-trained CUHK dataset. Hence, an enlargement of the SERI dataset was performed to extend the number of images that are necessary to enhance the generalizability of the learned model. As a result, fine-tuning an architecture trained on retinal OCT image classification has been explored. This fine-tuning method aims to recognize OCT images with pathology. It also involves formerly a pre-processing step for denoising and flattening OCT images, which is not a common task in the deep learning community.

Considering fine-tuning a network pre-trained on the CUHK dataset for retinal SD-OCT image classification has not been examined. However, in this chapter, we conducted an analysis of the application of a deep TL model based on the network architecture presented in chapter 4, to classify SD-OCT images with DME, for presenting a timely and accurate diagnosis of the most emerge diabetic pathology. The flowchart of the proposed fined-tuned paradigm is shown in Fig. 68. This model took advantage of features extracted from the CUHK OCT dataset using the pre-trained method to fine-tune another similar dataset. The proposed paradigm can be conceivably implemented to ophthalmology clinics to assist clinicians to make a symptomatic decision and spread it to an extensive scope of utilization in medical imaging approaches to assist the ophthalmologist to achieve more accurate examination and proper treatments.

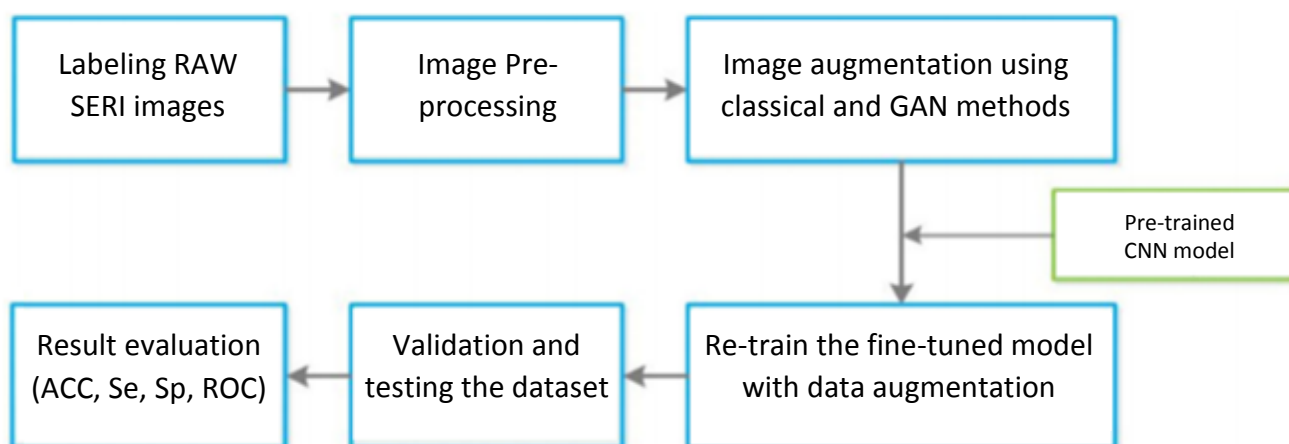


Fig. 68: The proposed approach's workflow diagram for classifying

Firstly, the images were captured, acquired, and labeled by volumes. After, we practiced image pre-processing techniques including normalization, denoising, flattening, and cropping. Posteriorly, the SERI dataset was augmented by mirroring

and synthesizing the pre-processing set. Consequently, shuffled and divided the entire labeled data into training, validation, and testing dataset. Finally, the prediction for the input image was determined by the output from the pre-trained CNN network. The proposed method exploited the OCT known features of the CUHK images and automatically learned the differentiation of the newly learned features from SERI to compose a significant learned feature to increase classification results.

This pre-trained architecture as presented previously in section 4.3.1. is formed of 5 convolutional layers with 3×3 kernel's size following by Relu activation, 3 max-pooling with 2×2 kernel's size, and stride is set to 1 after each of the CNN groups, and several filters varying from 16 to 256. The input to this model is fixed to 256x256 pixels grayscale images for the CUHK dataset. Fig. 69 depicts a presentation of the model trained using the CUHK dataset.

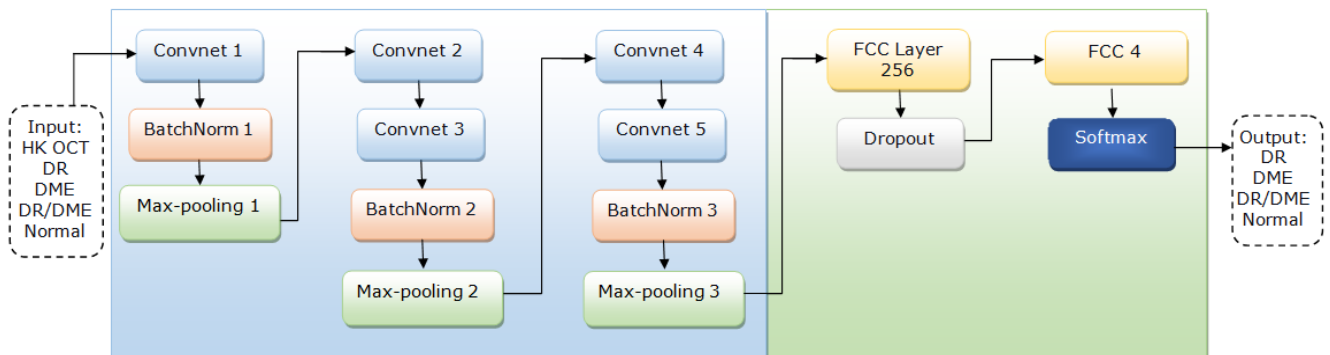


Fig. 69: Flowchart of the model trained over CUHK dataset

For fine-tuning, we froze firstly all CNN layers and the relevant max-pooling layers as feature extractors, therefore only the new classification layers update their weights. Otherwise, the randomly initialized weights in the new classification layers will lead to a large error rate that affects the weights in the trainable CNN layers by back-propagation. As a result, in the re-training process, we manage to unfreeze and update the pre-trained weights by back-propagation in the SERI OCT images to fine-tune the CNN layers to avoid over-fitting. Fig. 70 presents an overview of the new classification layers added to the pre-trained CNN layers.

The CNN layers were initialized by loading the pre-trained saved weights to accelerate the training. The CNN layers are followed by a new classification layer including two adapted Fully-Connected (FCC) layers: the first has 1000 channels; the second performs 2-way classifications adjusted to 2 output classes for DME and NORMAL instead of the 4 output classes of the CUHK. Moreover, between FC layers, a dropout layer has been set to 0.2 to avoid over-fitting when using a new dataset.

Finally, the Softmax classifier layer has been utilized as the model prediction for DME and normal image diagnosis.

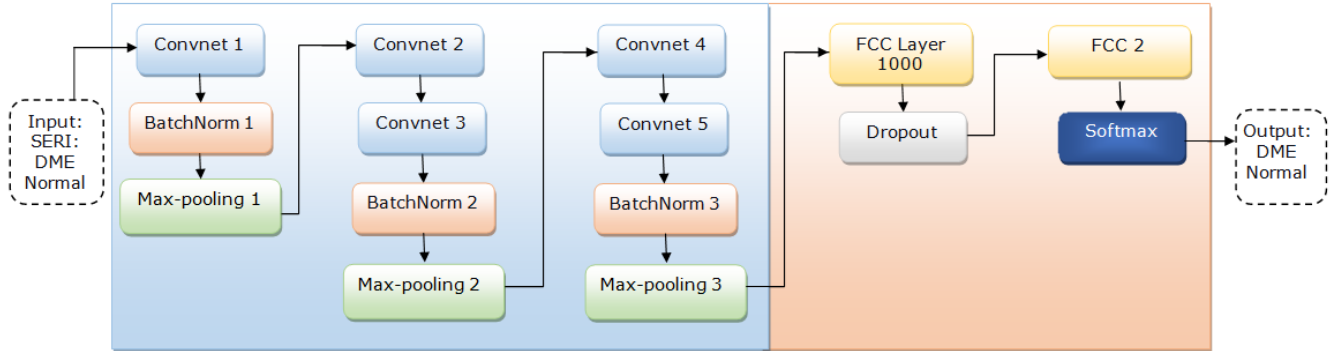


Fig. 70: Flowchart of the pre-trained model over SERI dataset in addition to newly classification (red section)

The newly initialized model was trained on an Intel Xeon E5-2670 2.3 GHz workstation with NVIDIA Tesla K40c GPU, 12 GB GDDR5 with association of Google-Colab Server. The hyper-parameters have been based on the pre-trained model and adapted carefully and by applying the trial and error, where the following values provide the optimal results: We used ADAM optimizer to train layers with a low learning rate of 0.0001, a momentum of 0.5. After 120 epochs, the training was suspended since both accuracy and cross-entropy loss would not be improved any more.

5.4. Performance Analysis

In this chapter, all performance evaluations were achieved using Python including NumPy and scikit.learn modules with Keras. The classification accuracy for the training and validation sets was measured. Next, the confusion matrix is computed alongside with the classification sensitivity and specificity for the validation set. Later, ROC curves were plotted. The formulas for computing these values are defined and explained in chapter 3 section 3.2.5.

The evaluation of the model was conducted by applying the SERI retinal SD-OCT image dataset. Classifying a patients' volume as infected with DME helps him to be redirected to an ophthalmologist immediately. Otherwise, his vision is compromised to be lost.

For this experiment, The SERI dataset consisted originally, as presented in Table 14, of 4096 images from 32 adult patients, divided equally into 16 volumes for DME and 16 volumes for healthy cases. Each volume consists of 128 images where the labeling methodology took into consideration the majority of the volume as an overall case decision despite the possibility of discovering that not all B-scans are labeled as to its volume.

Consequently, assuming that all images within a volume are labeled the same as their volume label, therefore, 2048 images were labeled as DME, and 2048 represented healthy cases. Later, the data was augmented and duplicated producing a total of 49152 images from both cases.

For model evaluation, we shuffled and split the data into 80% for training and 20% for validation where the validation set is divided equally for each category from the SD-OCT dataset resulting to a total of 9830 images for the validation, and the rest images were used for training with a total of 39322 images. We allocated images from the same class a single unique label.

Table 15: Images distribution for the training and validation sets

Number of SD-OCT images	DME	Normal	Total
Training set	19661	19661	39322
Validation Set	4915	4915	9830

The training set was used to re-train the weights of the pre-trained network while evaluating the performance using the validation set. Training and validation set distribution are shown in Table 15.

As presented in Table 16, a confusion matrix for DME and Normal images was identified accurately. By comparing both classes, we observe that 9791 of 9830 images on the validation set were correctly predicted with an accuracy of 99.6%. Alongside, as seen in Table 16, it can be noted that 13 images of DME were incorrectly categorized as normal and 26 images were also incorrectly classified as DME.

Table 16: The confusion matrix of our model's classification on the validation

Confusion Matrix		Predicted Class		
		DME	Normal	Total
Actual Class	Class			
	DME	4889 99.47%	26	4915
	Normal	13	4902 99.73%	4915
	Total	4902	4928	9830

Table 17 depicts the recall, specificity, precision and F1-score for the validation set. The ROC curve reflected the performance measurement of the model to distinguish between DME and NORMAL. The AUC produced by plotting sensitivity versus 1-specificity attained 99.874% as depicted in Fig. 71.

Table 17: Classification report showing the validation ratios

	Recall	Specificity	Precision	F1-Score
DME	99.47%	99.87%	99.73%	99.6%
Normal	99.74%	99.74%	99.47%	99.6%

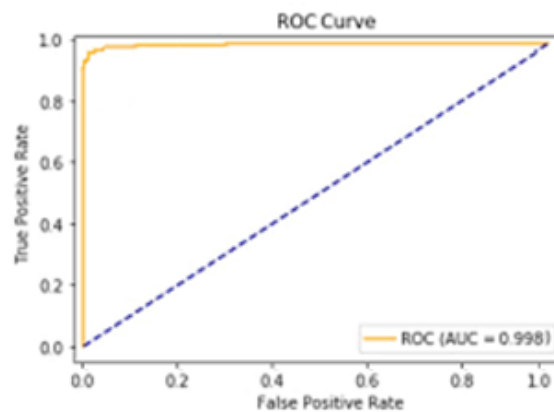


Fig. 71: The ROC curve showed the classification performance

Two other datasets were used as a test for evaluating the prediction model result. The DUKE and the NOOR datasets have been pre-processed to fit the input condition for the reason of the difference in quantity, quality, and size; thus, each dataset has been adapted separately to test the model. Therefore, in order to test the model prediction effectiveness. The DUKE Dataset was priority set to be experimented. The dataset proposed by [86] includes 45 SD-OCT volumes: (15 normal and 15 DME patients) besides 15 AMD cases been excluded from this study. All volumes contain several B-scans that vary from 31 to 97 with a resolution of around 496x768 pixels with some deformation in capturing mechanism. Fig. 72 presents a sample example of the dataset.

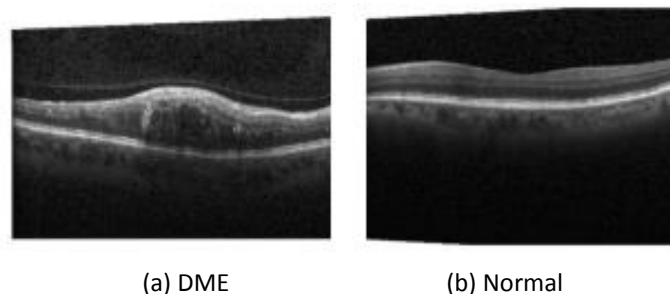


Fig. 72: Example B-scans from Normal, and DME

Firstly, the dataset has been pre-processed including normalization, resizing, denoising, flattening and cropping where all deformed scans have been fixed and reduced to match the model correspondent input dimension fixed to 256x256. Next, every image from each volume has been passed through the model to achieve a prediction result. After counting the prediction result, a basic rule to make the final prediction is to consider a patient as infected if the half+1 of the total predicted image of the volume have been predicted as DME and/or a sequence of infected images has been detected (for example, a chain of 10 to 15 consecutive detected slices).

Both DME and Normal volumes have been placed to be tested where each volume contains usually 97 OCT B-scan, therefore, the patient is predicted as infected if at least 45 to 50 of 97 images are predicted as DME, either a series of scans showing DME where it simulates the situation of eye blinking where DME will not appear in OCT scans. As shown in Table 18, a slice by slice representation for every volume of DME showing the slice of DME in green in the traverse of normal slices presented in red. The results showed that 14 of 15 patients have been predicted as infected with an accuracy of 93.34% and 15 of 15 are predicted as normal cases with an accuracy of 100%.

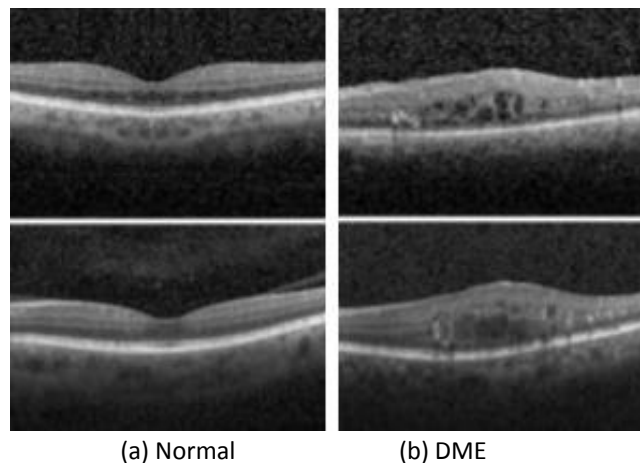


Fig. 73: Example B-scans from Normal, and DME

Another experiment was tested using Noor Eye Hospital Dataset. It consists of 50 normal and 50 DME OCTs volume, where 19, 25, 31, and 61 B-scans per volume are acquired from different patients. Fig. 73 presents an example of OCT images. The dataset was also pre-processed as previously achieved using DUKE Dataset with respecting the characteristic of acquired images where the images were relatively more adequate to use. The result showed that 48 of 50 patients have been predicted as DME infected with an accuracy of 96% and 49 of 50 patients are predicted as normal cases with an accuracy of 98%.

5.5. Discussion

In this chapter, we explored a transfer learning method based on a pre-trained model to recognize DME automatically using pre-processed SD-OCT retinal images that show better result compared to raw data. This method avoided a large number of epochs for convergence with a considerable amount of SERI images, by fine-tuning the pre-trained model, and achieved a relatively equivalent performance to specialists in diagnosing retinal OCT volumes. The model was validated with an adequate validation set incorporating 9830 images from both categories that can additionally yield high performance in accuracy, specificity, sensitivity, and AUC for reaching the accurate prediction.

Thereby, the highly powerful classification and prediction abilities of the TL method were represented, even though handling a relatively mid-size augmented training set. The proposed model can be potentially applied in the pre-diagnosis of other retinal pathologies such as CNV, DR, and glaucoma. OCT, a widely use and reliable technique in the field of ophthalmology, can present high-resolution cross-sectional images of the layers captured in the retina to accurately detect DME. As a general paradigm, the pre-trained network can be utilized to a broad category of SD-OCT image classifications without understanding the fundamental eye pathology process.

The results of the proposed method, when compared with other related models based on DL architectures discussed previously in the-state-of-the-art, showed an advantage that we could notice. Our method focused on re-training a pre-trained model on an OCT-based dataset relatively similar to the one held to be re-trained after augmentation. Both datasets concentrated on DME that is considered the most pathology leading to blindness in the retinal eye diabetic.

Almost all related works exploit a very deep pre-trained well-known architecture. However [95] proposed a very deep model based on GoogLe-Net pre-trained network using ImageNet dataset. This model is intended to classify OCT based on a dissimilar dataset. The model is implemented to classify 3 cases by fine-tuning GoogLe-Net. Additionally, it involves firstly pre-processing the OCT scans. The prediction result of testing showed a ratio of 96% of model accuracy via SVM classifier using cross-validation, this method compared to the proposed model showed less accurate result using a very deep architecture pre-trained over an unrelated dataset.

[108] proposed a transfer learning-based model using a pre-trained VGG-Net for classifying DME using OCT scan volumes. In order to evaluate the model different classifiers have been examined like KNN and RF. The performance evaluation based on the LOPO-CV method and the majority rule has been applied to classify the volume. The experiments, after pre-processing OCT images, showed that the accuracy obtained

was 93%. The proposed fine-tuned model compared to this pre-trained model without fine-tuning, showed better results since the adjustment of the weights via back-propagation is a necessary step to minimize loss.

Table 18: A demonstration for all DME volumes where the number of slices showing DME prediction are colored in "green" compared to false negative colored in "red"

Volume	DME Slice Count	Predicted as DME
DME1	92/97	TRUE
DME2	62/97	TRUE
DME3	59/97	TRUE
DME4	94/97	TRUE
DME5	97/97	TRUE
DME6	93/97	TRUE
DME7	61/61	TRUE
DME8	61/61	TRUE
DME9	29/61	TRUE
DME10	2/31	FALSE
DME11	41/61	TRUE
DME12	53/61	TRUE
DME13	39/61	TRUE
DME14	19/61	TRUE
DME15	60/61	TRUE

Moreover, [110] introduced a deep transfer learned model based on the pre-trained VGG16 CNN network for classifying AMD and DME using OCT images. After pre-processing a separating the dataset, the results of the fine-tuned model showed an accuracy ratio of 98.6%. However, compared to the work proposed in this thesis, the proposed work showed better accuracy results in detecting two types of pathologies (DME and DR) with less network complexity and converge faster using a similar

dataset. [111] presented a work uniting between several pre-trained popular CNN models to merge the reduced features. After a pre-processing step over OCT volumes, the results, using an "LTPO-CV" and different classifiers, showed an accuracy of 93.75% using SVM.

5.6. Conclusion

In this work, a representation of a fined-tuned CNN classification model has been proposed with the ability to detect patterns for diabetic macular edema using OCT images. In the opposite of manual retinal-disease diagnosis based on clinical objective examination and analysis of OCT images, this method showed the capability to automatically predict DME with a high accuracy rate. The result showed that the proposed method has been highly effective for predicting images from an untrained and unseen dataset. The experimental results showed an accuracy of classification over 99.6% using holdout validation method, which is exceptionally promising in the field of early detection of diabetic diseases such as DME to aid ophthalmologists in biomedical technologies. The proposed model showed a better classification in detecting healthy cases compared to DME that has been improved in comparison to other work presented in the state of arts.

Finally, the data pre-processing techniques showed improvement in the results, where the images flattening method helped other methods to standardize the format of model input. Although the augmented images aided to generate a larger dataset driving to avoid over-fitting.

General Conclusion

In this thesis, some deep learning architectures and their applications have been discussed. In particular, attention was provided to Convolutional Neural Network. CNN is formed of an aggregation of convolution and sub-sampling layers and terminated by a prediction in order to determine the class label of the image input to the network. Despite the significant attention that DL models are gaining and their application in many studies in many domain fields, they are yet barely considered in others. This may be proved by the evidence that DL models are based on a neural network structure that requires an enormous dataset to train. Moreover, the main goal of deep models is to extract high-level characteristics and features to utilize them for classification.

Proceeding from this principle, we have generated a deep method based on CNN in order to deal with problems of both multi-labeling and binary classification of diabetic diseases such as DR and DME, which are the direct causes of vision problem, leading to total blindness in the age of working citizens of most countries, which affects the healthy lifestyle of patients for decades. Consequently, to avoid vision problems, this thesis presents an automated system for detecting patients with diabetic retinopathy and macular edema using OCT scans since they are the common scanning method applied in ophthalmology in almost all countries.

In the following, we give highlights emphasizing the proposed deep models. For additional details, we guide the reader to the corresponding chapters.

In chapter 4, a simple new model has been presented based on Convolutional Neural Network and fully trained from scratch for the classification of OCT imaging, which helps the ophthalmologist to discover and treat the patient before it is too late. The proposed CNN model has been trained using the CUHK dataset with 3 blocks of convolutional layers followed by max-pooling layers. The last two fully connected layers have consisted of 256 units and 4 units for the classification. The proposed model was trained and examined using a mid-size dataset consisting of pre-processed 24192 images.

The result showed that the proposed method has been useful for predicting images with pathologies. Compared to other models, the results showed that the proposed method performed better and achieved a test accuracy rate of 99.02%. The proposed method showed a classification advance in patients with and DR in an efficient way to detect both diseases together with a percentage of over 98%.

In chapter 5, a deep transfer learned model has been presented where a fine-tuned CNN classification model has been applied with the capability to detect DME

anomalies using OCT images. The model used SERI dataset in order to retrain the model where this dataset was augmented via classical augmentation method then using GAN to synthesis another amount of images. Therefore, 4096 images were labeled as DME and Normal equally. Later, the data augmentation produced a total of 49152 images from both cases.

This method showed a potential performance to automatically predict Normal cases with a high accuracy rate. For model evaluation, we shuffled and split the data into 80% for training and 20% for validation where the validation set is divided equally for each category from the OCT dataset resulting to a total amount of 9830 images for the validation, and for training, the rest images were used with a total of 39322 images.

The result showed that the proposed method has been highly effective for predicting images from an untrained and unseen dataset. By using the holdout validation method, the experimental results depicted classification accuracy over 99.6%, which is remarkably helpful in the domain of diagnosing diabetic disease to support ophthalmologists.

Conclusion Général

Dans cette thèse, certaines architectures d'apprentissage en profond et leurs applications ont été discutées. En particulier, une attention a été accordée au réseau de neurones convolutifs "Convolutional Neural Network". CNN est formé d'une agrégation de couches de convolution et de sous-échantillonnage et se termine par une prédiction afin de déterminer l'étiquette de classe de l'image d'entrée du réseau. Malgré l'attention significative que les modèles DL gagnent et leur application dans de nombreuses études dans de nombreux domaines, ils sont encore à peine considérés dans d'autres. Cela peut être prouvé par la preuve que les modèles DL sont basés sur une structure de réseau neuronal qui nécessite un énorme ensemble de données pour s'entraîner. De plus, l'objectif principal des modèles profonds est d'extraire des caractéristiques et des fonctionnalités de haut niveau pour les utiliser pour la classification.

Partant de ce principe, nous avons généré une méthode approfondie basée sur CNN afin de traiter les problèmes de multi-étiquetage et de classification binaire des maladies diabétiques telles que DR et DME, qui sont les causes directes des problèmes de vision, conduisant à la cécité totale. À l'âge des citoyens actifs de la plupart des pays, ce qui affecte le mode de vie sain des patients pendant des décennies. Par conséquent, pour éviter les problèmes de vision, cette thèse présente un système automatisé de détection des patients atteints de rétinopathie diabétique et d'œdème maculaire à l'aide de scans OCT car ils sont la méthode de numérisation courante appliquée en ophtalmologie dans presque tous les pays.

Dans ce qui suit, nous mettons en évidence les modèles profonds proposés. Pour plus de détails, nous guidons le lecteur vers les chapitres correspondants.

Dans le chapitre 4, un nouveau modèle simple a été présenté basé sur Convolutional Neural Network et entièrement formé à partir de début pour la classification de l'imagerie OCT, ce qui aide l'ophtalmologiste à découvrir et à traiter le patient avant qu'il ne soit trop tard. Le modèle CNN proposé a été formé à l'aide de l'ensemble de données CUHK avec 3 blocs de couches convolutives suivies de couches de "max-pooling". Les deux dernières couches entièrement connectées ont consisté en 256 unités et 4 unités pour la classification. Le modèle proposé a été formé et examiné à l'aide d'un ensemble de données de taille moyenne composé de 24192 images prétraitées. Le résultat a montré que la méthode proposée a été utile pour prédire des images avec des pathologies. Par rapport aux autres modèles, les résultats ont montré que la méthode proposée fonctionnait mieux et atteignait un taux d'exactitude des tests de 99,02%. La méthode proposée a montré une avancée dans la classification des patients avec et DR d'une manière efficace pour détecter les deux maladies avec un pourcentage de plus de 98%.

Dans le chapitre 5, un modèle d'apprentissage par transfert profond a été présenté dans lequel un modèle de classification CNN affiné a été appliqué avec la capacité de détecter les anomalies DME à l'aide d'images OCT. Le modèle a utilisé l'ensemble de données SERI afin de recycler le modèle où cet ensemble de données a été augmenté via une méthode d'augmentation classique, puis en utilisant GAN pour synthétiser une autre quantité d'images. Par conséquent, 4096 images ont été étiquetées comme DME et Normal de la même manière. Plus tard, l'augmentation des données a produit un total de 49152 images des deux cas.

Cette méthode a montré une performance potentielle pour prédire automatiquement les cas normaux avec un taux de précision élevé. Pour l'évaluation du modèle, nous avons mélangé et divisé les données en 80% pour la formation et 20% pour la validation où l'ensemble de validation est divisé également pour chaque catégorie de l'ensemble de données OCT, ce qui donne un total de 9830 images pour la validation et pour la formation, les autres images ont été utilisées avec un total de 39322 images.

Le résultat a montré que la méthode proposée s'est avérée très efficace pour prédire des images à partir d'un ensemble de données non formé et invisible. En utilisant la méthode de validation d'exclusion, les résultats expérimentaux ont montré une précision de classification supérieure à 99,6%, ce qui est remarquablement utile dans le domaine du diagnostic de la maladie diabétique pour soutenir les ophtalmologistes.

Future Perspectives

In the future works, we propose as perspective some methods based on our methods presented in this work. Among these applications, we can mention:

- Using different type of networks: such as Auto-Encoder (AE) Neural Network, RNNs, R-CNN, Fast R-CNN, and FCNN.
- Using "One-shot" learning makes use of prior knowledge of learnt categories and allows for learning on minimal training examples.
- Using metadata, like properties of the learning problem, algorithm properties or patterns previously derived from the data; it is possible to learn, select, alter or combine different learning algorithms to effectively solve an OCT learning problem.
- We plan to improve GAN works by using different models with a variety of numbers of synthesized images.
- We plan to gather all datasets from different institutions and hospitals to build a standard dataset containing almost all diabetic pathologies.

List of Publications

Conference Article:

- Safarjalani, Rami & Sidibé, Désiré & Ainouz, Samia & Shahin, Ahmad & Meriaudeau, Fabrice. (2018). Diabetic Retinal Tomographical Image Classification Using Convolutional Neural Network. 1-6. 10.1109/ICIAS.2018.8540616.

Submitted Article:

- Safarjalani, Rami & Sidibé, Désiré & Ainouz, Samia & Shahin, Ahmad & Meriaudeau, Fabrice. (2020). Convolutional Neural Network classification model for Diabetic Macular Edema using OCT. Biomedical Signal Processing and Control.

Bibliography

- [1] National institute of diabetes and digestive and kidney diseases, <https://bit.ly/3h3kDH1>
- [2] International Diabetes Federation. <https://bit.ly/3i2WS37>
- [3] Centers for Disease Control and Prevention (CDC) "National Diabetes Statistics Report, 2020". <https://bit.ly/330murb>
- [4] W. Shiel, "Medical Definition of Eye" .<https://bit.ly/2DygvkF>
- [5] Photoreceptors. <https://bit.ly/3lOeNgt>
- [6] Normal Retinal Anatomy. <https://bit.ly/3hbaHeV>
- [7] What is macular degeneration. <https://bit.ly/35aW7l4>
- [8] B. Klein, R. Klein. "Cataracts and macular degeneration in older americans". Archives of Ophthalmology, 100(4): 571, 1982.
- [9] D. Turbert. "What is Cytomegalovirus Retinitis" <https://bit.ly/3h56Uj3>
- [10] Cytomegalovirus retinitis. <https://bit.ly/3jQGD9S>
- [11] Retinal Detachment. <https://mayocl.in/358mGHv>
- [12] Collaborative Normal-Tension Glaucoma Study Group et al. "Comparison of glaucomatous progression between untreated patients with normal-tension glaucoma and patients with therapeutically reduced intraocular pressures". American journal of ophthalmology, 126(4):487–497, 1998.
- [13] What is glaucoma. <https://bit.ly/321onV7>
- [14] What is glaucoma. <https://bit.ly/35aWrjM>
- [15] TY. Wong, A. Shankar, et al. "Prospective cohort study of retinal vessel diameters and risk of hypertension" BMJ. 329 (7457): 79. 2004.
- [16] J. TIMMER, "AI trained to spot heart disease risks using retina scan". <https://bit.ly/3i27IMt>
- [17] Lucentis Ranibimuzab injection. <https://bit.ly/3h2a1Is>
- [18] Diabetic retinopathy (The Basics). <https://bit.ly/3jKNMsc>
- [19] T. Wong, C. Cheung, M. Larsen, et al. "Diabetic retinopathy", Nature Rev. Disease Primers. 2016.
- [20] P. Porwal, S. Pachade, M. Kokare, et al. "Indian diabetic retinopathy image dataset (IDRiD): Diabetic Retinopathy – Segmentation and Grading Challenge", Medical Image Analysis, 59, 2020.
- [21] Y.R. Sharma, R. Karthikeya R, et al. "Recent Advances in Management of Diabetic Macular Edema", Current Diabetes Reviews 11, no. 2:79–97. 2015.
- [22] Diabetic Retinopathy Guidelines, "The Royal College of Ophthalmologists", December 2012, <https://bit.ly/2Dyhi57>

- 23 J. Wells, A. Glassman, A. Ayala et al. "Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema", *The New England Journal of Medicine*, 372, (193–1203), 2015.
- [24] F. Bandello, M.B. Parodia, et al. "Diabetic Macular Edema" *Dev Ophthalmol. Basel, Karger*, 47: 73–110. 2010.
- [25] Y. Chung, Y. Ho Kim, S. Ha, et al. "Role of Inflammation in Classification of Diabetic Macular Edema by Optical Coherence Tomography", *Journal of Diabetes Research*, Article ID 8164250, 2019.
- [26] National Eye Institute. <https://bit.ly/3bulpKZ>
- [27] B. Cense, N. Nassif, T. Chen, M. Pierce, S. Yun "Ultrahigh-resolution high-speed retinal imaging using spectral-domain optical coherence tomography". *Optics Express*, 12(11):2435–2447, 2004.
- [28] W. Jackman, J. Webster "On photographing the eye of the living human retina". *Phila. Photogr*, 23, 340–341. 1886.
- [29] M. Karmel "Retinal Imaging: Choosing the Right Method". *EyeNet Magazine*. July 2014.
- [30] Normal Fundus - adult. <https://bit.ly/3gZqxcP>
- [31] P. Godara, A. Dubis, A. Roorda, et al. "Adaptive optics retinal imaging: emerging clinical applications." *Optom Vis Sci*; 87(12):930-41. 2010.
- [32] Carroll J, Kay DB, et al. "Adaptive optics retinal imaging-clinical opportunities and challenges", *Curr Eye Res*; 38(7):709-21. 2013.
- [33] P. Godara, A.M. Dubis, et al. "Adaptive Optics Retinal Imaging : Emerging Clinical Applications". *Optom Vis Sci*. 87(12) : 930–941. 2010.
- [34] Z. Hu, Q. Liu, Y.M. Paulus, " New frontiers in retinal imaging". *Int. J. Ophthalmic Res*, 2, 148–158. 2016.
- [35] K. Takayama, S. Ooto, et al. "High-Resolution Imaging of the Retinal Nerve Fiber Layer in Normal Eyes Using Adaptive Optics Scanning Laser Ophthalmoscopy" . *PLoS ONE* 7(3): e33158. 2012.
- [36] C. Tian, T. Feng, et al. "Non-Contact Photoacoustic Imaging Using a Commercial Heterodyne Interferometer", *IEEE Sens J*; 16(23): 8381-8388. 2016.
- [37] C. Tian, w. Zhang W, et al. "Noninvasive chorioretinal imaging in living rabbits using integrated photoacoustic microscopy and optical coherence tomography". *Opt. Express*, 25, 15947–15955. 2017.
- [38] Y. Li, X. Xia, Y.M. Paulus, "Advances in Retinal Optical Imaging", *Photonics*; 5(2): 9. 2018.
- [39] R. Silverman, "High-resolution ultrasound imaging of the eye – a review", *Clin Experiment Ophthalmol*. 2009.
- [40] A. Roka, J.A. Rubero, et al. "Non-traumatic Retinal Detachment Diagnosed via Bedside Ultrasonography". *Cureus* 10(6): e2771. 2018.

- [41] T.Q. Duong, E.R. Muir "Magnetic Resonance Imaging of the Retina" *Jpn J Ophthalmol.* 53(4): 352–367. 2010.
- [42] Sajda P, "Machine learning for detection and diagnosis of disease". *Annu Rev Biomed Eng* 8:537–565, 2006.
- [43] Povazay, B., K. Bizheva, A.H. Unterhuber, W. Drexler, A. Apolonski, et al.. 2002." Submicrometer axial resolution optical coherence tomography" *Opt. Lett.* 27: 1800. Crossref, Medline, CAS.
- [44] Swanson, E.A., J.A. Izatt, M.R. Hee and J.G. Fujimoto. "In-vivo retinal imaging by optical coherence tomography". *Opt. Lett.* 18: 1864–1866. Crossref, Medline, CAS, 1993.
- [45] J. Fujimoto, E. Swanson. "The development, commercialization and impact of optical coherence tomography". *Invest Ophthalmol Vis Sci*; 57(9). 2016.
- [46] S. Aumann, S. Donner, J. Fischer et al. "Optical Coherence Tomography (OCT): Principle and Technical Realization". *High Resolution Imaging in Microscopy and Ophthalmology*. Springer, Cham, 2019.
- [47] D. Huang, et al. "Optical coherence tomography". *Science*; 254(5035): 1178–81. 1991.
- [48] A.F. Fercher, et al. "Optical Coherence Tomography - principles and applications". *Rep Prog Phys.*; 66(2): 239–303. 2003.
- [49] Global Leader in computer Vision and Deep Learning "Explaining OCT scans". <https://bit.ly/3hGwj2X>
- [50] AF. Fercher et al. "Measurement of intraocular distances by backscattering spectral interferometry". *Opt Commun*; 117: 43–8. 1995.
- [51] W. Drexler, et al. "Optical coherence tomography today: speed, contrast, and multimodality". *J Biomed Opt*; 19(7):071412. 2014.
- [52] Ophthalmology Web. <https://bit.ly/3IW6FdB>
- [53] Heidelberg Engineering. <https://bit.ly/2ZdlwFY>
- [54] S. Aumann, S. Donner, J. Fischer and F. Müller "Optical Coherence Tomography (OCT): Principle and Technical Realization. In: Bille J. (eds) *High Resolution Imaging in Microscopy and Ophthalmology*". Springer, Cham. 2019.
- [55] D. Gildea "The diagnostic value of optical coherence tomography angiography in diabetic retinopathy: a systematic review". *International Ophthalmology.* 39 (10): 2413–2433. 2019.
- [56] E. Talisa, K Nadia, et al. "A review of optical coherence tomography angiography (OCTA)". *International Journal of Retina and Vitreous.* 1 (1): 5. 2015.
- [57] G. Hinton, S. Osindero and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," in *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, July 2006.

- [58] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks" in Proc. Neural Inf. Process. Syst., Cambridge, MA, USA, p153–160, 2007.
- [59] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
- [60] Y. Li, X. Bai, X. Liang, L. Jiao "Sparse Restricted Boltzmann Machine Based on Multiobjective Optimization". In: Shi Y. et al. (eds) Simulated Evolution and Learning. SEAL 2017. Lecture Notes in Computer Science, vol 10593. Springer, Cham. 2017.
- [61] Autoencoders. Stanford University. <https://stanford.io/3bx8zNc>
- [62] Implementing of Convolutional Neural Network using Keras. <https://bit.ly/321O9bZ>
- [63] J. Tang, C. Deng, G. B. Huang and B. Zhao, "Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine," in IEEE Transactions on Geo-science and Remote Sensing, 53(3), p. 1174-1185, 2015.
- [64] Liu, Ying & Cui, Hong-Yuan & Kuang, Zheng & Li, Guoqing. "Ship Detection and Classification on Optical Remote Sensing Images Using Deep Learning". ITM Web of Conferences. 12. 05012. (2017).
- [65] S. Barra, S.M. Carta et al. "Deep Learning and Time Series-to-Image Encoding for Financial Forecasting," IEEE/CAA J. Autom. Sinica, vol. 7, no. 3, pp. 683-692, 2020.
- [66] Shu-Yu Chen, Wanchao Su, Lin Gao, Shihong Xia, and Hongbo Fu. "DeepFaceDrawing: Deep Generation of Face Images from Sketches" Hong Kong University. <https://bit.ly/35b98LA>
- [67] M.A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani and R. R Yager, "Deep learning approach for active classification of electrocardiogram signals," Information Sciences 345, 340–354, 2016.
- [68] Lin Zhou, Yan Yan, Xingbin Qin, Chan Yuan, Dashun Que and Lei Wang, "Deep learning-based classification of massive electrocardiography data," 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, pp. 780-785. 2016.
- [69] A. Ben Said, A. Mohamed, T. Elfouly, K. Harras and Z. J. Wang, "Multimodal Deep Learning Approach for Joint EEG-EMG Data Compression and Classification," 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, pp. 1-6. 2017.
- [70] C. Li and C. Yang, "The research on traffic sign recognition based on deep learning," 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, pp. 156-161. 2016.
- [71] X. Peng, N. Ratha and S. Pankanti, "Learning face recognition from limited training data using deep neural networks," 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 1442- 1447. 2016.

- [72] D. Menotti et al., "Deep Representations for Iris, Face, and Fingerprint Spoofing Detection," in IEEE Transactions on Information Forensics and Security, vol. 10, no. 4, pp. 864-879, 2015.
- [73] B. Bena, J. Kalita, "Introducing Aspects of Creativity in Automatic Poetry Generation", Submitted on 6 Feb 2020. <https://bit.ly/3igxibj>
- [74] M. Hampson, "New AI Dupes Humans Into Believing Synthesized Sound Effects Are Real". 2020. <https://bit.ly/2QUqBj0>
- [75] S. Ghose, J. J. Prevost, "AutoFoley: Artificial Synthesis of Synchronized Sound Tracks for Silent Videos with Deep Learning," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2020.3005033. 2020.
- [76] The Free Library, "Dubai gets world's first AI-composed theme song", Syndi-Gate Media. 2018. <https://bit.ly/2DxhWzQ>
- [77] Mordvintsev, Alexander; Olah, Christopher; Tyka, Mike, "DeepDream - a code example for visualizing Neural Networks". Google Research. Archived from the original on 2015-07-08.
- [78] P. Tully, L. Foster. "Repurposing Neural Networks to Generate Synthetic Media for Information Operations". 2020. <https://bit.ly/3bxh2jt>
- [79] Softmax Regression. <https://bit.ly/2Zd66m5>
- [80] Artificial Neural Networks. <https://bit.ly/2DBMf8G>
- [81] Decision Trees. <https://bit.ly/3h35TrN>
- [82] Random Forest. <https://bit.ly/3h4AkhI>
- [83] J.M. Ashfaq, A. Iqbal. "Introduction to Support Vector Machines and Kernel Methods". April 2019.
- [84] ROC curve analysis. <https://bit.ly/3IVTOYT>
- [85] S. Sankar, D. Sidibe, F. Meriaudeau, et al, "An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images", Computer methods and programs in biomedicine, vol. 139, p. 109-117. 2017.
- [86] PP. Srinivasan, LA. Kim, PS. Mettu et al. " Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images". Biomed Opt Express; 5(10):3568–77. 2014.
- [87] k. Alsaih, G. Lemaitre, et al. "Machine learning techniques for diabetic macular edema classification on SD-OCT images," Biomedical engineering online, vol. 16, no. 1, pp. 68, 2017.
- [88] G. Lemaitre, M. Rastgoo, et al. "Classification of SD-OCT Volumes using Local Binary Patterns: Experimental Validation for DME Detection", Journal of Ophthalmology, May 2016.

- [89] Y-Y. Liu, M. Chen, H. Ishikawa et al. "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding". *Medical Image Anal*; 15:748–59. 2011.
- [90] D. Ciresan, J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," pp. 2852–2860, Dec. 2012.
- [91] C.S. Lee, A.J Tyring, et al. "Deep-learning based, automated segmentation of macular edema in optical coherence tomography," *bioRxiv*, p. 135640, 2017.
- [92] X. Liu, T. Fu, Z. Pan et al. "Automated layer segmentation of retinal optical coherence tomography images using a deep feature enhanced structured random forests classifier", *IEEE journal of biomedical and health informatics*, 23(4), 1404–1416, 2018.
- [93] D. Lu, M. Heisler et al. "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network." *Medical image analysis* vol. 54. 2019.
- [94] G. Karthik, J. Sivaswamy. "Segmentation of Retinal Cysts from Optical Coherence Tomography Volumes via Selective Enhancement." *IEEE journal of biomedical and health informatics* vol. 23, 1: 273-282. 2019.
- [95] S.P.K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomedical Optics Express*, vol. 8, pp. 579-592, 2017.
- [96] C. Szegedy, W. Liu, Y. Jia et al. "Going Deeper with Convolutions", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
- [97] L. Fang, Y. Jin, L. Huang, et al. "Iterative fusion convolutional neural networks for classification of optical coherence tomography images". *J Vis Commun Image Represent* 59:327–333. 2019.
- [98] R. Rasti, H. Rabbani et al. "Macular OCT classification using a multi-scale convolutional neural network ensemble". *IEEE Trans Med Imaging* 37(4):1024–1034. 2019.
- [99] P. Mehta, AY. Lee, C. Lee et al. "Multilabel multiclass classification of OCT images augmented with age, gender and visual acuity data". *bioRxiv*: 316349. 2018.
- [100] G.C.Chan, S.A. Shah, F. Meriaudeau, et al. "Deep features and data reduction for classification of sd-oct images: Application to diabetic macular edema". In *IEEE 2018 International Conference on Intelligent and Advanced System (ICIAS)*, 2018.
- [101] Q. Ji, W. He, J. Huang, Y. Sun. "Efficient Deep Learning-Based Automated Pathology Identification in Retinal Optical Coherence Tomography Images". *Algorithms*, 11, 88. 2018.

- [102] DK. Hwang, CC. Hsu, et al. "Artificial intelligence-based decision-making for age-related macular degeneration". *Theranostics* 9(1):232–245. 2019.
- [103] K. Nugroho "A comparison of handcrafted and deep neural network feature extraction for classifying optical coherence tomography (OCT) images". In: second International Conference on Informatics and Computational Sciences (ICICoS), IEEE, pp 1-6. 2018.
- [104] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1: 886–893. 2005.
- [105] T. Ojala, M. Pietikainen, and D. Harwood. "Performance evaluation of texture measures with classification based on kullback discrimination of distributions". IEEE Comput. Soc. 12th International Conference on Pattern Recognition, 1: 582–585. 1994.
- [106] K. He, X. Zhang et al. "Deep residual learning for image recognition". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. 2016.
- [107] G. Huang, Z. Liu et al. "Densely connected convolutional networks". arXiv: 1608.06993. 2016.
- [108] M. Awais, H. Mller, et al. "Classification of sd-oct images using a deep learning approach," in 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 489–492, Sept 2017.
- [109] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition" arXiv preprint arXiv: 1409.1556, 2014.
- [110] Li. Feng, C. Hua et al. "Fully automated detection of retinal disorders by image-based deep learning", *Graefe's Archive for Clinical and Experimental Ophthalmology*. 2019.
- [111] G.C.Y. Chan, R. Kamble, S.A.A. Shah et al. "Fusion of Deep Learning Architectures for Automatic Classification of Normal and Diabetic Macular Edema (DME) Detection on Optical Coherence Tomography (OCT) Images," submitted to EMBC International Conference of IEEE Engineering in Medicine and Biology Society. 2018.
- [112] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks", 25th International Conference on Neural Information Processing Systems, 1: (1097–1105). 2012.
- [113] O. Perdomo, S. Otalora, et al, "Oct-net: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes," in 2018 IEEE International Symposium on Biomedical Imaging (ISBI), Accepted, April 2018.
- [114] S.A. Kamran, S. Saha et al "Optic-Net: A Novel Convolutional Neural Network for Diagnosis of Retinal Diseases from Optical Tomography Images" 18th IEEE International Conference on Machine Learning and Applications (ICMLA). 2019.

- [115] "Department of Ophthalmology and Visual Sciences". The Chinese University of Hong Kong.
<https://bit.ly/32ZHQ89>
- [116] J. M. Schmitt, S. H. Xiang, and K. M. Yung, "Speckle in optical coherence tomography," *Journal of Biomedical Optics*, 4(1), 95–105, 1999.
- [117] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 60–65, 2005.
- [118] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Transactions on Image Processing*, 18(10), 2221–2229, 2009.
- [119] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, "Image denoising with block-matching and 3d filtering" in *Proceedings of SPIE*, vol. 6064. 2006.
- [120] N. Srivastava, G. Hinton, et al, "Dropout: a simple way to prevent neural networks from overfitting", *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [121] GridSearchCV from scikit sklearn model selection. <https://bit.ly/3jOy7lp>
- [122] L. Bottou, F. E. Curtis, J. Nocedal. *Optimization Methods for Large-Scale Machine Learning*. Technical Report 1606.04838, arXiv, 2016.
- [123] T. Tieleman, G. Hinton. Lecture 6.5. RMSPROP: Divide the gradient by a running average of its recent magnitude. COURSERA: *Neural Networks for Machine Learning*, 2012.
- [124] *Optimizing Gradient Descent*. <https://bit.ly/2DyjY2F>
- [125] M.D. Zeiler. ADADELTA: an adaptive learning rate method. *CoRR*, abs/1212.5701, 2012.
- [126] D. Kingma, J. Ba "Adam: A Method for Stochastic Optimization", latest version 30 Jan 2017.
- [127] An overview of gradient descent optimization algorithms. <https://bit.ly/3h9Vdb2>
- [128] Usage of initializers. <https://bit.ly/2Dy9Eb3>
- [129] Usage of activations. <https://bit.ly/3IRkkTj>
- [130] NVIDIA Corporation. *Nvidia tesla product literature*. <https://bit.ly/323DNIF>
- [131] Usage of callbacks. <https://bit.ly/3ITltdc>
- [132] B Cense, TC Chen, et al, "Ultra-high speed and ultra-high resolution spectral-domain optical coherence tomography and optical doppler tomography in ophthalmology". *Bulletin-Societe Belge D Ophthalmologie*, 302:123, 2006.
- [133] J. Massich, M. Rastgoo, G. Lemaître, C.Y. Cheung, T.Y. Wong, D. Sidibé, F. Mériaudeau, "Classifying DME vs normal SD-OCT volumes: A review", *Pattern Recognition (ICPR)*, 23rd International Conference on, 1297-1302. 2016.

- [134] P. Srinivasan, L. Kim, P.S. Mettu et al, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express* 5, 3568-3577. 2014.
- [135] R. Rasti, H. Rabbanin, A. Mehri, F. Hajizadeh, "Macular OCT Classification using a Multi-Scale Convolutional Neural Network Ensemble", *IEEE Transactions on Medical Imaging*, 37(4), 1024-1034, 2018.
- [136] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [137] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv: 1511.06434, 2015.
- [138] k. Tero, A. Timo, L. Samuli, L. Jaakko, "Progressive growing of GANs for improved quality, stability, and variation". CoRR, abs/1710.10196, 2017.
- [139] P. Luis, W. Jason. "The effectiveness of data augmentation in image classification using deep learning". In: Stanford University research report, 2017.
- [140] J. Lemley, S. Barzrafkan, P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy". In: *IEEE Access*. 2017.
- [141] Ekin DC, Barret Z, Dandelion M, Vijay V, Quoc VL. AutoAugment: learning augmentation policies from data. ArXiv preprint. 2018.
- [142] Maayan F-A, Eyal K, Jacob G, Hayit G. GAN-based data augmentation for improved liver lesion classification. arXiv preprint. 2018.
- [143] S. Chintala, E. Denton, M. Arjovsky, and M. Mathieu. "How to train a gan? Tips and tricks to make gans work". 2016.