جامعة سيدي محمد بن عبد الله بفاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES



كلية العلوم ظهر المهراز فاس
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

N° d'ordre : 48/2024

Centre des Etudes : ***Sciences et Techniques et Sciences Médicales***

Formation Doctorale : ***Sciences et Technologies de l'Information et de la Communication***

Discipline : ***Informatique.***

Spécialité : ***Informatique.***

Laboratoire : ***Laboratoire d'Informatique, Signaux, Automatique et Cognitivisme.***

THÈSE DE DOCTORAT

Titre de la thèse

Diabetic Retinopathy Prediction Based on Deep Learning Architectures

Présentée par :

OULHADJ Mohammed

Soutenue le 27/09/2023 devant le jury :

Nom et prénom	Etablissement	Grade	Qualité
Pr. TAIRI Hamid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Président
Pr. BELLACH Benaissa	Ecole Nationale des Sciences Appliquées, UMP, Oujda	PES	Rapporteur & Examinateur
Pr. TAIME Abderazzak	Ecole Supérieure de Technologie, Khénifra	MCH	Rapporteur & Examinateur
Pr. EL FAZAZY Khalid	Faculté des Sciences Dhar El Mahraz, Fès	PES	Rapporteur & Examinateur
Pr. BENATIYA ANDALOUSSI Idriss	Ophtalmology Department Hassan II Hospital, Fès	PES	Examinateur
Pr. AHERRAHROU Noura	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinatrice
Pr. HJOUJI Amal	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinatrice
Pr. MAHRAZ Mohamed Adnane	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Examinateur
Pr. YAHYAYOUY Ali	Faculté des Sciences Dhar El Mahraz, Fès	PES	Examinateur
Pr. RIFFI Jamal	Faculté des Sciences Dhar El Mahraz, Fès	MCH	Directeur de thèse

Année Universitaire : 2023-2024

Résumé de la thèse

La rétinopathie diabétique est l'une des complications les plus graves du diabète, pouvant affecter la vue et entraîner la cécité. À l'échelle mondiale, elle figure parmi les principales causes de cécité chez les personnes âgées. En fonction du degré de gravité de la rétinopathie diabétique, l'ophtalmologue adapte le traitement en conséquence. Un dépistage précoce ainsi qu'un suivi régulier sont essentiels pour évaluer avec précision l'évolution de la maladie. Cependant, la détection manuelle de cette pathologie demeure très coûteuse en termes de temps et d'expertise. Au cours de la dernière décennie, l'apprentissage profond a réalisé des avancées significatives dans l'automatisation de la détection et de la prédiction des maladies. Cela constitue donc une solution optimale pour automatiser la détection précoce de la rétinopathie diabétique à partir d'images médicales.

Cette thèse vise à développer des méthodes automatisées pour la détection précoce de la rétinopathie diabétique et à évaluer précisément le stade de la maladie, dans le but d'aider les ophtalmologues à offrir un traitement personnalisé aux patients en fonction de leur état clinique. La première approche propose une méthode automatique pour la détection de la rétinopathie diabétique en utilisant l'apprentissage profond et le recalage déformable. Cette méthode inclut une nouvelle étape de prétraitement, consistant à appliquer le recalage déformable à la rétine afin qu'elle occupe entièrement l'espace de l'image. Les images prétraitées sont ensuite introduites dans quatre modèles d'apprentissage profond. Enfin, une technique de vote d'ensemble est utilisée lors de la phase de test pour générer une prédiction unique basée sur les résultats des quatre modèles. La seconde approche repose sur un réseau de capsules modifié, récemment proposé, combiné à un bloc d'Inception modifié, pour créer une méthode automatique de prédiction de la rétinopathie diabétique. Cette méthode intègre également une étape de prétraitement avancée qui applique un algorithme de transformée en ondelettes discrètes à l'image du fond de l'œil. En exploitant la structure hiérarchique de la transformée en ondelettes discrètes, cette approche génère trois niveaux de décomposition de la même image, fournissant ainsi des informations plus détaillées à partir d'une seule image. Enfin, deux approches novatrices pour la détection automatique de la rétinopathie diabétique ont été proposées. La première approche combine un modèle Vision Transformer pré-entraîné avec un réseau de capsules modifié pour l'extraction des caractéristiques, intégrés à un modèle de concaténation permettant de prédire le stade de la rétinopathie diabétique. La seconde approche fait appel au sous-modèle ViTIncep, au bloc CBAM et au bloc SE pour l'extraction des caractéristiques, lesquelles sont ensuite introduites dans un sous-modèle de classification afin de déterminer le stade de la maladie. Les deux approches incluent une étape de prétraitement visant à améliorer la précision et l'efficacité du processus de détection.

Mots clés: Deep Learning, Diabetic Retinopathy, Medical Imaging, Image Classification.

Abstract

Diabetic retinopathy is one of the most serious complications of diabetes, affecting eyesight and leading to blindness. Worldwide, it is one of the leading causes of blindness in the elderly. Depending on the severity of diabetic retinopathy, the ophthalmologist adapts the treatment accordingly. Early detection and regular follow-up are essential to accurately assess the progression of the disease. However, manual detection of this pathology remains very costly in terms of time and expertise. Over the past decade, deep learning has made significant advances in automating disease detection and prediction. This makes it an ideal solution for automating the early detection of diabetic retinopathy from medical images.

The aim of this thesis is to develop automated methods for the early detection of diabetic retinopathy and to accurately assess the stage of the disease, in order to help ophthalmologists offer patients personalized treatment based on their clinical condition. The first approach proposes an automatic method for the detection of diabetic retinopathy using deep learning and deformable registration. This method includes a new pre-processing step, consisting in applying deformable registration to the retina so that it occupies the entire image space. The pre-processed images are then fed into four deep learning models. Finally, an ensemble voting technique is used in the test phase to generate a single prediction based on the results of all four models. The second approach is based on a recently proposed modified capsule network, combined with a modified Inception block, to create an automatic method for predicting diabetic retinopathy. This method also incorporates an advanced pre-processing step that applies a discrete wavelet transform algorithm to the fundus image. By exploiting the hierarchical structure of the discrete wavelet transform, this approach generates three levels of decomposition of the same image, providing more detailed information from a single image. Finally, two innovative approaches for the automatic detection of diabetic retinopathy have been proposed. The first approach combines a pre-trained Vision Transformer model with a modified capsule network for feature extraction, integrated with a concatenation model to predict the stage of diabetic retinopathy. The second approach uses the ViTIncep submodel, CBAM block and SE block for feature extraction, which are then fed into a classification submodel to determine disease stage. Both approaches include a pre-processing step to improve the accuracy and efficiency of the detection process.

Keywords: Deep Learning, Diabetic Retinopathy, Medical Imaging, Image Classification.

This thesis is dedicated to my parents, who have always believed in me and encouraged me to pursue my dreams. Their love and support have been my constant source of strength. I also dedicate this work to my sisters and brother for their support.

Acknowledgements

This thesis has been made possible through the invaluable advice and support of many people to whom I would like to express my deepest gratitude. First of all, I would like to express my gratitude to My supervisor, Prof. Jamal Riffi, for his guidance, advice, comments, support, and encouragement that were instrumental throughout this thesis from start to finish. Without his help, this research would never have seen the light of day. I learned a lot from our regular meetings and discussions, his wisdom, experience, expertise, diligence, and passion for research. It has been an honor and privilege to work with him.

I would like to extend my gratitude to the jury members, Prof.Hamid Tairi as president, Prof.Benaissa Bellach, Prof.Abderazzak Taime, and Prof. Khalid El Fazazy as rapporteurs of this thesis. I also would like to thank Prof.Idriss Benatiya Andaloussi, Prof.Noura Aherrahrou, Prof.Amal Hjouji, Prof.Adnane Mahraz Mohamed, and Prof.Ali Yahyayouy for agreeing to review and evaluate my work. Special thanks to my Professors from the Faculty of Sciences Dhar El Mahraz, University of Mohamed Ben Abdallah, for the valuable feedback, advice, and encouragement they gave me during my studies. We express our gratitude to our colleagues in the Ophthalmology Department in the Medicine and Pharmacy Faculty at the University of Mohamed Ben Abdallah in Fez, Morocco, who provided invaluable experience that greatly assisted in this work.

I would like to express my deepest thanks to my parents, sisters, and brother for their support throughout my studies; nothing would have been possible without their encouragement. Special thanks to my colleagues and friends who provided valuable feedback and encouragement.

Abstract

Diabetic retinopathy is one of the most serious complications of diabetes, affecting eyesight and leading to blindness. Worldwide, it is one of the leading causes of blindness in the elderly. Depending on the severity of diabetic retinopathy, the ophthalmologist adapts the treatment accordingly. Early detection and regular follow-up are essential to accurately assess the progression of the disease. However, manual detection of this pathology remains very costly in terms of time and expertise. Over the past decade, deep learning has made significant advances in automating disease detection and prediction. This makes it an ideal solution for automating the early detection of diabetic retinopathy from medical images.

The aim of this thesis is to develop automated methods for the early detection of diabetic retinopathy and to accurately assess the stage of the disease, in order to help ophthalmologists offer patients personalized treatment based on their clinical condition. The first approach proposes an automatic method for the detection of diabetic retinopathy using deep learning and deformable registration. This method includes a new pre-processing step, consisting in applying deformable registration to the retina so that it occupies the entire image space. The pre-processed images are then fed into four deep learning models. Finally, an ensemble voting technique is used in the test phase to generate a single prediction based on the results of all four models.

The second approach is based on a recently proposed modified capsule network, combined with a modified Inception block, to create an automatic method for predicting diabetic retinopathy. This method also incorporates an advanced pre-processing step that applies a discrete wavelet transform algorithm to the fundus image. By exploiting the hierarchical structure of the discrete wavelet transform, this approach generates three levels of decomposition of the same image, providing more detailed information from a single image. Finally, two innovative approaches for the automatic detection of diabetic retinopathy have been proposed. The first approach combines a pre-trained Vision Transformer model with a modified capsule network for feature extraction, integrated with a concatenation model to predict the stage of diabetic retinopathy. The second approach uses the ViTIncep submodel, CBAM block and SE block for feature extraction, which are then fed into a classification submodel to determine disease stage. Both approaches include a pre-processing step to improve the accuracy and efficiency of the detection process.

Keywords: Deep Learning, Diabetic Retinopathy, Medical Imaging, Image Classification.

Contents

Abstract **1**

Introduction General **1**

1 Generalities on Diabetic Retinopathy and AI Applications in Medical Imaging. **8**

1.1 Introduction 8

1.2 Diabetic Retinopathy 9

1.2.1 Diabetic retinopathy symptoms 9

1.2.2 Diabetic retinopathy stages 10

1.2.3 Diabetic retinopathy detection 12

1.3 Diabetic Retinopathy in Morocco: Al-Khwarizmi Program Initiative 13

1.4 Artificial Intelligence in Healthcare 14

1.4.1 machine learning approaches 15

1.4.1.1 Supervised machine learning algorithms 15

1.4.1.2 Unsupervised machine learning algorithms 16

1.4.1.3 Semi-supervised machine learning algorithms . . 17

1.4.2 Deep learning 18

1.4.3 Deep learning domains of applications 19

1.4.4 Deep Learning Algorithms 19

1.4.4.1 Convolutional Neural Networks 20
1.4.4.2 Recurrent neural networks 20
1.4.4.3 Autoencoders 21
1.4.5 Advanced techniques algorithm 21
1.4.5.1 Attention mechanism 22
1.4.5.2 Transformer architecture 22
1.4.5.3 Generative Adversarial Networks 24
1.4.6 Deep Learning strategies 25
1.4.6.1 Transfer learning 25
1.4.6.2 Ensemble Learning 26
1.5 Preprocessing in medical image analysis 27
1.5.1 Cropping images 27
1.5.2 Noise reduction 27
1.5.3 Intensity normalization 28
1.5.4 Images registration 29
1.5.5 images segmentation 30
1.6 Conclusion 30

2 Medical Background and datasets 31

2.1 Introduction 31
2.2 Retina Fundus Image Dataset 31
2.2.1 APTOS dataset 32
2.2.2 EyeSPACS dataset 32
2.2.3 Messidor-2 dataset 33
2.2.4 DDR dataset 33
2.3 Evaluation Metrics 33
2.3.1 Accuracy 34
2.3.2 Precision 34

2.3.3 Recall . . . 34
2.3.4 F1-score . . . 35
2.3.5 Kappa Score . . . 35
2.3.6 Confusion matrix . . . 35
2.4 State of the art on diabetic retinopathy prediction . . . 36
2.4.1 Supervised and Unsupervised Machine Learning . . . 36
2.4.1.1 Support vector machines . . . 36
2.4.1.2 fuzzy classifier . . . 38
2.4.1.3 K-means cluster . . . 38
2.4.1.4 K-Nearest Neighbors . . . 39
2.4.2 Pre-processing based approaches . . . 39
2.4.3 Deep Learning based approaches . . . 42
2.4.4 Advanced deep learning based approaches . . . 48
2.5 conclusion . . . 50

3 Diabetic retinopathy prediction based on deep learning and deformable registration 52
3.1 Introduction . . . 52
3.2 Materials and Methods . . . 53
3.2.1 Deformable Registration . . . 53
3.2.2 Transfer learning . . . 54
3.2.3 Inception-V3 . . . 55
3.2.4 DenseNet-121 . . . 56
3.2.5 ResNet . . . 56
3.2.6 Xception . . . 57
3.3 Proposed approach . . . 57
3.4 Proposed Experimental Results and Discussion . . . 60
3.4.1 Dataset . . . 60

3.4.2 Evaluation Metrics 61
3.4.3 Performance Evaluation 61
3.4.3.1 Analysis of model performance during training and validation phase 61
3.4.3.2 Analysis of model performance during the testing phase 64
3.5 Conclusion 67

4 Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network 68
4.1 Introduction 68
4.2 Materials and Methods 69
4.2.1 Discrete Wavelet Transform 70
4.2.2 Inception block 70
4.2.3 Capsule Network 72
4.3 Proposed approach 72
4.3.1 Pre-processing 74
4.3.2 Inception block 75
4.3.3 Deep Hybrid model 76
4.3.4 Capsule network 77
4.3.4.1 Primary capsule: 77
4.3.4.2 DigitCaps: 78
4.3.4.3 Classification layer: 79
4.4 Experimental results and dataset 80
4.4.1 Dataset 80
4.4.2 Evaluation Metrics 81
4.4.3 Experimental results 81
4.5 Conclusion 83

5 Advanced Deep Learning Architectures for Diabetic Retinopathy Prediction. 85

5.1 Introduction . . . 85

5.2 Materials and methods . . . 86

5.2.1 Power Law Transformation . . . 86

5.2.2 Contrast Limited Adaptive Histogram Equalization . . . 87

5.2.3 Vision Transformer . . . 87

5.2.4 Capsule Network . . . 88

5.2.5 Adjust the image contrast and brightness . . . 88

5.2.6 Convolutional Block Attention Module (CBAM) . . . 89

5.2.7 Squeeze and Excitation . . . 91

5.2.8 Inception Block . . . 92

5.3 Diabetic Retinopathy Prediction Based On Vision Transformer and Modified Capsule Network . . . 92

5.3.0.1 Preprocessing . . . 93

5.3.0.2 Fine-tuning Vision Transformer . . . 94

5.3.0.3 Modified Capsule Network . . . 97

5.3.0.4 Classification model . . . 98

5.4 Hybrid Deep Learning Architecture For Diabetic Retinopathy Prediction Using Medical Images . . . 99

5.4.1 pre-processing . . . 100

5.4.2 Proposed deep Hybrid model . . . 101

5.5 Experimental results and dataset . . . 103

5.5.1 Dataset . . . 103

5.5.2 Evaluation Metrics . . . 103

5.5.3 Experimental results . . . 104

5.5.4 Diabetic Retinopathy Prediction Based On Vision Transformer and Modified Capsule Network 104
5.5.5 Hybrid Deep Learning Architecture For Diabetic Retinopathy Prediction Using Medical Images 106
5.6 Comparing finding results with the state-of-the-art 107
5.7 conclusion 109

Conclusions General and perspectives **113**

References **129**

List of Figures

1.1 Diabetic retinopathy symptoms. 10
1.2 Diabetic retinopathy stages. 11
1.3 Principe of supervised learning. 16
1.4 Principe of unsupervised learning. 17
1.5 Principe of semi-supervised learning. 18
1.6 Principe of deep learning. 19
1.7 Principe of deep learning. 20
1.8 Principe of Attention mechanism. 22
1.9 The transformer architecture 23
1.10 The Generative Adversarial Networks architecture. 24
1.11 The transfer learning technique. 25
1.12 Ensemble learning technique. 26
1.13 Cropping operation. 28

3.1 The global architecture of the proposed approach. 58
3.2 Example of an Image Dataset: (A): before applying deformable registration; (A'): after applying deformable registration. 59

3.3 (a) The accuracy curve during the training and validation phases of the Xception classifier across epochs. (b) The loss curve during the training and validation phases of the Xception classifier across epochs. 62
3.4 (a) The accuracy curve during the training and validation phases of the Inception-v3 classifier across epochs. (b) The loss curve during the training and validation phases of the Inception-v3 classifier across epochs. 63
3.5 (a) The accuracy curve during the training and validation phases of the DenseNet-121 classifier across epochs. (b) The loss curve during the training and validation phases of the DenseNet-121 classifier across epochs. 63
3.6 (a) The accuracy curve during the training and validation phases of the ResNet50 classifier across epochs. (b) The loss curve during the training and validation phases of the ResNet50 classifier across epochs. 64
3.7 The confusion matrix of the test phase after using the voting technique. 66

4.1 The pyramid hierarchy architecture of the discrete wavelet transform. 71
4.2 Illustrated the original version of the inception block. 71
4.3 An illustration of the overall structure of the proposed approach. . 73
4.4 An illustration of the overall structure of decomposition rounds used in the proposed approach. 75
4.5 A visualization showing the suggested modifed inception block proposed in our approach. 76
4.6 An illustration of the Primary capsule layer section. 78
4.7 An illustration of the digitCaps layer and classification layer sections. 80

4.8 The confusion matrix during the test process of the proposed approach.. 82

5.1 Illustration of the Convolutional Block Attention Module mechanism. 89
5.2 Illustration of the Spatial Attention Module architecture. 90
5.3 Illustration of the channel attention module architecture. 91
5.4 The Squeeze and Excitation Block. 91
5.5 An illustration of the overall structure of the proposed approach. . 93
5.6 The structure of the overall proposed pre-processing step. 95
5.7 The Multi-Head Attention Mechanism. 96
5.8 Illustration of the proposed classification model architecture. . . . 99
5.9 An illustration of the overall structure of the proposed approach. . 100
5.10 Confusion matrix of the proposed approach during the test phase across the four datasets utilized. 106
5.11 The confusion matrix of our approach in the testing phase. (A) for the APTOS dataset. (B) for the DDR dataset. 108

List of Tables

3.1 The distribution of the used dataset during the training, validation, and test processes. 61

3.2 The performance demonstrated by each model during the test process. 65

3.3 The performance obtained in the test phase after using the voting technique. 66

3.4 The comparison of the performance of our proposed approach with the state of the art during the test phase. 67

4.1 Overview of the deep hybrid model. 77

4.2 The performance demonstrated by each model during the test process. 80

4.3 The performance obtained during the test phase. 81

4.4 The comparison of the performance of the current study with other relevant earlier research during the test phase. 83

5.1 The performance of the proposed approach during the test phase using Ac, Rec, Pr, Ka, and F1 denotes accuracy, recall, precision, Kappa score, and F1-score. On the four datasets used. (unit:%). . 105

5.2 The results obtained for the grading of DR on the APTOS and DDR datasets in the testing phase, Ac, Ma Pr, Ma Rec, Ka, and F1 denote accuracy, macro precision, macro recall, Kappa score, and F1-score., respectively. (unit:%) 107
5.3 The comparison of our proposed approach with relevant earlier research (unit:%). 108

Introduction General

Diabetes is a chronic condition characterized by elevated blood sugar levels. If left untreated, it can lead to serious health complications. According to the International Diabetes Federation (IDF), over 537 million adults aged 20-79 years had diabetes in 2021, and this number is expected to rise to 783 million by 2045. One of the most severe complications of diabetes is diabetic retinopathy. As noted by Gargeya et al. [1], between 40% to 45% of diabetes patients may develop diabetic retinopathy (DR) at some point in their lives. Diabetic retinopathy, also referred to as diabetic eye disease, is a prevalent diabetes complication resulting from vascular damage to the retina caused by elevated blood sugar levels [2]. As diabetic retinopathy progresses through its stages—from mild NPDR to moderate NPDR, then severe NPDR, and finally to proliferative DR—new blood vessels may become blocked, leak on the retinal surface, or grow abnormally. This progression can ultimately lead to blindness. Therefore, regular monitoring by healthcare professionals is essential to prevent the progression of diabetic retinopathy and manage it effectively through appropriate treatment. Relying only on manual methods is insufficient, as they are time-consuming, require highly skilled clinicians, and the disease often presents no symptoms in its early stages, making early detection challenging. Which make. Hence, there is a significant need for automatic detection methods to improve the efficiency and accuracy of identifying and managing this condition in its early stages. These methods can also play a

crucial role in assisting doctors in the prediction and diagnosis process, ensuring timely and precise interventions. In this regard, Morocco's Ministry of Education and the Digital Development Agency (ADD) launched the Al-Khawarizmi program in March 2019 in collaboration with the National Center for Scientific and Technical Research (CNRST) to finance practical projects in the fields of artificial intelligence (AI) and big data to improve healthcare in Morocco. Our project is one of the projects financed by this program, which is titled " Towards a Democratization of the Moroccan Healthcare System: Development of a Predictive System for Early Detection of Diabetic Retinopathy based on Deep Learning." This project aims to propose an automatic system for the early detection of diabetic retinopathy using deep learning techniques. In the last decade, deep learning has demonstrated outstanding results in analyzing medical images. It has been successfully applied in predicting diseases, detecting and segmenting tumors, and classifying medical images into various categories or stages of disease severity. This makes deep learning a highly promising solution for diabetic retinopathy detection, enabling early diagnosis and accurate determination of the disease's stage. In fact, several automatic approaches for diabetic retinopathy prediction based on deep learning have been suggested. Numerous studies leverage advanced deep learning architectures such as vision transformers [3], capsule networks [4], and attention mechanisms [5][6]. Other approaches rely on traditional deep learning architectures [7][1], and are enhanced with techniques like transfer learning [8][9] or ensemble learning [10]. Additionally, some research focuses on the pre-processing process [11][12] to improve the quality of retina fundus images before feeding them into the deep learning model. Overall, these diverse strategies underscore the potential of deep learning to revolutionize diabetic retinopathy detection and improve early diagnosis and treatment. However, further improvements are still needed, especially for multiclass issues of classifi-

cation tasks and in enhancing early detection to better support patients in their treatment process.

Objectives and Contributions

The primary contribution of this thesis is the development of automatic diabetic retinopathy prediction methods leveraging advanced deep learning architectures. These approaches aim to enhance the performance of disease detection and provide valuable assistance to doctors in the diagnostic process. Our major contributions are summarized as follows:

1. First, we introduced an automatic diabetic retinopathy prediction approach utilizing deep learning combined with deformable registration techniques [13]. This method involves deformable registration of retinal images to ensure that the entire image area is utilized, preserving crucial information at the edges and minimizing issues such as black spaces that can interrupt the training process. Unlike cropping methods, which may lead to information loss or leaving black space around the retina, which affect the training process, this strategy ensures the complete retention of relevant image details. The transformed images are then fed into an ensemble of four pre-trained deep learning models (Xception, Inception-V3, ResNet-50, and DenseNet-121), each trained separately. We modified the final classification layers of these models to align with our specific task and employed transfer learning to avoid starting the training process from scratch. For the test phase, a majority voting technique was used, where the final decision is based on the most frequently occurring output among the models or the output admitted by all the models.

2. Second, we proposed an automatic diabetic retinopathy prediction approach that is based on wavelet decomposition in addition to an improved capsule network [14]. This method utilizes discrete wavelet decomposition to exploit the

hierarchical pyramid structure of retinal fundus images, allowing the image to be decomposed into three levels, each providing additional detailed information compared to a single image. The decomposed images are then processed through a classification model starting with a modified inception block. Then, the features extracted in this block are further enhanced by a deep hybrid model, which plays a crucial role in connecting the modified inception block with the proposed improved capsule network. Finally, these enhanced features are fed into the modified capsule network to accurately predict the diabetic retinopathy severity levels present in the image.

3. Third, we presented a new automatic method for diagnosing the severity of diabetic retinopathy utilizing a combination of vision transformers, modified capsule networks, and a novel sequence of computer vision techniques [15]. The pre-processing step is designed to enhance the retinal fundus image using the Power Law Transformation to produce the first enhanced image, subsequently by Contrast Limiting Adaptive Histogram Equalization to create a second enhanced image. These enhanced images are then input into a hybrid deep learning model comprising a pre-trained vision transformer, an improved capsule network, and a classification model. The first enhanced image is processed by the vision transformer, while the second is processed by the improved capsule network. The feature maps generated by these two models are then fused using the classification model to precisely forecast the severity stage of DR present in the image.

4. Fourth, we developed an innovative automatic approach for diabetic retinopathy prediction utilizing a hybrid deep learning model that incorporates Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation (SE), and ViT-Incep sub-model, alongside a pre-processing step. Our approach begins with a pre-processing stage designed to adjust the contrast and brightness of retinal fundus images, thereby improving their quality and reducing noise. The deep hybrid

model employs the ViTIncep sub-model to generate meaningful feature maps, which are then refined through the CBAM and SE blocks. The CBAM block enhances the feature maps by focusing on critical regions within the retina image, while the SE block further refines these features using its squeeze-and-excitation technique. Finally, a fusion method and a classification sub-model are integrated into the deep hybrid model to facilitate accurate diabetic retinopathy grading.

Overview of the Thesis

The structure of the thesis outline is as follows:

Chapter 1: This chapter provides basic information regarding the stages and symptoms of diabetic retinopathy. Next, some Moroccan statistics are given. Followed by the most popular pre-processing operation in medical imaging, as well as a few artificial intelligence architectures in the healthcare area.

Chapter 2: This chapter introduces widely used retina fundus image datasets and common evaluation metrics for assessing the performance of automatic approaches. Additionally, it provides a literature review of existing automatic methods for diabetic retinopathy prediction based on artificial intelligence.

Chapter 3: This chapter explains in more detail the proposed approach for diabetic retinopathy prediction based on deep learning and deformable registration. Besides, it presents the result achieved. In addition, a comparison between the proposed approaches and the SOTA methods is presented in this chapter.

Chapter 4: This chapter presents the proposed approach based on the capsule network and discrete wavelet decomposition for diabetic retinopathy prediction. Besides, it presents the result achieved. In addition, a comparison between the proposed approaches and the SOTA methods is presented in this chapter.

Chapter 5: This chapter introduces two approaches for diabetic retinopathy prediction. First, diabetic retinopathy prediction is based on a vision transformer

and a modified capsule network. Second, a hybrid deep learning architecture for diabetic retinopathy prediction using image medical. In addition, a comparison between the proposed approaches and the SOTA methods is presented in this chapter.

Publication

Journal and conference papers published and under review during this thesis:

• Oulhadj, Mohammed, Jamal Riffi, Khodriss Chaimae, Adnane Mohamed Mahraz, Bennis Ahmed, Ali Yahyaouy, Chraibi Fouad, Abdellaoui Meriem, Benatiya Andaloussi Idriss, and Hamid Tairi. "Diabetic retinopathy prediction based on deep learning and deformable registration." Multimedia Tools and Applications 81, no. 20 (2022): 28709-28727.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ahmed Bennis, Ali Yahyaouy, Fouad Chraibi, Meriem Abdellaoui, Idriss Benatiya Andaloussi, and Hamid Tairi. "Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network." Journal of Digital Imaging 36, no. 4 (2023): 1739-1751.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ali Yahyaouy, Meriem Abdellaoui, Idriss Benatiya Andaloussi, and Hamid Tairi. "Diabetic retinopathy prediction based on vision transformer and modified capsule network." Computers in Biology and Medicine 175 (2024): 108523.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ali Yahyaouy, Meriem Abdellaoui, Idriss Benatiya Andaloussi, and Hamid Tairi. "Hybrid deep learning architecture for diabetic retinopathy prediction using image medical". **(under review)**.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ahmed Bennis, Ali Yahyaouy, Fouad Chraibi, Meriem Abdellaoui, Idriss

Benatiya Andsaloussi, and Hamid Tairi. "Diabetic Retinopathy Prediction Based on Transfer Learning and Ensemble Voting." In International Conference on Digital Technologies and Applications, pp. 929-937. Cham: Springer Nature Switzerland, 2023.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ali Yahyaouy, Meriem Abdellaoui, Idriss Benatiya Andaloussi, and Hamid Tairi. "Diabetic Retinopathy Prediction Based on a Hybrid Deep Learning Approach." In 2024 International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1-5. IEEE, 2024.

• Oulhadj, Mohammed, Jamal Riffi, Chaimae Khodriss, Adnane Mohamed Mahraz, Ali Yahyaouy, Meriem Abdellaoui, Idriss Benatiya Andaloussi, and Hamid Tairi. "Enhancement of Diabetic Retinopathy Prediction-Based Capsule Network and DenseNet-121 Utilizing Wavelet Decomposition". In International Conference on Mediterranean Smart Cities Conference: MSCC2024. Presented in Tetouan, Morocco on May 3, 2024 **(Not published yet)**.

Chapter 1

Generalities on Diabetic Retinopathy and AI Applications in Medical Imaging.

1.1 Introduction

In this chapter, we explore the various aspects of diabetic retinopathy and its detection, particularly focusing on the advancements in artificial intelligence (AI) and their application in healthcare. We begin with an overview of diabetic retinopathy, outlining its causes, symptoms, and stages. This is followed by a discussion on the prevalence and management of diabetic retinopathy in Morocco, highlighting the AL-KHWARIZMI program's contributions. Next, we delve into the role of artificial intelligence in healthcare, emphasizing its potential to revolutionize medical diagnostics and treatment. Finally, we examine the importance of preprocessing in medical image analysis, which is crucial for enhancing the accuracy of AI models in detecting diabetic retinopathy.

1.2 Diabetic Retinopathy

Diabetic retinopathy, also recognized as diabetic eye disease, is a frequent diabetes condition marked by retinal vascular damage, due to high blood sugar levels [2]. This condition arises from blood leakage on the retinal surface, resulting from the fragility of blood vessels, and can ultimately lead to blindness [16]. Diabetic retinopathy is considered a long-term condition seen in diabetic patients; Its progression increases as long as the patient has diabetes.

Diabetic retinopathy is one of the main causes of loss of sight in people of working age [17]. Besides, this pathology comes with no symptoms in its early stages or may have the same minor vision problems, but as the disease gets worse, it may cause loss of vision [18]. In addition, the appropriate treatment of this disease is based on early detection, which performs an essential role in preventing the development of that disease and preventing many individuals from losing their vision.

1.2.1 Diabetic retinopathy symptoms

Microaneurysms: Microaneurysms (MAs) are the earliest diabetic retinopathy signs visible by clinicals in retina fundus images (Fig.1.1). They usually appear as tiny bulges of blood and look like small red dots on the retina with sharp edges. This symptom causes damage to the blood vessels to leak into the retina. [19].

Hard exudates: Hard exudates (HE) are lipids that appear as a yellowish color in the retina and appear as a result of leakage from retinal blood vessels (Fig.1.1). The latter symptoms appear as spots and may be in different shapes and sizes [19].

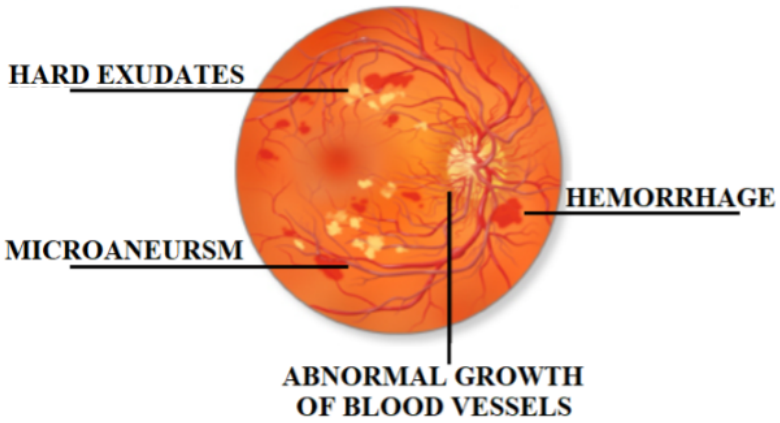

Figure 1.1: Diabetic retinopathy symptoms.

Hemorrhages: As the disease progresses, new symptoms appear on the surface of the retina as hemorrhages (Fig.1.1). Hemorrhages in diabetic retinopathy are bleeding events that occur within the retina due to damage and elevated fragility of the retinal blood vessels. The latter symptoms can take various forms.

1.2.2 Diabetic retinopathy stages

Diabetic retinopathy progresses through several stages; each stage represents the degree of danger and has specific retinal changes. The stages of diabetic retinopathy are generally categorized as follows (Fig.1.2):

Mild Non-Proliferative Diabetic Retinopathy (NPDR): is the earliest stage of diabetic retinopathy; in this stage, patients typically do not experience noticeable vision changes. This stage is outlined by the beginning of the presence of microaneurysms (MAs), which are the earliest visible signs of damage to the retinal blood vessels.

Moderate Non-Proliferative Diabetic Retinopathy (NPDR): is the second stage of diabetic retinopathy. At the moment, some retinal vessels may swell, and there may appear some new symptoms including hard exudates, hemorrhages, and cotton wool spots, which may cause physical changes to the retina.

Severe Non-Proliferative Diabetic Retinopathy (NPDR): Severe NPDR is an advanced stage of non-proliferative diabetic retinopathy. At this stage, patients start to notice more significant visual disturbances, such as dark spots, and the retina reveals the extensive presence of microaneurysms, hemorrhages, and hard exudates. The development of DR over this stage increases the risk of lost vision, especially if the condition progresses to the next level.

Proliferative Diabetic Retinopathy (PDR): is the most serious stage of diabetic retinopathy. It is defined by the proliferation of new abnormal blood vessels on the surface of the optic disc and retina. These new blood vessels are fragile and prone to bleeding, which may result in complications such as vision loss. The patients may see spots or floaters in their vision due to bleeding from these new blood vessels.

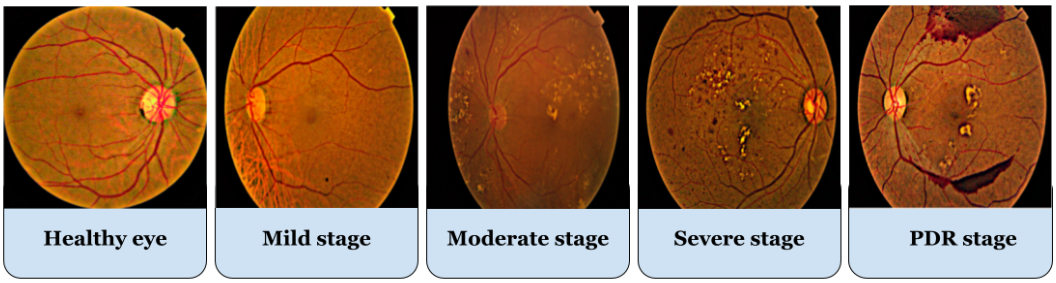

Figure 1.2: Diabetic retinopathy stages.

1.2.3 Diabetic retinopathy detection

Early detection of diabetic retinopathy and determining the exact stage a patient is in is essential to providing proper treatment to avoid loss of sight. Nonetheless, the early diagnosis of diabetic retinopathy remains a challenging process requiring skilled human retina fundus image interpretation for vigilant patient follow-up.

Expert ophthalmologists use the manual screening method to detect diabetic retinopathy. They visually examine the retina fundus image to identify abnormalities such as diabetic retinopathy symptoms. This manual procedure is time-consuming and may be inefficient for a large amount of patients. In addition, it requires professionals, doctors, and skilled ophthalmologists [20]. The proposed solution to address this challenge is to propose automatic approaches and methods to detect the DR based on deep learning and machine learning techniques.

In the last decade, deep and machine learning methods have shown great results in detecting and classifying medical images, especially diabetic retinopathy. In the context of diabetic retinopathy, several methods have been proposed for automatic detection of DR, which provide multiple benefits, including quicker diagnosis compared to manual examinations by ophthalmologists. They efficiently manage large volumes of retinal images, facilitating the early detection of subtle signs of DR. This capability not only reduces the necessity for frequent manual assessments but also ensures timely management of the disease, which may avoid eyesight loss.

1.3 Diabetic Retinopathy in Morocco: Al-Khwarizmi Program Initiative

According to the findings of the 10th edition of the International Diabetes Federation (IDF) [21], diabetes is one of the fastest growing global health emergencies around the world. They estimated that 537 million people having diabetes in 2021, and by 2030, the number may increase to 643 million people around the world. As pointed out by [22], almost all patients with diabetes are at high risk of developing diabetic retinopathy. In the same regard, Morocco is one of the countries that suffers from this disease. According to [23], diabetic retinopathy is a serious and frequent complication, and it is one of the major causes of blindness. It threatens the health of at least one million individuals [23].

As indicated in a study proposed by Hammoudi et al.[24], the study occupied 244 diabetes patients, including 122 patients suffering from diabetic retinopathy, of whom 122 were healthy, in Oujda during the period between December 2018 and July 2019. This study showed that 57.4% of these patients were suffering from the worst stage of diabetic retinopathy, which is called proliferative diabetic retinopathy (PDR). Besides, non-proliferative diabetic retinopathy (NPDR) represents 42.6%, which is classified as 8.2% having the mild stage, 26.2% suffering from the moderate stage, and the rest of 8.2% having the severe stage.

Another study [25] in Tetouan, Morocco, involved 500 patients who consulted the same ophthalmologist between June 2022 and December 2022. The patients have an average age of 62 years and an average duration of diabetes of 11.5 years, divided into 217 men and 283 women. They find that between every 3 patients, there is one who is unaware that diabetes is a blinding disease, and just 1 of 8 patients has an annual ophthalmological check-up. Besides, the limited number of available centers for diabetic retinopathy early detection. For all this reason,

utilizing and using automatic methods became essential, especially in remote areas.

In regard to helping people detect the disease in its early stages and predict its exact stage, Morocco's Ministry of Education and the Digital Development Agency (ADD) launched the Al Khawarizmi program in March 2019 in collaboration with the National Center for Scientific and Technical Research (CNRST) to finance practical projects in the fields of artificial intelligence (AI) and big data. To improve and help accelerate Morocco's digital transition in various domains, such as the healthcare area. In this regard, the AL-Khwarizmi program finances our project, "Towards a Democratization of the Moroccan Healthcare System: Development of a Predictive System for Early Detection of Diabetic Retinopathy based on Deep Learning," to help patients suffering from diabetic retinopathy in the process of the early detection and grading of DR.

1.4 Artificial Intelligence in Healthcare

Artificial intelligence (AI) has been widely used in the healthcare field in the last decade, and it has shown great results in tasks such as medical image analysis, segmenting, and detection by developing automatic approaches using machine learning and deep learning algorithms. Automatic approaches that are capable of processing the input image as a retina fundus image in order to detect the diabetic retinopathy symptoms. Besides, classifying these medical images into the severity stage of DR. In addition, detecting this pathology in its early stages using a retinal fundus image or optical coherence tomography (OCT) image remains a hard task even for the expert ophthalmologist.

1.4.1 machine learning approaches

Over decades, machine learning (ML) has been used to develop approaches to analyzing the medical image. by learning features from input data and performing an outcome prediction, facilitating early diagnosis and detection. The ML algorithm can be divided into four main categories: supervised, semi-supervised, unsupervised learning, and reinforcement learning (RL).

1.4.1.1 Supervised machine learning algorithms

Supervised machine learning (SML) is a type of machine learning that relies on labeled training data to make decisions based on this labeled data. The principal of this method is based on training a model using an input image paired with its label. Then, during the training process, the model learns to extract features from the input image. Once the model is trained, we use this trained model to make new predictions about the unlabeled test set. SML algorithms are indeed utilized to address both classification and regression problems (Fig.1.3).

Classification problem: Aid in categorizing data into discrete class labels or categorical outputs. As in diabetic retinopathy detection, it can be used to diagnose this pathology, classifying its severity.

Regression problem: It is employed to predict continuous values based on input variables, which can be used in grading diabetic retinopathy by predicting values instead of directly predicting the class and then using thresholding to predict the exact class.

Supervised machine learning (SML) is based on several algorithms, such as the support vector machine algorithm, the random forest algorithm, the decision tree algorithm, the logistic regression algorithm, and the multi-layer perception.

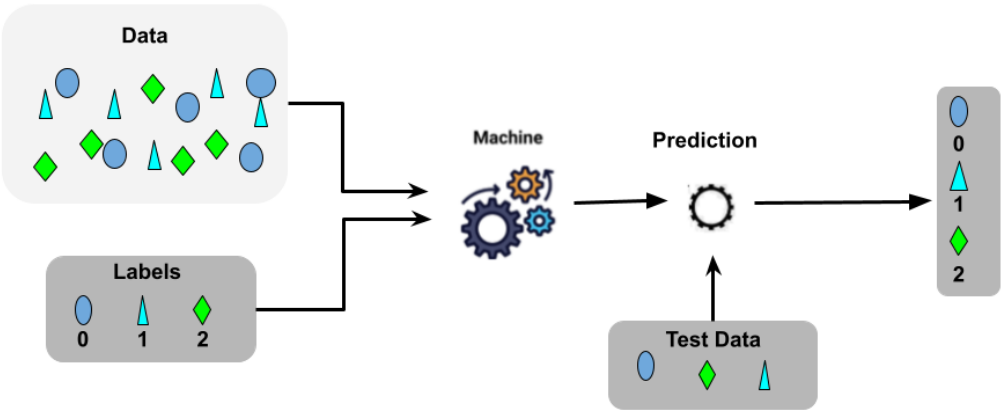

Figure 1.3: Principe of supervised learning.

1.4.1.2 Unsupervised machine learning algorithms

Unsupervised machine learning (UML), also known as unsupervised learning, consists of analyzing input data based on an unlabeled dataset. It consists of training a machine learning model without pairing the input image with its labels; instead, they learn by identifying the natural structure within the data. The model learns by itself and predicts the output based on the training data. The UML algorithms are indeed utilized to address both clustering and association problems.

Clustering problems a key aspect of unsupervised machine learning, involves grouping objects based on similarities with the aim of placing similar objects within the same cluster. This technique is particularly valuable in medical image analysis, where it facilitates the identification of patterns and structures within the data, thereby contributing to the diagnosis, treatment, and comprehension of various medical conditions.

Association problem Association mining plays a critical role in medical image analysis by uncovering meaningful relationships and patterns within large datasets of medical images. These techniques focus on identifying dependencies and mapping variables to optimize outcomes effectively. Unsupervised machine learning (UML) is based on several algorithms, such as the K-means algorithm, Bayesian networks, probabilistic graphical models, and the and the DBSCAN algorithm (Fig.1.4).

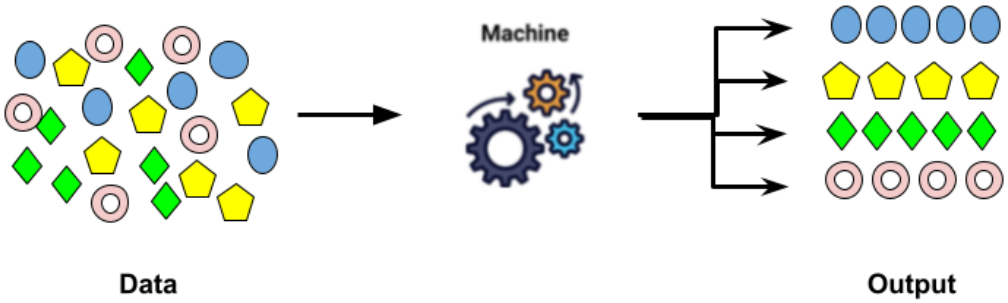

Figure 1.4: Principe of unsupervised learning.

1.4.1.3 Semi-supervised machine learning algorithms

Semi-supervised machine learning is a combination of supervised and unsupervised machine learning. It aims to combine labeled and unlabeled datasets during the training process. This approach combines a small amount of labeled data with a large amount of unlabeled data. By utilizing both types of data, semi-supervised learning can achieve better performance compared to purely supervised or unsupervised approaches, especially when faced with the limitation of an insufficiently labeled dataset. Besides, it can be used for labeling unlabeled data (Fig.1.5).

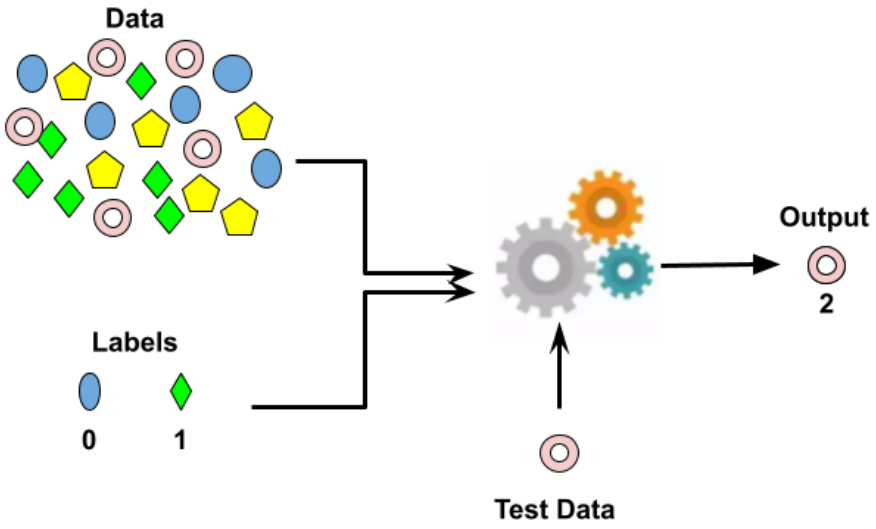

Figure 1.5: Principe of semi-supervised learning.

1.4.2 Deep learning

Deep learning is a subset of machine learning that has come to solve many problems that traditional machine learning was unable to face. Problems such as recognizing objects in images, detecting and localizing multiple objects within images or video frames, and effectively segmenting images into different regions for detailed analysis.

Deep learning is based on a neural network mechanism to directly extract features from large amounts of data to generate a trained model. This trained model can be used to make predictions or classifications on a test dataset that wasn't used in the training process. The Fig.1.6 represents an example of deep learning architecture

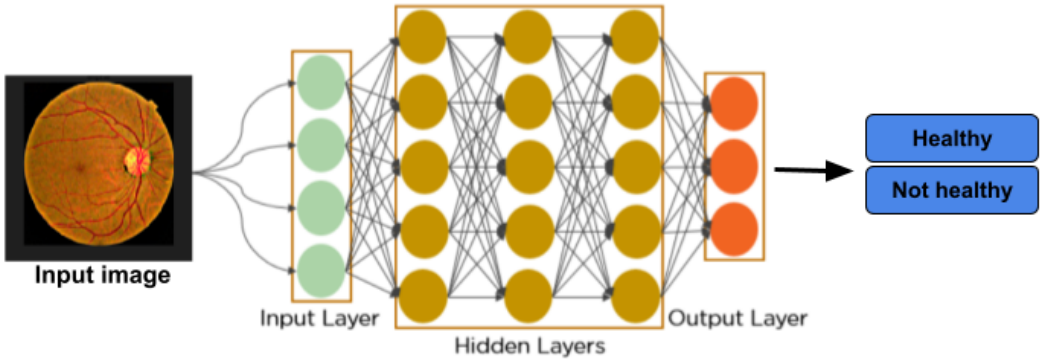

Figure 1.6: Principe of deep learning.

1.4.3 Deep learning domains of applications

Deep learning, as a subset of artificial intelligence, is useful in a wide range of contexts due to its capability to learn from complex patterns of data. Over the past decade, deep learning has shown great performance in computer vision, where many studies have proposed it in the fields of image classification, object detection, and image segmentation. In the same vein, DL has demonstrated excellent results in natural language processing (NLP). Several studies have been suggested for the purposes of speech recognition, machine translation, and sentiment analysis. Moreover, DL proved excellent performance in the healthcare domain; numerous studies have been suggested with the aim of identifying diseases and tumors, aiding in decision-making. In a similar way, DL has proven exceptional achievement in the smart city area by enabling vehicles to perceive their environment and navigate autonomously (self-driving) and enhancing robotic perception, control, and decision-making capabilities (as robots and drones).

1.4.4 Deep Learning Algorithms

In this section, we will provide a brief description of some of the most frequently employed and popular deep learning algorithms and techniques.

1.4.4.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) are a type of deep learning algorithm that was proposed for the first time to process inputs as images. It soon began to be used in a number of different aspects, including videos, speech, text,...

The CNNs are based on several operations, such as convolution, which is a mathematical operation that consists of applying several filters to the input to extract feature maps. In addition, pooling operations reduce the dimensionality and computation of the feature maps. Moreover, many types of activation functions can be applied, such as the ReLU function. Furthermore, combining feature maps using a fully connected layer to produce the final classification or regression result, as illustrated in Fig.1.7.

During the last decade, CNNs have been widely used in several tasks, such as image classification, object detection, image segmentation, face recognition, medical image analysis, video analysis, and natural language processing (NLP).

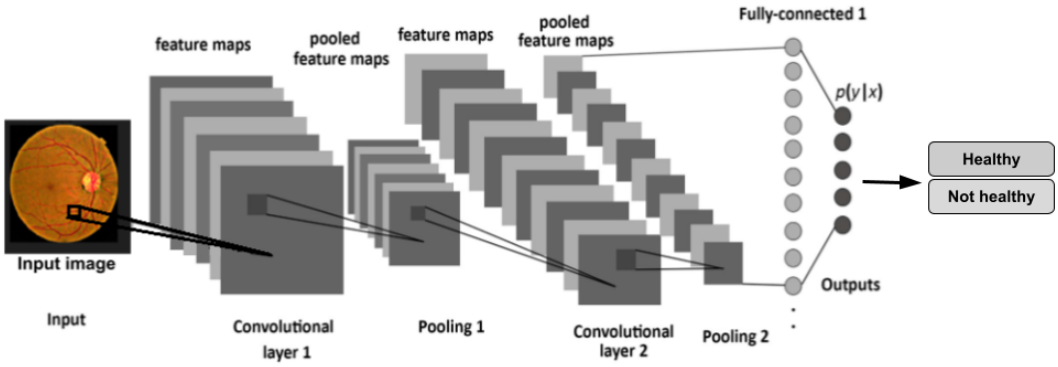

Figure 1.7: Principe of deep learning.

1.4.4.2 Recurrent neural networks

Recurrent Neural Networks (RNNs) are a type of deep learning algorithm that are widely used in the context of sequential data due to their suitability for

tasks involving sequences and their reliance on recurrent connections to identify sequential features and patterns in data. Which make them suitable for tasks such as time series forecasting, natural language processing, and speech recognition.

Recurrent Neural Networks (RNNs) consist of several key components as hidden states that capture information from previous inputs and carry it forward. The recurrent layer processes each element of the sequence sequentially. The output layer generates predictions or classifications for each time step. These components together make RNNs powerful for tasks such as natural language processing, time series forecasting, and speech recognition.

1.4.4.3 Autoencoders

Autoencoders are a type of deep learning algorithm that is widely used in medical imaging. The main objective of this approach involves compressing the input into a laten-space representation during the encoder operation and then reconstructing the output from this representation during the decoder operation.

In medical imaging, autoencoders are being extensively used for various tasks, such as noise encoding to remove noise from input images. Dimensionality reduction to reduces the number of features in input data while preserving important information. Generative modeling is used to create new data that can be used in training DL models or as augmented data.

1.4.5 Advanced techniques algorithm

In recent years, many new and advanced deep-learning architectures have been proposed, such as Capsule networks, Vision transformers, Attention mechanisms, transformers,... Those new architectures provide its abilities and strength in healthcare areas, by enhancing and proving tasks such as classification, segmentation, and detection of tumors and diseases. In this section, we provide a brief

description of some of the most frequently employed and popular advanced deep-learning algorithms and techniques.

1.4.5.1 Attention mechanism

The attention mechanism is a new deep learning concept that gives the model the possibility to focus on a specific part of the input image and give more importance to the information in those parts during the training process. This mechanism has become an essential substance in many advanced deep-learning approaches, particularly in computer vision, due to its strength in handling large images while objects of interest, such as symptoms and tumors, are visible in small parts of the overall image.

The main architecture of the attention mechanism is illustrated in Fig.1.8.

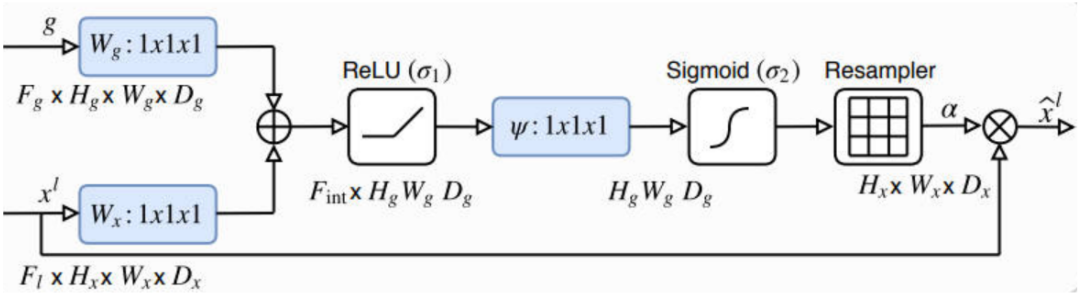

Figure 1.8: Principe of Attention mechanism.

1.4.5.2 Transformer architecture

The transformer model in a recent deep learning architecture was introduced for the first time in the paper "Attention is All You Need" by Vaswani et al. in 2017. Due to its performance in several areas and multi-task uses, it became the foundation for many state-of-the-art models in computer vision.

The transformer core is based on several components, including input embeddings, which convert the inputs into vectors of fixed size and convert the image into patches. Besides, the position encodings, which add position information to the input patches, In addition, the self-attention mechanism, which uses the attention mechanism for each pair of patches and position, focuses on relevant parts of the input to make predictions. Also, the multi-head attention, which runs multiple self-attention mechanisms in parallel, allows the model to capture different aspects of relationships within the data.

The main architecture of the transformer architecture is illustrated in Fig.1.9 as presented in the original paper.

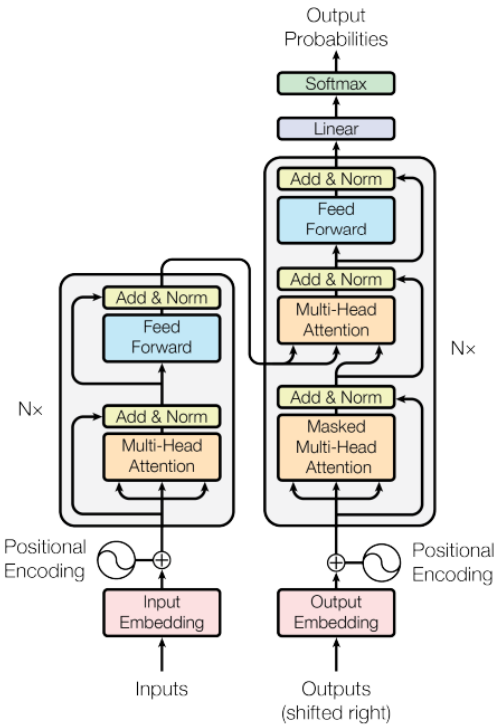

Figure 1.9: The transformer architecture .

1.4.5.3 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a type of deep learning model. GANs consist of two neural networks: the generator and the discriminator, which are trained simultaneously in a competitive manner. The GANs model can generate and create new simples, such as images, videos, texts, or audios, based on its inputs. In this task, the generator and the discriminator work in a collaborative way. Where the generator takes random noise as input and generates synthetic data that is similar to real data, Its goal is to produce outputs that are identical to the original data. While the discriminator acts as a classifier, distinguishing between real data samples and those generated by the generator, It learns to classify whether an input data sample is real or fake (Fig.1.10).

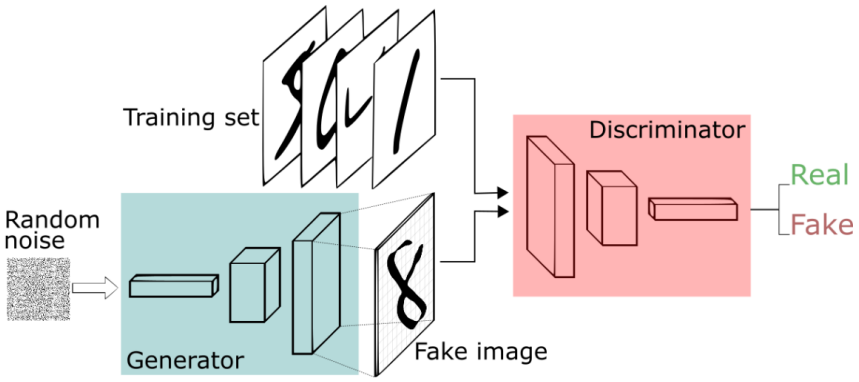

Figure 1.10: The Generative Adversarial Networks architecture.

Generative adversarial networks (GANs) have a wide range of applications across various domains due to their ability to generate realistic data. GANs are widely used in important domains, such as data augmentation to generate additional training data; text-to-image to create images from text; deepfake to create new fake images and videos; medical imaging to generate synthetic medical images for training and augmenting datasets and improving the quality of medical images; and art, where they generate new artworks and styles.

1.4.6 Deep Learning strategies

Even with the strong performance of deep learning, incorporating new strategies and operations can often further enhance the performance of deep learning models. Especially when facing challenges such as resource limitations, performance constraints, or insufficient training data. Several strategies can be employed in this regard, such as:

1.4.6.1 Transfer learning

Transfer learning is a powerful technique in deep learning that involves utilizing pre-trained models on other datasets and for other tasks to reuse them on new datasets and new tasks. in order to improve the performance of a model on a new task and avoid training the models from scratch. This technique is very useful when training a model on an insufficient amount of training data or with limited resources to train high model parameters. The transfer learning technique is illustrated in Fig.1.11.

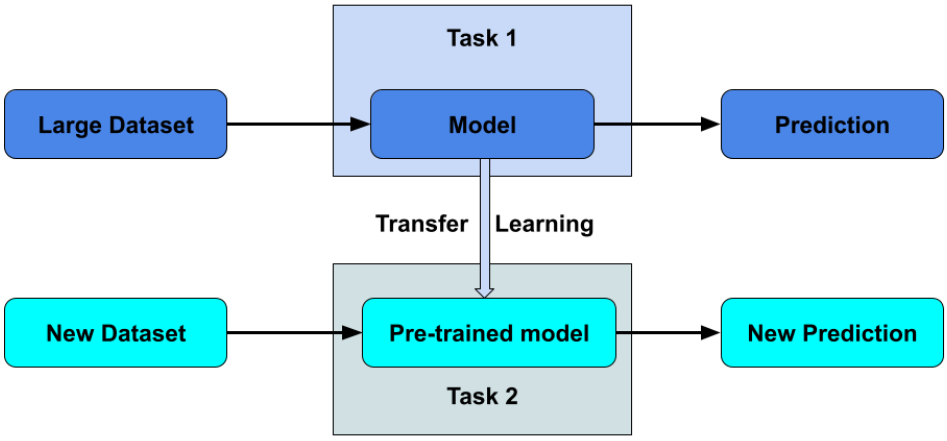

Figure 1.11: The transfer learning technique.

1.4.6.2 Ensemble Learning

Ensemble learning (EL) is a powerful technique that involves merging the predictions of multiple models to improve overall performance in comparison to individual models. The fundamental idea is to leverage the strengths of diverse models and offset their weaknesses (Fig.1.12). Ensemble learning has several benefits, such as the improvement of accuracy, where an ensemble method can achieve a higher accuracy score than a single mode. Besides, it increases the robustness of the noise in the data by reducing the risk of overfitting. However, EL faces some disadvantages due to the increased complexity of the model. Its capacity to combine the strengths of several models makes this technique widely used in several domains, such as healthcare, for tasks such as classification, regression, and detection.

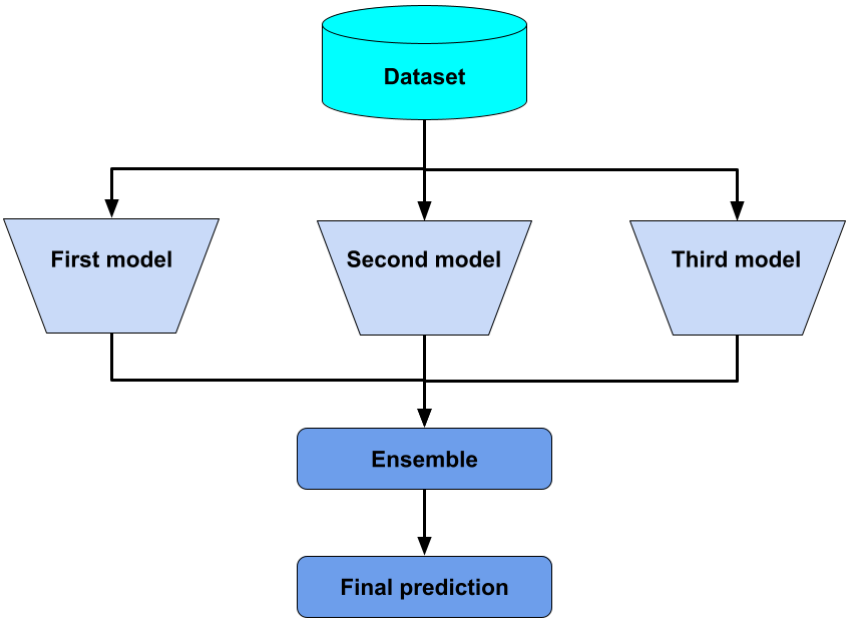

Figure 1.12: Ensemble learning technique.

1.5 Preprocessing in medical image analysis

The pre-processing step plays an essential role in medical image classification and disease detection. It involves a series of techniques to enhance the quality of the images and extract relevant features, ultimately improving the performance of machine learning algorithms and deep learning models. In diabetic retinopathy detection-based retina fundus images, those images most often come with poor quality and may contain noise or be out of focus, which makes the pre-processing step an essential step before feeding the image into the deep learning model. Here are some commonly used pre-processing techniques.

1.5.1 Cropping images

Image cropping is a process that allows us to remove unimportant backgrounds or irrelevant parts of an image. This process has several benefits, such as enhancing model performance by focusing on just the important parts of the image. Additionally, it reduces computational requirements by ignoring unimportant parts, ultimately speeding up the classification process. By reducing the size of the images, cropping decreases the amount of data the model needs to process, thereby lowering computational costs and memory usage and speeding up training. Furthermore, retina fundus images often come with background space around the retina, making the removal of this space an important task for the diabetic retinopathy (DR) detection process (Fig.1.13).

1.5.2 Noise reduction

In image classification, noise reduction is an essential preprocessing step that guarantees consistency across datasets, accelerates convergence, boosts model accuracy, and improves the quality of the training data. This is particularly im-

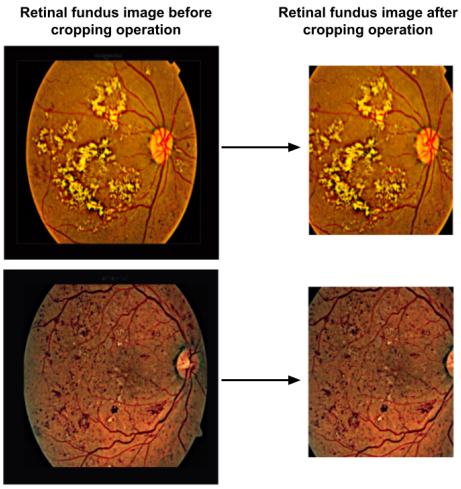

Figure 1.13: Cropping operation.

portant for diabetic retinopathy prediction, as retina fundus images often suffer from noise and poor quality due to the conditions in which they were taken. Common noise reduction techniques include Gaussian filtering, which helps reduce high-frequency noise by smoothing the image. In addition to applying the median filter, which replaces each pixel with the median values of its neighbors, decomposing the image into different levels reduces the noise using the wavelet transformation.

1.5.3 Intensity normalization

One of the most commonly used pre-processing operations in the field of deep and machine learning is the intensity normalization technique. This technique helps standardize the pixel intensity values across images. Through this method, we can improve the performance of image classification tasks and enhance the model's performance. For some of the most commonly used techniques for intensity normalization, we find min-max normalization, which ensures that all pixels's values scale to a specific range, usually between 0 and 1, using the equation (Eq1.1).

$$I_{value} = \frac{I_{value} - Min}{Max - Min} \tag{1.1}$$

Furthermore, Z-score normalization also scales the pixels into a specific range, where it consists to has a mean of 0 and a standard deviation of 1, using the equation (Eq1.2).

$$Z - score = \frac{I_{value} - \mu}{\sigma} \tag{1.2}$$

Besides, histogram equalization is also a normalization method that consists of adjusting the contrast of the image by redistributing pixel intensity values, creating a uniform histogram. Moreover, percentile normalization is another technique for normalization, which consists of normalizing the data into a specific target percentile range.

1.5.4 Images registration

Image registration is a preprocessing technique that aligns two or more images into a common coordinate system. The registration technique has several benefits, such as aligning images that have been taken at different times and from different viewpoints. Besides, it's a crucial technique to follow the development of tumors or diseases in images. Image registration plays an essential role by enabling precise alignment of retinal fundus images taken over different visits. This technique is crucial for tracking the progression of the disease in a patient. Besides, by aligning images, clinicians can precisely monitor the size, shape, and location of DR symptoms such as blood vessels, microaneurysms, hemorrhages, and exudates., making it easier to assess disease progression and make informed treatment decisions.

1.5.5 images segmentation

Image segmentation is a crucial preprocessing step in the medical imaging area, specifically in disease detection and classification. It involves dividing the input image into multiple regions based on several characteristics, such as texture, intensity, and color. In the context of diabetic retinopathy (DR) detection, segmentation focuses on isolating and identifying critical features and DR symptoms within retinal fundus images, such as blood vessels, microaneurysms, hemorrhages, and exudates. This makes this technique crucial to tracking the progression of the disease in a patient.

1.6 Conclusion

Diabetic retinopathy stands out as one of the leading causes of blindness among the global working-age population. Several automatic methods have been proposed to assist doctors and ophthalmologists in diagnosing and detecting this disease in order to enhance the manual process, especially since its treatment is based on early detection, which is still a hard task even for experts, and predicting the exact level of DR that the patient suffers from.

Chapter 2

Medical Background and datasets

2.1 Introduction

In this chapter, we explore various studies on the state-of-the-art techniques for grading diabetic retinopathy using computer vision techniques. These studies focus on enhancing their approaches by leveraging powerful machine learning techniques, introducing new pre-processing techniques to improve the quality of retinal fundus images, and employing deep learning models, including hybrid models and advanced deep learning models, to achieve more accurate and reliable diabetic retinopathy grading. Besides, we represent the famous datasets used in this field of DR detection.

2.2 Retina Fundus Image Dataset

In this section, we present and detail the datasets used to evaluate the performance of all our contributions. All the datasets used are public and popular in the field of diabetic retinopathy detection. The datasets contain retinal fundus images and were split into training, validation, and test sets using the Sklearn

operation "train_test_split.".

All the retina fundus images in those datasets are rated into five classes by experts to represent the stage of diabetic retinopathy. The images are rated from 0 to 4 as follows: Label zero, or no DR, for the patient that has no DR; label one, or mild DR, for the patient that suffers from a mild NPDR level; label two, or moderate DR, for the patient that has a moderate NPDR level of this disease; label three, or severe DR, for the patient that suffers from a severe NPDR stage; and label four, or proliferative stage, for the patient that suffers from the proliferative DR (PDR) stage.

2.2.1 APTOS dataset

APTOS [26] stands for the Asia Pacific Tele-Ophthalmology Society dataset. The APTOS dataset is a set of retinal fundus images collected from various hospitals using different cameras over an extended period of time. It comprises 3,662 retinal images, graded from 0 to 4 by clinicians based on the severity of diabetic retinopathy. These images may exhibit noise, artifacts, focus issues, and variations in size, orientation, and shape. The dataset is rated as follows: no diabetic retinopathy (class 0), mild DR (class 1), moderate DR (class 2), severe DR (class 3), and proliferative DR (class 4).

2.2.2 EyeSPACS dataset

The EyePACS dataset [27] consists of a large collection of retina fundus images intended for diabetic retinopathy analysis. These images were captured using various camera types and models, resulting in potential issues such as focus problems, underexposure, overexposure, and noise. The dataset is divided into labeled and unlabeled sets, with our focus on the labeled set, which contains approximately 35,000 retinal fundus images for classification tasks. The EyePACS dataset is

rated as follows: no diabetic retinopathy (class 0), mild DR (class 1), moderate DR (class 2), severe DR (class 3), and proliferative diabetic retinopathy (class 4). In our contribution, we focused only on the labeled images and partitioned the images into training, validation, and test sets.

2.2.3 Messidor-2 dataset

The Messidor-2[28] dataset comprises retinal fundus images of diabetic retinopathy patients from the ophthalmology department at Brest University Hospital (France). Building upon the original Messidor-1 dataset [29], which includes around 1200 images, Messidor-2 expands this collection with approximately 1748 additional retinal fundus images obtained from the same hospital. Images in the Messidor-2 dataset are categorized into severity levels: no diabetic retinopathy (class 0), mild DR (class 1), moderate DR (class 2), severe DR (class 3), and proliferative diabetic retinopathy (class 4).

2.2.4 DDR dataset

We also utilized the DDR dataset, introduced by Li et al. [30], to assess our models's performance. This dataset comprises over 13,673 retinal fundus images collected from 147 hospitals across China. The DDR dataset includes annotations for DR grading levels, along with a sixth category for images of poor quality that were not utilized in our evaluation.

2.3 Evaluation Metrics

In this section, we illustrate the evaluation metrics we employed to prove the performance and quality of our proposed contribution. Due to our focus on

classification tasks, we used metrics that are essential for measuring how well our models perform across the classification tasks.

2.3.1 Accuracy

Accuracy is an evaluation metric that measures the proportion of correctly classified instances among all instances evaluated. Besides, it is an indication of how well a model's predictions match the true values in a dataset. The equation (2.1) provides the used equation to compute the accuracy, which consists of dividing the total of the true predictions by the total number of predictions.

$$Accuracy = \frac{Number\ of\ correctly\ predicted}{Number\ of\ total\ prediction} \tag{2.1}$$

2.3.2 Precision

Precision is an evaluation metric for binary and multiclass classification tasks. It is measured by dividing the correctly predicted instances of a label by all predicted instances of the same label. The equation (2.2) is used to calculate the precision.

$$Precision = \frac{Number\ of\ Correctly\ predicted\ of\ given\ label}{All\ number\ predicted\ of\ the\ same\ label} \tag{2.2}$$

2.3.3 Recall

Recall is also an evaluation metric for the binary and multiclass classification tasks. It clarified how many true positive cases we were able to predict correctly. It is calculated using the equation (2.3), which is done by dividing the number of correctly predicted simples by the number of true examples of the same label.

$$Recall = \frac{TP}{TP + FN} \tag{2.3}$$

2.3.4 F1-score

F1-score is a measures metrics that merge recall and precision into one measure to determine a balanced assessment of the model's performance. It is computed using the equation (2.4).

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{2.4}$$

2.3.5 Kappa Score

The kappa score, also known as Cohen's Kappa coefficient, is a statistic that measures the agreement between the predicted and actual classifications made by the model. The equation (2.5) represents the mathematical equations used to compute the kappa score.

$$Kappa = \frac{(p_o - p_e)}{(1 - p_e)} \tag{2.5}$$

Where Pe is the expected agreement and Po represents the observed agreement.

2.3.6 Confusion matrix

A confusion matrix is a table that is used to evaluate the performance of a classification model. where the number of lines and columns are the number of classes used during the classification task. It organizes the model's predictions and the actual labels across all classes in a tabular format.

2.4 State of the art on diabetic retinopathy prediction

This section explores various studies on state-of-the-art techniques for grading diabetic retinopathy using computer vision techniques. These studies focus on enhancing their approaches by leveraging powerful machine learning techniques, introducing new pre-processing techniques to improve the quality of retinal fundus images, and employing deep learning models, including hybrid models and advanced deep learning models, to achieve more accurate and reliable diabetic retinopathy grading.

2.4.1 Supervised and Unsupervised Machine Learning

Machine learning models have demonstrated significant advancements in achieving high-performance accuracy for the early detection of diabetic retinopathy (DR) through the analysis of medical images, showcasing their potential to revolutionize the early diagnosis and management of this condition. These models leverage machine learning techniques to analyze intricate patterns within medical images, enabling them to identify subtle indicators of DR at an early stage with remarkable precision and reliability.

2.4.1.1 Support vector machines

The healthcare domain uses Support Vector Machine (SVM), a powerful algorithm of supervised machine learning models, for regression and classification. SVM can efficiently classify complex medical data, such as identifying tumors and detecting diseases using images as input or distinguishing between different stages of disease.

In the spirit of detecting diabetic retinopathy prediction, Carrera et al. [31]

introduced an automated method for detecting diabetic retinopathy using support vector machines (SVM). Their approach involves isolating blood vessels, microaneurysms, and hard exudates to extract features utilized by a support vector machine. This model, a significant application of machine learning algorithms, classifies 400 labeled retinal images on a 4-stage of diabetic retinopathy. They achieved a sensitivity of 95% and an average accuracy of 85%. Moreover, Chetoui et al. [32] have introduced a machine learning-based approach for diabetic retinopathy identification. The proposed approach is based on an SVM model with two output labels to classify retina fundus mages with or without diabetic retinopathy (DR). They utilized LTP (Local Ternary Pattern) and LESH (Local Energy-based Shape Histogram) as features on the MESSIDOR dataset [29]. Their method demonstrated strong performance for a small dataset like MESSIDOR and binary output, achieving an accuracy of 90.4% and an AUC (area under curve) of 93.1%. This highlights the effectiveness of their approach in accurately identifying the presence of diabetic retinopathy in retinal images, showcasing its potential for aiding in the early diagnosis and management of this condition.

In the same vein of grading DR and detection using SVM, Adal et al.[33] introduced a method for identifying diabetic retinopathy (DR) in images by detecting small red lesions. Their approach involved determining DR-related changes based on various density and shape features using a support vector machine classifier. This method achieved a sensitivity of 80% in identifying DR-related lesions. Furthermore, their work demonstrates the potential for accurately detecting DR-related changes in retinal images, which is crucial for the early diagnosis and effective management of diabetic retinopathy. Finally, Qomariah et al.[34] presented a new study for diabetic retinopathy prediction based on support vector machines (SVM). They extracted features from the last fully connected layer of

the transferred learning convolutional neural network and fed them to the support vector machine as input. Their approach reduces the computation time required by the classification process using the convolutional neural network with fine-tuning.

2.4.1.2 fuzzy classifier

Another type of machine learning algorithm is the fuzzy classifier, which is widely used in the healthcare domain. Fuzzy classifiers are applied to a variety of tasks, such as disease diagnosis, medical image analysis, and patient risk assessment. For example, they can be used to determine the severity of a disease based on a range of symptoms and test results.

In this regard, Jahiruzzaman et al.[35] introduced a powerful method for the detection of exudates and hemorrhages in retinal images in order to predict diabetic retinopathy. The approach utilizes a k-means color compression operation to cluster fundus images into distinct regions of interest, decreasing color dimensions. Subsequently, the diabetic fundus is segmented and analyzed based on regional property attributes. The recognition of diabetic retinopathy is then performed using a knowledge-based fuzzy inference system (FIS) with these effective attributes through experimentation and trial. This approach demonstrates the potential to accurately identify and classify diabetic retinopathy, offering a cost-effective solution for early detection and diagnosis.

2.4.1.3 K-means cluster

K-means is a popular unsupervised clustering algorithm used to partition a dataset into K-distinct clusters based on feature similarity. K-means clustering has found diverse applications in the healthcare area, offering valuable insights and solutions for various healthcare-related challenges. Such as tumor segmentation, disease

prediction, and analysis of medical images.

In the diabetic retinopathy detection task, K-means has shown great results in the detection of DR symptoms and symptoms segmentation. Such as Biyani et al.[36] presented a paper based on K-means clustering as a classifier and morphological technique as pre-processing to detect the exudates in retina images. They used the public DIARETDB1 dataset to test their approach.

2.4.1.4 K-Nearest Neighbors

The k-nearest Neighbors model in supervised machine learning relies on proximity to produce data with similar attributes. It consists of calculating the distance between all input data using specific distance metrics, such as Euclidean. Then, it identifies the 'k' closest data point to the new data. Finally, determine the label of the new data based on the majority class between the 'k' nearest neighbors.

in the task of diabetic retinopathy prediction, Furthermore, Wu et al. [37] proposed an approach for detecting microaneurysms (MAs) in the eye, focusing on pre-processing phase, candidate extraction (MAs), feature extraction, and classification. They also employed three models, namely K-Nearest Neighbors, Naïve Bayes, and AdaBoost, to achieve optimal results. The performance of their method was assessed using two publicly available datasets: E-Optha and ROC. The outcomes demonstrate that the KNN and Adaboost classifiers have similar classification performance and outperform Naïve Bayes classifier.

2.4.2 Pre-processing based approaches

Several studies based on distinct medical image analysis techniques have been proposed to offer a more accurate diagnosis of diabetic retinopathy, focusing on improving and enhancing the pre-processing process through the utilization of multiple medical imaging techniques. These efforts aim to enhance the quality of

the classification and detection of diabetic retinopathy by enhancing the quality of the image dataset. Jamal et al. [38], among the researchers focusing on the pre-processing process, enhanced the accuracy of their approach by improving the quality of retinal images in the pre-processing phase. Their strategy effectively splits the background space and noisy areas of the retinal fundus image by employing both fine and coarse segmentation techniques. This novel approach demonstrates the potential for significantly enhancing the retinal images, which is essential for the proper analysis and diagnosis of diabetic retinopathy. In the same regard, Kar and Maity [39] suggested an automatic lesion detection method for DR images, which is based on four main steps. The first one involves the extraction of blood vessels and the removal of the optic disc. Subsequently, they separate the dark lesions from the poorly illuminated retinal background. Next, they focus on the detection of hemorrhages (HEMs), exudates (EXs), and microaneurysms (MAs). In the end, a post-processing stage is performed To eliminate candidate pixels that were mistakenly detected. In addition, Orujov et al.[12] proposed an image preprocessing technique for retinal segmentation. Their approach, which relies on contour detection and fuzzy rules, was applied to three different datasets in order to identify blood vessels in images: DRIVE (Retinal Images for Vessel Extraction), STARE (Structured Analysis of the Retina), and the CHASEDB datasets. The accuracy of their method was quite satisfactory, with scores of 86.5%, 93.9%, and 95% for the STARE, DRIVE, and CHASEdb datasets, respectively.

In the same vein of grading DR and detection, Vijayan et al.[11] introduced a novel approach for grading and detection of diabetic retinopathy utilizing the Efficientnet-B0 model. Their method includes a preprocessing step to normalize and minimize the impact of variations in illumination. The approach demonstrated an accuracy of 84.8% on the DDR dataset and 86.2% on the APTOS

dataset. However, their study's limitations are associated with the interrelationships between different severity degrees in the model. In the same regard of proposing an architecture based on pre-progressing for grading DR and detection, Islam et al.[40] implemented a row-stage training supervised contrastive learning (SCL) method with a supervised contrastive loss function, utilizing the pre-trained Xception model as the encoder with transfer learning. Additionally, they applied contrast-limited Adaptive Histogram Equalization (CLAHE) in the pre-processing phase to improve the quality of retinal fundus images. Their approach was validated on the Messidor-2 and APTOS datasets, achieving a test accuracy of 84.356% for multitask classification and 98.36% for binary classification. Furthermore, Mondal et al. [10] introduced an Ensemble Deep Learning for Diabetic Retinopathy (EDLDR) approach for grading diabetic retinopathy. Their method is based on a pre-processing step where it starts with the application of the Contrast Limited Adaptive Histogram Equalization (CLAHE) method for histogram equalization, followed by the use of a GAN-based augmentation technique for data augmentation. Subsequently, the processed images were fed into an ensemble deep learning model comprising modified ResNeXt and DenseNet101. The EDLDR method attained an accuracy of 96.98% for diabetic retinopathy identification and 86.08% for diabetic retinopathy severity detection. However, a major limitation of this approach is the imbalance of limited data for the five classes, which may impact its generalizability and robustness.

Besides, Das et al.[7] introduced a novel deep architecture that leverages segmented images for the classification of diabetic retinopathy images. The preprocessing phase involves employing the maximum principal curvature technique to extract blood vasculature from the images, followed by adaptive histogram equalization and morphological opening for image enhancement. For the training phase, they employed two separate CNN models to classify the images as

either normal or abnormal. Their method demonstrated high performance on the DIARETDB1 dataset, by achieving a precision of 97.2% and an accuracy of 98.7%. Ramasamy et al.[41] developed an automated approach for detecting diabetic retinopathy (DR) by extracting features from retinal images based on textural gray-level, co-occurrence, run-length matrix, and the coefficient of the continuous ridgelet transform. In the pre-processing step, they applied a series of pre-processing operations, including image conversion, filtering, morphological operations, and segmentation. They used the green channel of the retina fundus image to apply filtering to denoising the image using the Laplacian filter. In addition, morphological closure is performed to clean the main blood vessels. In the classification step, they utilized the Sequential Minimal Optimization (SMO) classification to classify the DIARETDB1 and KAGGLE datasets into two classes, achieving impressive results. Specifically, they attained a sensitivity of 98.87%, a specificity of 95.24%, and an accuracy of 97.05% on the DIARETDB1 dataset, and a sensitivity of 90.9%, a specificity of 91.0%, and an accuracy of 91.0% on the KAGGLE dataset. This innovative approach demonstrates the potential of using processing techniques to improve the effectiveness and performance of the proposed approach.

2.4.3 Deep Learning based approaches

In the last decade, deep learning has shown great effectiveness in the classification of medical imaging and the detection and prediction of diseases. In this regard, several studies have been presented. These studies provide an effective and powerful automatic approach for the severity grades of diabetic retinopathy based on convolutional neural networks and the traditional deep learning approach.

Several approaches based on convolutional neural networks are proposed for the task of DR severity level detection, such as Pratt et al. [42] utilized a CNN

model to grade the retinal fundus images to distinct levels of diabetic retinopathy (Non-DR, Mild, Moderate, Severe, and PDR). Their model achieved a sensitivity of 30%, a specificity of 95%, and an accuracy of 75% during validation, by validating their approach using the EyePACS[27] Kaggle dataset. Yang et al. [43] introduced a strategy for diabetic retinopathy detection based on deep convolutional neural networks. Their approach aimed to identify abnormal lesions and severity grades of diabetic retinopathy in fundus images using the same dataset. Additionally, Sharma et al.[44] suggested an automated method to detect diabetic retinopathy in a huge dataset comprising 35,000 retina fundus images. They employed a DCNN model to classify the image dataset into five stages and attained an accuracy of 74.04%. Besides, Gargeya et al.[1] developed a CNN that extracts features from retina fundus images and utilizes a decision tree model for classification. The model categorizes the features into binary classifications (with or without DR) and was tested on public datasets. It achieved a sensitivity of 90% and a specificity of 94% on the E-Optha dataset, as well as a sensitivity of 93% and a specificity of 87% on the MESSIDOR2 dataset..

In the same regard of grading DR utilizing the deep learning models, Jain et al.[45] utilized three deep learning architectures (VGG19, VGG16, and Inception-V3) for both grading diabetic retinopathy and binary classification, aiming to automate DR detection. They focused on data augmentation to enhance performance. For the 5-grade task, the model yielded an accuracy of 76.9%, while for the binary task, they achieved an accuracy of 80.4%. Their emphasis on data augmentation reflects a commitment to improving model robustness and generalization, essential for accurate and reliable DR classification. Moreover, Murugan et al.[46] proposed an automatic approach based on a deep convolutional neural network. They tested their proposed method on the IDRID [47], ROC [48] datasets, and a local dataset, achieving a, sensitivity, accuracy, and specificity

of, 98.45%,98.2%, and 98.56%, respectively. Their approach focuses only on the binary classification of diabetic retinopathy without considering the five stages of DR. Furthermore, Khan et al.[49] employed a new VGG-NiN approach to classify diabetic retinopathy (DR) using a combination of VGG architecture and a spatial pyramid pooling layer, along with a network-in-network model. Their method aimed to classify the EyePACS dataset into five grades of DR and achieved an accuracy of 85%. The proposed model focused on reducing model training time and convergence time in DR classification.

Despite the effectiveness of deep learning approaches in medical image classification, this approach suffers from several issues. Among the main problems with deep learning are the requirements for a large amount of data, the need for substantial computational resources to train a deep learning model, and the time-consuming nature of the training process. One of the most effective techniques to address these challenges is the use of transfer learning methods, which leverage pre-trained models to improve the performance of deep learning models in medical image analysis. Transfer learning strategies have been shown to be beneficial in mitigating the data and resource requirements while improving the efficiency of the training process, making them a valuable tool in the advancement of medical image classification using deep learning. Pak et al.[50] conducted a comparative analysis of conventional architectures, including DenseNet and ResNet, with the newly optimized EfficientNet. Their study focused on classifying retinal images in the APTOS dataset into five classes. The findings revealed that the most favorable results were achieved using the EfficientNet-b4 model, which attained a Kappa value of 79% during the validation phase. Recently, Wong et al.[8] introduced a novel approach, ECOC-SVM, based on transfer learning with optimized feature weights and parameters. This method involves utilizing selected features from the pooling layer of two transferred learning networks,

ShuffleNet and ResNet-18, which are then fed to an Error Correction Output Code (ECOC). The evaluation of this approach was conducted on two datasets: the APTOS dataset and a combination of the EyePac and Messifor-2 datasets. For the tasks of 5-class diabetic retinopathy (DR) grading and 2-class identification, the approach achieved an accuracy of 82% and 96% on the APTOS dataset, respectively. Additionally, for the task of 3-class DR, an accuracy of 75% was achieved with the combination of EyePac and Messidor-2. These results underscore the effectiveness of the ECOC-SVM approach in accurately classifying DR images across different datasets and classification tasks, highlighting its potential for enhancing DR diagnosis and identification. In this spirit, Dekhil et al.[51] employed VGG16 model with transfer learning technique to suggest an architecture for grading diabetic retinopathy fundus images into five classes. This architecture includes a pre-processing phase where the images are cropped before being fed to the proposed model. the approach validated on the APTOS dataset and attanted an accuracy of 77%. Besides, Gangwar et al.[9] utilized a pre-trained Inception-ResNet-v2 deep learning model to address the diabetic retinopathy detection challenge within the APTOS dataset. They used a custom layer block above of this DL model. As result their architecture achieved a notable accuracy of 82.18% on the APTOS dataset.

Moreover, Rao et al.[52] presented comprehensive experimental studies for the classification of diabetic retinopathy (DR) fundus images, utilizing five CNN classifiers. Their findings revealed that for binary classification, the ResNet50 architecture with the Adam optimizer achieved the highest accuracy of 96.59% on the test set. Additionally, for the 5-class classification task, the Inception-ResNetV2 architecture with the Adam optimizer attained the best accuracy of 85.02% on the test set. Furthermore, ZENG et al.[2] introduced a method based on deep learning for diagnosing diabetic retinopathy (DR) by classifying color

fundus images into two grades (0 without DR, 1 with DR). This approach leverages a pre-trained Inception-V3 model with a Siamese-like architecture, trained using transfer learning. The method achieved a kappa score of 82.9% on the EyePACE dataset. Additionally, Aatila et al. [53] presented a system based on deep learning for diabetic retinopathy (DR) classification. The system utilizes ResNet50 with transfer learning and achieved an accuracy of 70% on the EyePACS dataset. In the same regard, Tariq et al.[54] conducted classification of the EyePACS dataset using pre-trained ResNet50 and DenseNet121 models. They incorporated a clustering technique to enhance accuracy, resulting in an accuracy score of 70%. Besides, Ran et al.[55] introduced a novel architecture for diabetic retinopathy (DR) grading called SFADA, which is an abbreviation for source-free domain adaptation. Their approach consists of extracting information from a trained model to benefit from this information to produce a powerful DR grading model of ultra-wide-field fundus images. Their method achieved an accuracy of 85.36%, an F1-score of 83.71%, and a kappa score of 80.54%. Finally, Abbasi et al.[56] introduced a novel method for diabetic retinopathy (DR) classification, incorporating transfer learning, knowledge distillation, and the utilization of unlabeled data. Their approach was evaluated on the Messidor and EyePACS datasets, achieving the highest accuracy of 82.32% for the Messidor dataset and 76.84% for the EyePACS dataset. Finally, Furthermore, AL-ANTARY et al.[5] introduced the multi-scale attention network (MSA-Net) as an innovative and automated approach for detecting and classifying diabetic retinopathy. Their proposed approach is based on four blocks in addition to a pre-processing step. After the normalization and reduction of the illumination and intraclass variation effect of the retina image in the pre-processing step, they fed it to the four blocks of the MSA-Net architecture: starting with the encoder block for the high-level representation process, followed by a multi-level and multi-scale feature repre-

sentation block, then a multi-scale attention mechanism, and finally, the decoder block for generating the DR score. This method yielded an accuracy score of 79.9% and 84.6% for DR severity levels on the EyePACS and APTOS datasets, respectively. Additionally, for the task of binary classification of DR, they attained an accuracy of 87.5% and 98.% and on the same datasets. Abdelmaksoud et al.[57] proposed a method called E-DenseNet for diabetic retinopathy detection based on a hybrid model. Their approach involves the DenseNet model based on transfer learning and the EyeNet model. Which is done by inserting the dense block to customize the EyeNet and optimizing the hybrid E-denseNet hyperparameters. They validated their method on several datasets, such as the APTOS and IDRID datasets, and achieved an accuracy score of 84% and 93% respectively, for the task of grading DR. Finally, Butt et al.[58] proposed a hybrid model for diabetic retinopathy detection. They extracted features using a pre-trained CNN model on retinal fundus images. Based on these extracted features, they classify the retinal fundus images into binary and multiclass classifications of fundus images by passing this feature to several classifiers. They achieved an accuracy of 97.8% for the binary classification task and 89.29% for the 3-class task on the APTOS dataset. Finally, Chaturvedi et al.[59] introduced an automated method for detecting diabetic retinopathy (DR) using the pre-trained DensNet121 network. They employed a data augmentation technique in the pre-processing stage using the ImageDataGenerator function from Keras. This model was able to classify images from the APTOS dataset into five stages of DR. The results were promising, with a validation f1-score of 86%, a precision of 86%, a kappa score of 91.96%, and a recall of 87%.

2.4.4 Advanced deep learning based approaches

Over the past decade, deep learning has proven its effectiveness in the computer vision domain, expanding its applications across various medical fields. Numerous deep learning-based architectures have been proposed for disease segmentation and disease prediction, reflecting the growing influence of deep learning in the healthcare domain. Recently, advanced deep learning models such as attention mechanisms, capsule networks, vision transformers (ViT), and graph neural networks have been used and applied in this domain, which has revolutionized the healthcare domain by significantly enhancing diagnostic accuracy and disease prediction.

In this spirit, Hu et al.[60] introduced a graph adversarial transfer learning approach for diabetic retinopathy (DR) detection. Their model is based on self-supervised training; they utilize a graph neural network to extract potential features and employ adversarial training to improve model robustness. This methodology yielded an accuracy of 83.5% for multiple DR classifications on the APTOS 2019 datasets and 94.3% for binary DR classifications within the same dataset. However, the limitation of their method pertains to the challenge of addressing the fine classification of lesions due to the small size and similarity of DR lesions. In addition, Shaik and Cherukuri [61] introduced a model designed for the classification of diabetic retinopathy severity levels, featuring multiple attention stages known as Hinge Attention Networks (HA-Net). Their approach includes a pre-trained VGG16 base for extracting initial spatial illustrations from retinal fundus images, as well as a spatial attention autoencoder to learn lesion-specific latent illustrations in spatial dimensions, in addition to a channel attention-based hinge neural network to grab category-based discriminative features in channel dimensions and categorize the different stages of retinopathy. Besides, the model integrates a convolutional LSTM (Long short-term memory) to provide priority

to highly important spatial maps before forwarding to the hinge neural network. The model demonstrated strong performance, achieving an accuracy of 85.54% on the APTOS dataset and 66.41% on the IDRiD dataset. Furthermore, Li et al.[62] introduced a new cross-disease attention network (CANet) for jointly grading diabetic retinopathy (DR) and diabetic macular edema (DME) by exploring the internal relationship Within the maladies with only image-level supervision. Among their most important contribution are the disease-specific attention module to selectively learn useful information for individual diseases and a disease-dependent attention module to better understand the interdependent nature of the two diseases. The attention modules are integrated into a deep network to generate disease-specific and disease-dependent features, maximizing the overall performance for grading DR and DME. The approach achieved an accuracy of 65.1% on the IDRiD dataset and 92.6% on the Messidor dataset. In the same vein, Farag et al.[6] introduced an automated strategy for diagnosing diabetic retinopathy (DR) based on utilizing DenseNet-169's encoder to extract features in addition to a Convolutional Block Attention Module (CBAM) to refine these features. They validated their method on the APTOS dataset, achieving an impressive accuracy score of 97% for the binary task and 82% for severity DR grading.

In the same regard, Bodapati et al.[63] proposed a composite deep neural network with a gated-attention mechanism for grading diabetic retinopathy tasks. They achieved a kappa and accuracy scores of 97% and 82.54% on the APTOS dataset, respectively. Additionally, they introduced a diabetic retinopathy prediction method utilizing a stacked convolutional auto-encoder with spatial attention, achieving accuracies of 63.24% and 84.17% on the IDRiD and APTOS datasets, respectively. This innovative approach demonstrates the potential to significantly enhance the accuracy and efficiency of DR diagnosis and treatment, showcasing

promising results in the field of medical image analysis and classification. Besides, Mohan et al.[3] introduced ViT-DR, a vision transformer-based approach for detecting the severity of diabetic retinopathy (DR). Their method involves dividing fundus images into non-overlapping patches, converting these patches into sequences, and then subjecting them to a positional embedding process before feeding them to multiple multi-head attention layers. This innovative approach demonstrates the potential to significantly enhance the accuracy and efficiency of DR diagnosis and severity grading, showcasing promising results in using the ViT in medical image analysis and classification field. In addition, Nneji et al.[4] proposed a method for diabetic retinopathy classification utilizing a capsule network. They applied contrast-limited adaptive histogram equalization as a preprocessing step to enhance the retina fundus image, followed by feature extraction using the capsule network. Their approach, tested on the Messidor dataset, achieved an impressive accuracy of 99.1% using a sigmoid classifier. Finally, Wang et al.[64] introduced a CLC-Net method for the segmentation of four distinct types of fundus lesions associated with DR using a fully automatic deep convolutional neural network, including an attention mechanism. This method involves a contextual branch and a local branch to exploit multi-scale image information. In addition to using the attestation mechanism, Besides, add an auxiliary classification head at the bottom of the network encoder to classify the severity levels of diabetic retinopathy. Besides,

2.5 conclusion

In this chapter, we highlighted the importance of using pre-processing operations to enhance the proposed approaches to detecting diabetic retinopathy. Also, we presented several approaches that are based on deep learning and advanced deep

learning approaches. Besides, we presented the most used dataset in these fields and the evaluation metrics used to evaluate these approaches.

Chapter 3

Diabetic retinopathy prediction based on deep learning and deformable registration

3.1 Introduction

Diabetic retinopathy is an eye condition that impacts patients eyes, typically those suffering from diabetes. Its development may cause damage to the retina, which can lead to blindness. The appropriate treatment of this medical condition is based on the task of earlier detection and identifying the exact stage of that pathology. In fact, several works are proposed to improve the process of diabetic retinopathy prediction, through presenting new pre-processing strategies to improve the quality of retinal images or by strengthening and proposing new deep-learning models.

In this chapter, we suggest a novel approach for diabetic retinopathy detection using deep learning and deformable registration, which consists of two main steps. A primary step called the pre-processing phase consists of applying deformable

registration to enhance quality of the retina fundus image before feeding it to the following step. The second step is the classification step, which consists of categorizing the retina images into the exact stages of diabetic retinopathy that the patients are struggling with using several convolutional neural networks (CNN) architects and techniques.

The remainder of this chapter is organized as follows: the materials and methods that are employed are examined in the following section. The suggested approach is explained in the third section. In section four, experimental results and an assessment of the suggested approach are presented. The chapter's conclusion is found in Section five.

3.2 Materials and Methods

Diagnosing diabetic retinopathy is still a difficult procedure that requires experienced ophthalmologists. In this section, we presented the Materials and Methods utilized in our new proposed approach that combines powerful pre-processing operations and new deep-learning approaches.

3.2.1 Deformable Registration

One of the most commonly used techniques to follow the development of tumors or track the growth of disease is the image registration technique. This technique consists of matching a fixed image taken for the tumor in an early stage with a moving image of the same tumor after a period of time. According to the type of transformation, we can divide the registration technique into two main categories: rigid and non-rigid transformation. The rigid transformation involves applying simple transformations such as rotation, translation, and reflection without affecting the shape of the tumor or deforming it. Whereas, the

non-rigid transformation, which is also called the deformable registration, consists of reshaping and deforming the shape of the tumor by performing an elastic transformation based on the B-spline method.

Sorzano et al. [65] have optimized the B-spline registration algorithm by proposing a new regularization constraint, by minimizing the function of energy (3.1) of the deformable registration. Where E_{img} is the measure of dissimilarity between the target image and the warped source image, the soft landmark constraints E_{μ} a priori knowledge about the deformation field through the two independent measures E_{div} and E_{rot} that are related to the gradients of divergence and curl of the deformation field.

$$E = w_i E_{img} + w_{\mu} E + (w_d E_{div} + w_r E_{rot}) \tag{3.1}$$

As represented in [65; 66] the deformation field based on the B-splines (3.2): Where S_x and S_y are scalars controlling the degree of detail of the representation of the deformation field, β^3 means B-spline of degree 3, and $C_{k,l}$ are the B-spline coefficients.

$$\begin{aligned} g(x) =& g(x,y) \\ =& (g_1(x,y), g_2(x,y)) \\ =& \sum_{k,l \in \mathbb{Z}} \binom{C_{1,k,l}}{C_{2,k,l}} \beta^3 \left(\frac{y}{S_y} - l\right) \end{aligned} \tag{3.2}$$

3.2.2 Transfer learning

Training a large deep convolutional model for computer vision tasks such as image segmentation or image classification requires a large training dataset, and we frequently face the issue of the dataset's small size. One of the most commonly used solutions to such problems is the transfer learning (TL) technique, which

consists of training the deep-learning network on a huge dataset and then fine-tuning it on a smaller one. As stated by Yu et al.[67], Pre-training the models on a large-scale dataset such as the ImageNet dataset and fine-tuning them on the new dataset is a wise decision.

The transfer learning process consists of using the weight of a pre-trained model and using it as the starting weight of our base model to avoid training it from scratch. This pre-trained model, which was trained on a huge dataset, usually has more labels than the new classification task. Therefore, we remove the last layer, which is responsible for the classification process, and replace it with a new layer that is compatible with our case. Even more, we can add several new deep-learning layers. In addition to fine-tuning these transferred models, we train the model in the first stage by freezing the weight of the base model from updating except for the newly added layers, then we unfreeze the frozen layer and train the hall model. Transfer learning and the fine-tuning process can decrease the processing time.

3.2.3 Inception-V3

Inception-V3 is a type of deep learning architecture that is based on a convolutional neural network. It is considered the third version of the Inception architecture. Proposed and described by Szegedy et al.[68] in the research paper named "Rethinking the Inception Architecture for Computer Vision." This architecture is widely used in image classification and recognition, which makes it one of the best choices for medical image classification, such as diabetic retinopathy detection.

The architecture of Inception-V3 came with many benefits, such as reducing computational power while preserving efficiency and providing deeper networks without growth in the number of parameters, which made this architecture in-

troduce several innovations, including, Auxiliary classifiers, which help prevent overfitting. Factorized convolutions, which helps reduce the computational complexity of the convolution.

3.2.4 DenseNet-121

Dense Convolutional Network (DenseNet) is a type of deep learning architecture proposed in the research paper "Densely Connected Convolution Network" by Huang et al.[69]. This architecture is one of the greatest options for medical picture classification, such as diabetic retinopathy detection, because it is widely utilized in image classification and identification.

The proposition of DenseNet comes with many advantages, such as solving the problem of redundant layers by connecting all layers using the concept of dense blocks. Dense blocks enable each layer to connect to all other layers, including the output layer of each layer, which is connected with the output of all previous layers and comprises the input layer. This concept of dense blocks introduces the possibility of feature reuse.

DenseNet-121 is a specific version of the DenseNet architecture, where the number 121 refers to the number of layers. 121 layers are organized as presented in the research paper [70]; 117 convolution layers; 3 transition layers; and one classification layer.

3.2.5 ResNet

ResNet (Residual Network) is a type of deep learning architecture presented by He et al.[71] in the research paper "Deep Residual Learning for Image Recognition." The ResNet architecture is widely used in image classification and showed a promising result in disease detection, which made it a good choice for the task of DR early detection.

The introduction of ResNet addressed the problem of vanishing gradients, which face a lot of challenges in deep learning architecture. They solve this problem using a new concept called the residual connection, which is done by adding the input features of a layer to its output.

3.2.6 Xception

Xception is a type of deep convolutional neural network architecture inspired by the ideas of the Inception model, proposed by Francois Chollet [72]. The Xception model is based on a new concept called the depthwise separable convolution instead of the inception block that comes with the Inception-v1 architecture.

The depthwise separable convolution made the Xception architecture more effective and efficient by separating the standard convolution into two steps: depthwise convolution and pointwise convolution. The depthwise convolution applies a single filter to each input channel (depth), and the pointwise convolution uses a 1x1 convolution to combine the outputs of the depthwise convolution. This change in the main body of the architecture reduces the number of parameters and the computational cost.

3.3 Proposed approach

In this chapter, we a novel automatic approach for diabetic retinopathy detection based on deep learning and deformable registration. This proposed approach can solve one of the most common problems that can face the retina fundus image, which directly affects the task to detect the severity stage of diabetic retinopathy. That is to say, the detection process is affected by the background space around the retina in the image dataset.

The different sizes of each image in the dataset make this black space around

the retina hard to detect. To address such an issue, Numerous common sense methods can be applied, either cropping them or zooming in on the involved images. Unfortunately, these operations can overlook some significant areas which could potentially guide the classification phase. In this proposed approach, we used deformable registration, which allows us to remove the background while still accounting for all of the information contained in the retina by deforming the retina over the entire image in the image. In addition, there is a classification step to classify the severity level of DR. That is to say, our proposed approach starts with a pre-processing step to generate enhanced retinal fundus images. Then we feed the generated image into the classification step to detect the severity levels of DR, as shown in Fig. 4.1.

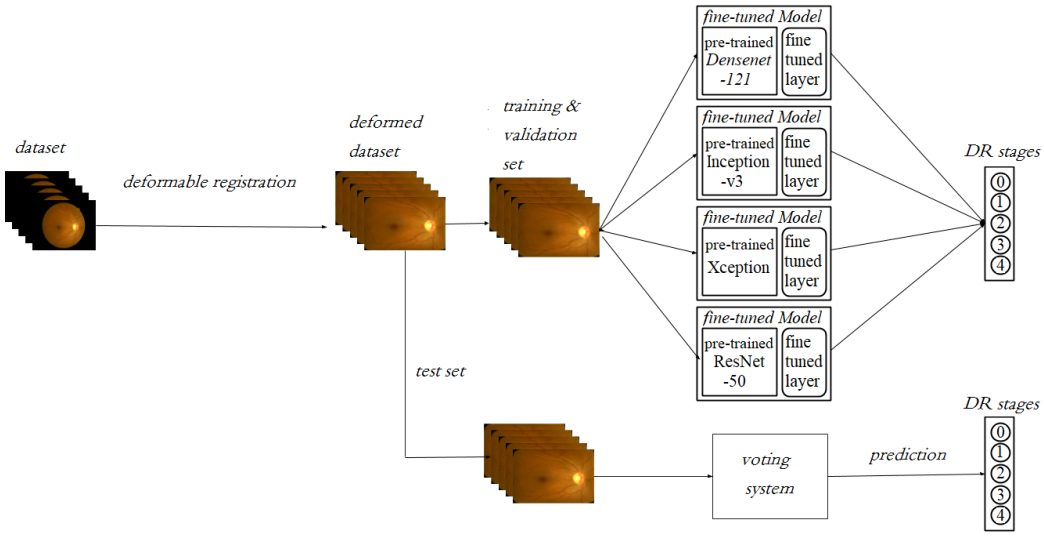

Figure 3.1: The global architecture of the proposed approach.

Preprocessing step : Since the image dataset has a different size, we start by resizing the image dataset to the same dimension. Next, we apply the deformable registration to the retina by reshaping the retina in the image to fill the entire space and remove the background, as presented in Fig.4.2. The deformable

registration we applied is based on B-Spline transformation. The deformable registration preserves all the characteristics of the image edges and conserves all the details in them. The edges contain crucial information that can help determine whether or not the image has DR. This assignment has a significant effect on the classification process, where any tiny details can influence the model's ability to make perfect decisions. After generating the enhanced image through the pre-processing step, we feed it to the classification step.

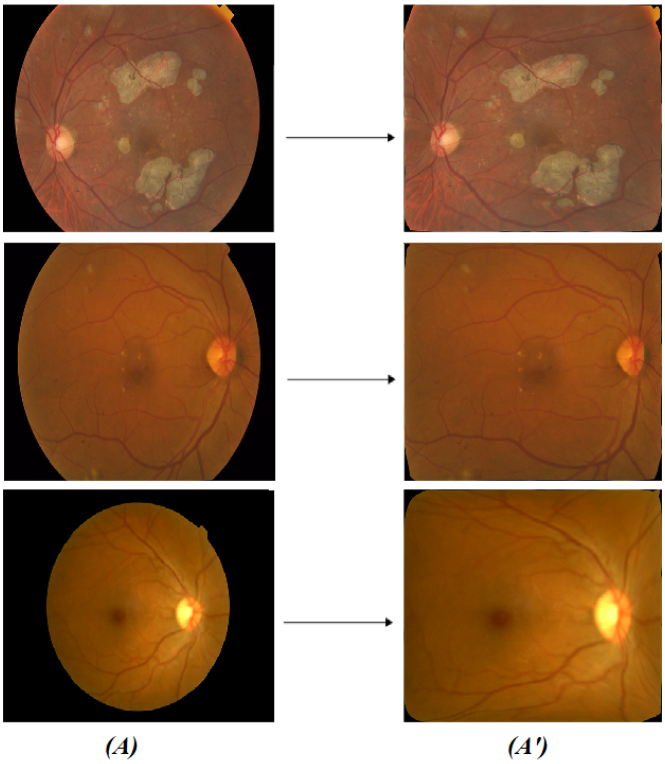

Figure 3.2: Example of an Image Dataset: (A): before applying deformable registration; (A'): after applying deformable registration.

Classification step : The classification step is responsible for the detection of the severity levels of diabetic retinopathy. We can divide this step into two main processes: the training and the testing processes. In the training process,

we used transfer learning on four deep convolution network models (DenseNet-121, Xception, InceptionV3, and ResNet-50). We used the weights of these pre-trained models to avoid training them from scratch, and we removed the last layer responsible for the classification task and replaced it with a dense layer with 1024 neurons, with a ReLu activation function, and an output layer with 5 units using SoftMax activation, which indicates the five grades in the APTOS dataset. Then we fine-tuned those models by freezing the top weights of these base model layers and updating the weights of the new layers. Every model (DenseNet-121, ResNet-50, InceptionV3, and Xception) trained and validated independently on the dataset and predicted its own decision in the test process. In the testing process, in order to give the final prediction, we used the voting by major technique, which consists of admitting the most duplicated output by each model in an effort to enhance the performance that each individual model is capable of.

3.4 Proposed Experimental Results and Discussion

3.4.1 Dataset

We validated the performance of the proposed approach during this chapter using the APTOS [26] dataset, which is widely used in the task of detecting DR. In addition, it is a public dataset and available on the Kaggle platform. The dataset is explained in more detail in Section 2.2.

The dataset was divided as follows: 2801 images for training, 494 for validation, and 367 images reserved for testing. The test set images were not used in either the training or validation phases, as shown in Table 3.1.

Table 3.1: The distribution of the used dataset during the training, validation, and test processes.

Label	Stage of DR	Train set	Valid set	Test set	Total
0	No DR	1383	251	171	1805
1	Mild DR	273	52	45	370
2	Moderate DR	773	126	100	999
3	Servere DR	152	26	15	193
4	Proliferative DR	220	39	36	295
Total	APTOS dataset	2801	494	367	3662

3.4.2 Evaluation Metrics

The performance of our proposed work was evaluated based on several metrics, such as accuracy, f1-score, recall, kappa score, and precision. All the measure metrics used in the chapter are explained in more detail in the section 2.3.

3.4.3 Performance Evaluation

We evaluated the performance of our proposed approach through analysis of model performance during the training and validation phases, in addition to the testing phase. In the end, we compared our obtained results with the state of the art.

3.4.3.1 Analysis of model performance during training and validation phase

In this subsection, we evaluate the performance of the fine-tuned models used in this approach during the training and validation process through the analysis of the loss and accuracy curves. The recorder accuracies and loss curves during the training and validation process provide crucial insights about the efficacy of the transfer learning technique and offer insightful information about the effectiveness

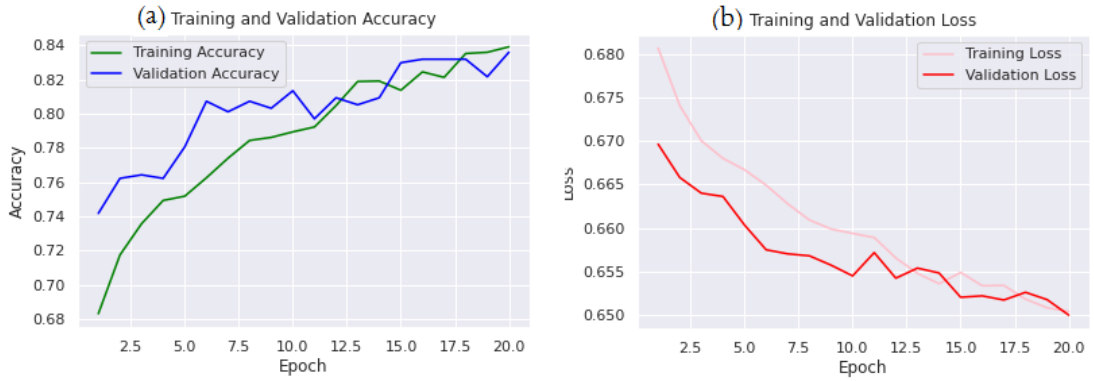

Figure 3.3: (a) The accuracy curve during the training and validation phases of the Xception classifier across epochs. (b) The loss curve during the training and validation phases of the Xception classifier across epochs.

of the improved model's performance.

Figure 5.7 represents the loss and accuracy of the fine-tuned Xception model in the training and validation process. During the training phase, the model yielded an accuracy of 83.92% with a corresponding loss of 0.65. Subsequently, in the validation phase, the model attained an accuracy of 83.61% with a loss of 0.64.

The performance evaluation of the fine-tuned Inception-v3 model is illustrated in Figure 5.10. The accuracy and loss of validation have been estimated at 83.17%, 0.65%, 81.76%, and 0.65%.

In the context of evaluating the performance of the fine-tuned DenseNet-121 model, the recorded metrics are depicted in Figure 5.8. where 83.23%, 0.65, 83.84%, and 0.65 represent the training accuracy, the training loss, the validation accuracy, and the validation loss, respectively.

The performance evaluation of the fine-tuned ResNet model is depicted in Figure 4.7, where the training accuracy, training loss, validation accuracy, and validation loss are reported as 79.63%, 0.65, 78.79%, and 0.65, respectively.

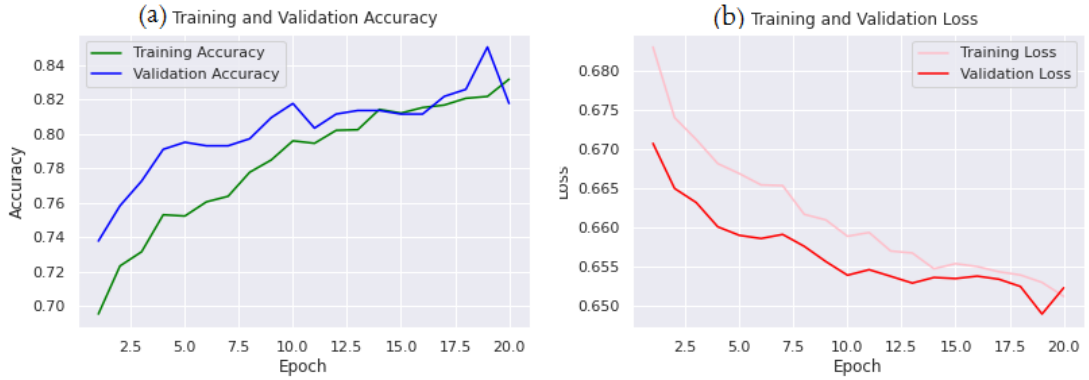

Figure 3.4: (a) The accuracy curve during the training and validation phases of the Inception-v3 classifier across epochs. (b) The loss curve during the training and validation phases of the Inception-v3 classifier across epochs.

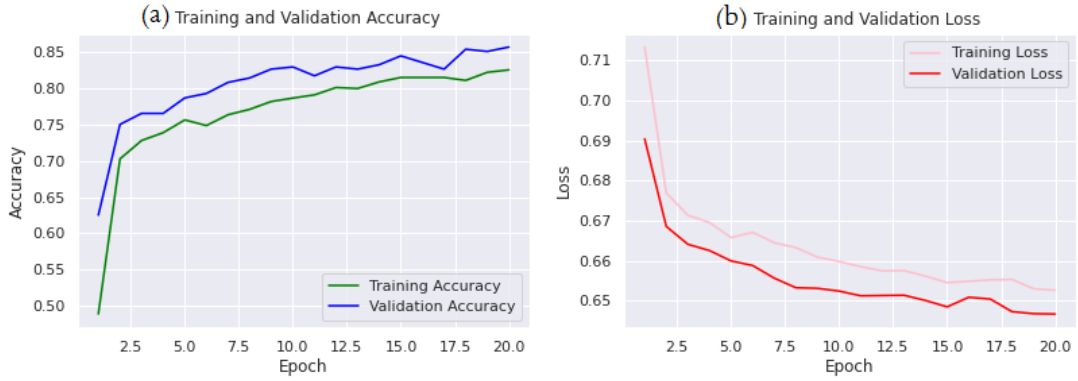

Figure 3.5: (a) The accuracy curve during the training and validation phases of the DenseNet-121 classifier across epochs. (b) The loss curve during the training and validation phases of the DenseNet-121 classifier across epochs.

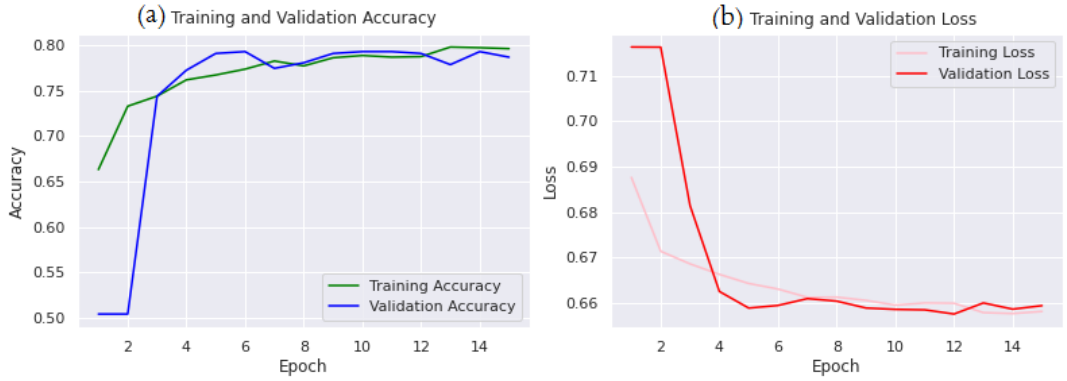

Figure 3.6: (a) The accuracy curve during the training and validation phases of the ResNet50 classifier across epochs. (b) The loss curve during the training and validation phases of the ResNet50 classifier across epochs.

3.4.3.2 Analysis of model performance during the testing phase

Several performance measures are used to evaluate the suggested model's effectiveness throughout the testing phase, including accuracy score, precision, kappa score, mean squared error (MSE), and f1-score. The testing phase performance of each model is displayed in Table 5.1. As we can observe, the accuracy of 83.92% has been obtained by the suggested technique that employs Densnet-121 as a classifier and provided the best MSE of 0.32, a kappa score of 75.92%, macro precision of 77.89%, macro recall of 71.09%, and f1-score of 83.33%. Similarly for the model used Xception as a classifier has also demonstrated excellent performance, with great accuracy results of 83.10%, an MSE of 0.40, the highest kappa score of 74.43%, macro precision of 75.11%, macro recall of 65.73%, and f1-score of 81.93%. Followed by the model used InceptionV3 as a classifier with an accuracy score of 82.01% and an MSE of 0.44. The lowest result is produced using ResNet-50 as a classifier, with an accuracy of 77.92% and an MSE of 0.67.

To achieve the final result, we used the ensemble voting by major technique. Which combines the output of these four models (ResNet-50, DenseNet-121,

model	inception-V3	Densnet-121	Resnet-50	Xception
Accuracy	0.8201	**0.8392**	0.7792	0.8310
MSE	0.4441	**0.3297**	0.6776	0.4059
kappa	0.7309	**0.7592**	0.6580	0.7443
Macro recall	0.7298	**0.7789**	0.4541	0.7511
Macro precision	0.6890	**0.7109**	0.4911	0.6573
Weighted f1-score	0.8145	**0.8333**	0.7244	0.8193
Macro f1-score	0.6966	**0.7343**	0.4637	0.6884

Table 3.2: The performance demonstrated by each model during the test process.

Inception-V3, and Xception) in order to provide a single result. The results of Table 4.1 indicate results of the voting system, where we obtained high test accuracy of 85.28%, kappa score of 77.78%, MSE error of 0.3242, macro recall of 69.97%, macro precision of 80.22%, f1-score of 84.33%, weighted specificity of 94%, and macro specificity of 96%,

Figure 4.8 shows the voting model's confusion matrix, while Table 4.1 shows their performance. Our method has produced the greatest results for the non-DR stage (label 0), correctly predicting 171 out of 171 images, as shown in Table 4.1. For the same label, the precision, recall, and f1-score are 0.94, 1.00, and 0.97, respectively. Regarding the moderate level (label 2), our methodology has produced excellent outcomes: Out of 100 photos, 91 are anticipated to be correct. For the moderate level, the corresponding values for precision, recall, and f1-score are 0.74, 0.91, and 0.82. The remaining forecasts for each stage are as follows: 18 for the proliferative-DR stage (label 4) out of 36, 8 out of 15 for the severe stage

(label 3), and 25 for the mild stage (label 1) out of 45.

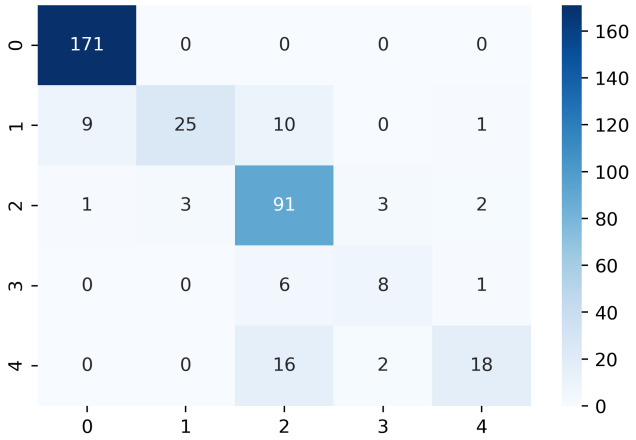

Figure 3.7: The confusion matrix of the test phase after using the voting technique.

Label	precision	recall	F1-score	specificity
0	**0.94**	**1.00**	**0.97**	0.948
1	0.89	0.56	0.68	**0.99**
2	0.74	0.91	0.82	0.880
3	0.63	0.53	0.57	0.985
4	0.82	0.50	0.62	0.987
Accuracy	-	-	**0.85**	-
Macro	0.80	0.70	0.73	**0.96**
Weighted	**0.86**	0.85	0.84	0.94

Table 3.3: The performance obtained in the test phase after using the voting technique.

We compared our proposed approach during this chapter with state-of-the-art approaches. All the presented approaches in this section are validated on the same dataset, and the results are for the testing process. In addition, all the presented results are for the multiclass classification of DR detection. Table 4.2 illustrates the comparison of the current study with state-of-the-art approaches; all those approaches are explained in more detail in Section 2.4.

architecture	Models	Validation accuracy(%)	Test accuracy(%)	kappa score(%)
Arora and Pandey [73]	DNN architecture	74.4	-	-
Dekhil et al. [51]	Hybred CNN	-	77	-
Kassani et al. [74]	Modified Xception	-	83.09	-
	Original Xception	-	79.59	-
	Inception v3	-	78.72	-
Kumar et al. [75]	DRDNet	-	80.59	-
Zhuang et al.[76]	Shallow neural network	67.05	69.03	-
	Re-trained Efficientnet-B3	80.85	77.87	-
Bodapati et al.[77]	Blended multi-modal	-	81.7	71.1
Gangwar and Ravi [78]	Hybrid Inception-ResNet-v2	-	82.18	-
Fan et al. [79]	Multi-scale features	-	85.32	77.26
Proposed method	Ensemble voting	-	85.28	77.78

Table 3.4: The comparison of the performance of our proposed approach with the state of the art during the test phase.

3.5 Conclusion

Providing the appropriate treatment for the patient by ophthalmologists is based on determining the exact levels of DR, in addition to early detection. To address such a problem, we suggest in this paper a new automatic approach based on deformable registration and deep learning for detecting DR severity levels. The deformable registration is used in the preprocessing step to enhance the retina fundus images by deforming the retina to occupy all the black space in the background and to illuminate the effect of the background on the classification process. The enhanced images serve as input to the classification step, in which we used four pre-trained and fine-tuned convolutional neural network architectures (DenseNet-121, Xception, Inception-V3, and ResNet-50) models to classify the retina fundus image, besides the ensemble major voting technique, to determine the exact levels of DR in the testing process. The voting technique consists of participating all the model outputs to make the final decision. We used the APTOS dataset to validate our approach, and we achieved an accuracy score of 85.28% which is on of the best accuracy in the SOTA.

Chapter 4

Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network

4.1 Introduction

One of the most common problems that a person of working age may have is loss of vision, particularly if it is incurable and can only be slowed down through medical monitoring. According to [17], diabetic retinopathy can be considered one of the leading causes of blindness in people of working age. Diabetic retinopathy is a disease developed as a result of either when blood vessels in the eye balloon out, causing them to close and preventing normal blood circulation, or when the fragility of the blood vessels makes the blood leak onto the surface of the retina. As stated in the statistics of the 10th edition of the International Diabetes Federation (IDF) Atlas [80], approximately 537 million adults have diabetes; in particular, those with type 2 diabetes are more susceptible to developing DR. Moreover, 80% of people with more than 20 years of diabetes history were

eventually affected by diabetic retinopathy.

Currently, various studies have been proposed in the field of prediction of diabetic retinopathy, either by improving the retina fundus image or by approving and developing new deep and machine learning models. In this chapter, we propose a new pre-processing technique to enhance the quality of retina fundus images that has never been used in this field, in addition to a new classification model that merges a modified inception block and a hybrid deep learning model with a modified capsule network. The pre-processing step is based on applying the discrete wavelet transformation (DWT) technique to the retina in order to exploit the hierarchy of DWT, which enables us to use three levels of decomposition of the same image, which means collecting more information from the same image. The generated images are fed into a new classification model to predict the exact levels of diabetic retinopathy, which consists of a modified inception block and a modified capsule network that are linked to each other with a hybrid deep learning model.

The remainder of this chapter is organized as follows: the materials and methods that are employed are examined in the following section. The suggested approach is explained in the third section. In section four, experimental results and an assessment of the suggested approach are presented. The chapter's conclusion is found in Section five.

4.2 Materials and Methods

Diagnosing diabetic retinopathy is still a difficult procedure that requires experienced ophthalmologists. In this section, we presented the Materials and Methods utilized in our new proposed approach that combines powerful pre-processing operations and new deep-learning approaches.

4.2.1 Discrete Wavelet Transform

Discrete wavelet transform (DWT) is mostly related to image denoising, which reduces and removes noise from images. In medical imaging, the DWT can be used to decompose the image into many different decomposition levels, which offers a hierarchical level of decomposition and provides more information than any single level can offer.

To benefit from an image's hierarchical decomposition, we use the discrete wavelet transform to provide a multiscale pyramid decomposition. The first decomposition of the image yields four bands, as shown in Figure 4.1, namely LL^1, LH^1, HL^1, and HH^1, where each band has the half-resolution and size of the original image. The four bands of decomposition are represented by: the low-low (LL) band, which is the approximated version of the original image; the low-high (LH) band, which represents the vertical edges; the high-low (HL) band, which performs the horizontal edges; and the high-high (HH) band, which performs the diagonal edges of the original image. In addition, to generate the second decomposition, we use LL^1 in the same way as the original image to generate four other bands, namely LL^2, LH^2, and HL^2, characterized by the half-resolution and size of LL^1.

The discrete wavelet transform principle is summed up with an example of three stages of decomposition in Figure 4.1.

4.2.2 Inception block

The fire appearance of the Inception-V1 architecture, also known as GoogleNet architecture, was described by Szegedy et al.[81], with which they won the 2014-ILSVRC competition. The essential concept of GoogleNet is reducing the processing cost while increasing accuracy, a problem that is resolved by the novel idea of the inception block. As illustrated in Figure 4.2, the inception block

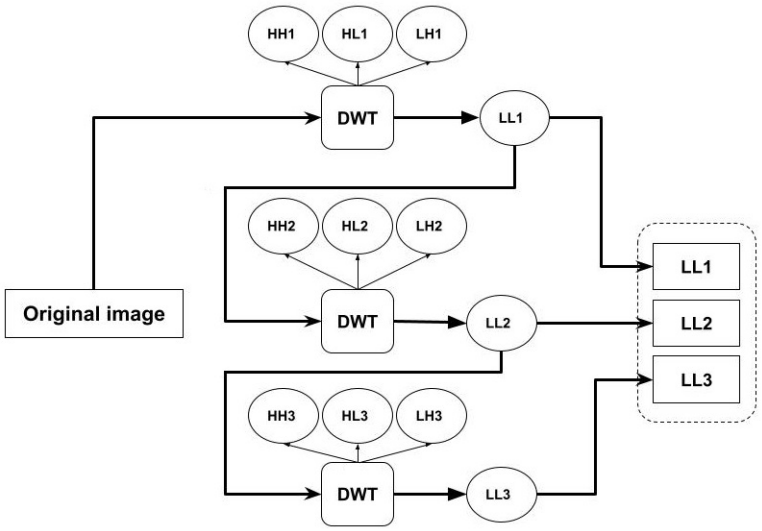

Figure 4.1: The pyramid hierarchy architecture of the discrete wavelet transform.

comprises a Max pooling layer and convolution with multi-size kernels (1x1, 3x3, and 5x5) to combine them into a single vector that will serve as the subsequent input. To reduce the number of trainable parameters, they also layered a layer of 1x1 convolutional filters before the 3x3 and 5x5 convolutional filters. To put it briefly, this idea of the inception block frees us from the restriction of using only one kernel size and lets us utilize several types in a single image.

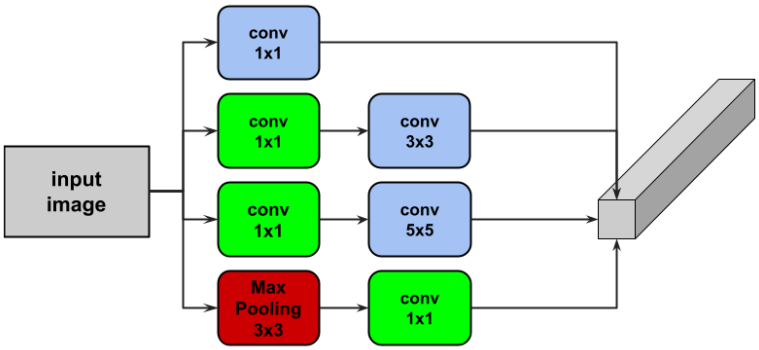

Figure 4.2: Illustrated the original version of the inception block.

4.2.3 Capsule Network

Convolutional neural networks (CNNs) have shown great results in the field of medical image processing, such as image classification, segmentation, and detection. Despite that, CNN faced various issues, such as the loss of information during the pooling operation. in addition to the unaffected CNN by the minor change in the input image. Besides, the CNN failed to distinguish the pose, texture, and deformation in the input image. All those problems were the main causes that led to the proposal of the capsule network concept by Sabour et al.[82]. The capsule network (CapsNet) is a group of capsules arranged to form a network, where each capsule is made of a tiny collection of neurons that each refer to a different component. Unlike standard CNNs, CapsNet keeps track of the location and orientation of each image component based on the new concept of the dynamic routing algorithm. Furthermore, the outcome of the capsule network is represented by a vector instead of a scalar as in the traditional deep learning algorithm. The original capsule network comes with two main objectives: the first is used for the classification task, and it comes with a special loss function to minimize the error. The second objective is reconstructing the input image task, which uses the Euclidean distance between the reconstructed image and the input image as a loss function. In this way, we can divide the original Capsule Network architecture into two parts: the encoder part, which is responsible for the classification task. and the decoder part, which is responsible for reconstructing the input image.

4.3 Proposed approach

In this chapter, we present a novel deep learning-based automated diabetic retinopathy prediction approach to track the disease's progression and predict the disease's

stage. Our suggested method is predicated on taking advantage of the wavelet transform of the retina fundus image's pyramid hierarchy. Precisely, we adopted a DWT algorithm to decompose the retina image into several stages. In our method, we divided the retina fundus image into three levels of wavelet transformation. These three levels serve as the input for our new architecture, which is built around a modified capsule network model on top of a newly modified inception block and a hybrid deep learning model. The entire pipeline of the approach we've suggested is depicted in Figure 4.3.

The proposed approach consists of four main steps, which are: preprocessing, the modified inception block, the hybrid deep learning model, and the modified capsule network, as pointed out in a more detailed way below:

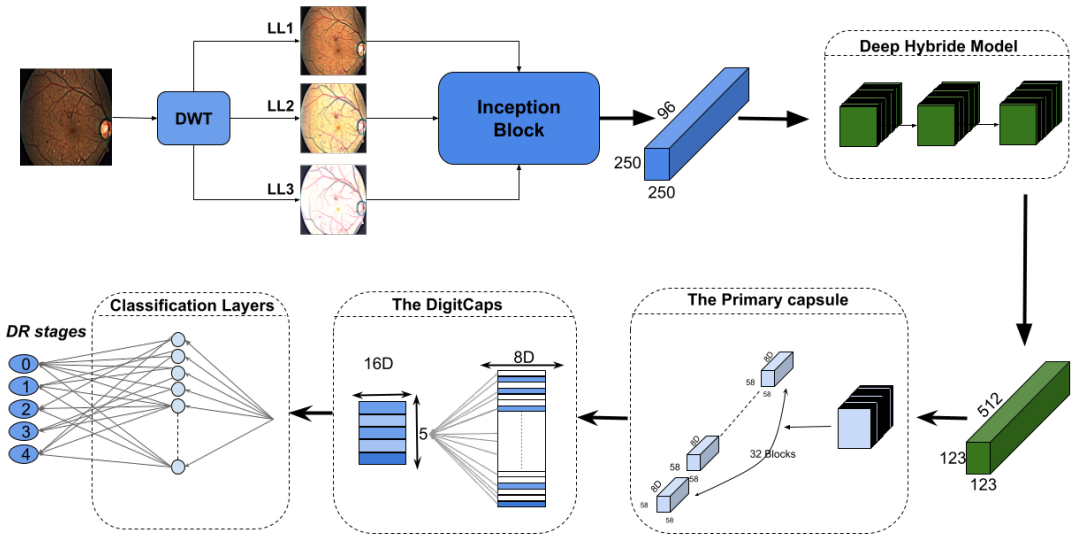

Figure 4.3: An illustration of the overall structure of the proposed approach.

4.3.1 Pre-processing

The pre-processing step started by applying the discrete wavelet transform (DWT) to the RGB image. which is done by first splitting the original image into three two-dimensional channels representing the red, green, and blue channels. We work on each channel of the retina fundus image separately in order to apply the DWT. Besides, we decompose each channel of the original image into three levels of wavelet transformation; each wavelet decomposition is mentioned by round in the following algorithm. In our approach, we acquire three rounds of decomposition; each round's algorithm can be summed up as follows:

Step 1: divided the original image into three channels of the RGB image.

Step 2: subjected the wavelet decomposition to one channel of the original image (let's say, the red channel), this decomposition produced four subband images with a size of half of the chosen channel. These images are referred to as the approximate coefficient (LL^1), detail coefficients diagonal (HH^1), detail coefficients vertical (HL^1), and detail coefficients horizontal (LH^1). Next, we carry out the identical procedure for the remaining channels (green and blue).

Step 3: To reconstruct the RGB image, which is the first level of the wavelet image decomposition, we fuse the approximate coefficient (LL^1) of the three original image channels.

Step 4: Similar to the original image, the approximate coefficient produced in step 3 is chosen as the input for the subsequent wavelet decomposition round.

Step 5: Repeat the algorithm for the two other rounds to generate the second and third levels of wavelet decomposition.

Figure 5.7 illustrates the DWT algorithm utilized in our approach.

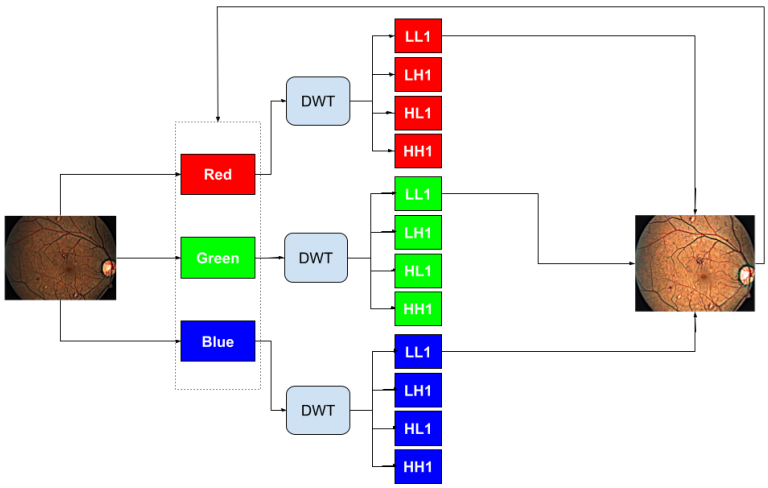

Figure 4.4: An illustration of the overall structure of decomposition rounds used in the proposed approach.

4.3.2 Inception block

Our architecture starts with the modified inception block. The modified inception block takes as input the three RGB images corresponding to three wavelet transformation levels of 250x250 dimensions generated in the pre-processing step. The latter is performed by three internal paths, each consisting of a convolutional layer with 32 channels. As illustrated in Figure 5.10, the first path consists of a convolution with a kernel size of (1x1x8) followed by (5x5x32), reserved for the first levels of the wavelet decomposition image. while the second path reserved for the wavelet decomposition image-level two consists of employing a kernel with a size of (1x1x16) followed by (3x3x32). The third path employs a kernel (1x1x32) reserved for the wavelet decomposition image-level three. Lastly, we combine all three path outputs into a single tensor to serve as input to the squeeze and excitation (SE) block. The SE block architecture is illustrated in Figure 5.10. As we can see, the SE block is comprised of two fully connected layers with ReLU and sigmoid activation, respectively, after a global pooling layer. This modified inception block played a crucial role in integrating the three wavelet levels of de-

composition in the proposed model and performing a convolution with multi-size kernels for spatial sizes and information extraction.

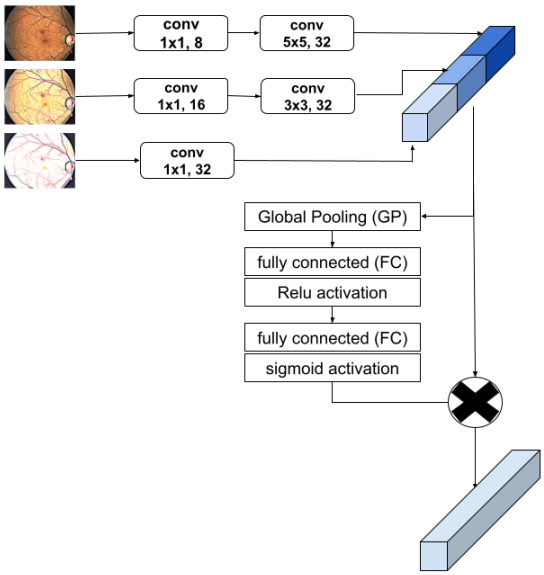

Figure 4.5: A visualization showing the suggested modifed inception block proposed in our approach.

4.3.3 Deep Hybrid model

The following block of the modified inception block is the hybrid deep learning model, which plays an essential role in linking the inception block with the capsule network. This hybrid model receives its input from the modified inception block, which comprises a tensor of size [250, 250, 96] to perform an output of size [123, 123, 512], which will serve as input to the capsule network after a series of convolution and pooling operations. The hybrid deep learning model is arranged as indicated in Table 4.1, which further displays the deep hybrid model that has been suggested in addition to all related information about it.

Table 4.1: Overview of the deep hybrid model.

Type	Patch size	padding	Output size
convolution	3x3	Same	250x250x96
convolution	3x3	Valid	246x246x128
Max Pool	2x2	Valid	123x123x128
batch normalization	-	-	123x123x128
convolution	3x3	Same	123x123x128
convolution	3x3	Same	123x123x256
Max Pool	2x2	Same	123x123x256
convolution	3x3	Same	123x123x256
batch normalization	-	-	123x123x256
convolution	3x3	Same	123x123x256
convolution	3x3	Same	123x123x512
Max Pool	2x2	Same	123x123x512
convolution	3x3	Same	123x123x512
batch normalization	-	-	123x123x512

4.3.4 Capsule network

The subsequent step of the deep hybrid model is the modified capsule network. This network is performed by the primary capsule layer section, the DigitCaps layer section, and the classification layer section, respectively.

4.3.4.1 Primary capsule:

The proposed modified capsule network starts with the primary capsule, which takes its input from the deep hybrid model's output. The process of the primary capsule is summarized in the convolution operation, reshape function, and squash function, respectively. The convolution operation consists of applying the convolution to a kernel of size 9x9 with a stride of 2 on the input tensor. Then the reshaping process, which comprises reshaping the output of the previous convolution [58, 58, 256] into 32 blocks of eight dimensions (58*58*32 = 107648), results in 32 groups of capsules with an 8-dimensional capsule for each group,

as illustrated in Figure 4.6. The last process is the squashing function. The primary goal of this formal function is to assist in the reduction of large vectors to unit vectors and tiny vectors to zero in order for each vector to represent the probability that the object is located or not in the given image.

The squashing function is computed using the expression (Eq4.1), where v_i is the output vector of a capsule j and S_i is the weighted sum of its total input calculated as presented in the equation (Eq4.2).

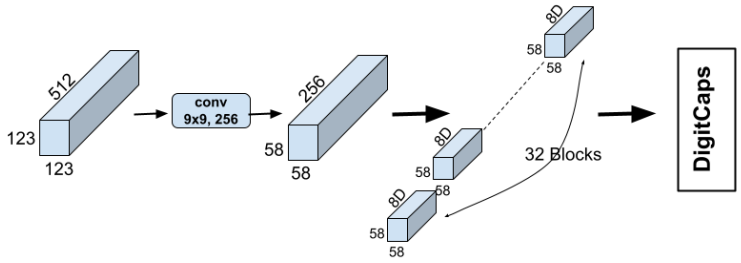

Figure 4.6: An illustration of the Primary capsule layer section.

$$v_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \frac{S_j}{\|S_j\|} \tag{4.1}$$

$$S_j = \sum_i c_{ij} \hat{u}_{i|j} \tag{4.2}$$

Where, c_{ij} is the coupling coefficient and $\hat{u}_{i|j} = W_{ij} u_i$ is the prediction vector.

4.3.4.2 DigitCaps:

The next part of the main capsule is called the digitCaps layer. digitCaps is using the dynamic routing algorithm between the primary capsule and digitCaps, which was proposed in the original capsule network paper by Sabour et al.[82]. The primary capsule section provides a group of capsules of the dimension [107648,8-D] as input for this section's. A dynamic routing algorithm is used to calculate

this layer's output. The result of this section is a vector of dimension [5, 16D], as shown in Figure 4.7, where 5 denotes the five phases of diabetic retinopathy and 16 denotes the length of each of the five vectors. The dynamic routing algorithm applied between the PrimaryCaps and the DigitCaps layers as presented in the original paper is illustrated below:

Algorithm 1 Routing algorithm.

PROCEDURE ROUTING($\hat{u}_{j|i}, r, l$)
for all capsule $i$in layer l and capsule j in layer $(l+1) : b_{ij} \leftarrow 0$.
for r iterations do
for all capsule i in layer $l : c_i \leftarrow softmax(b_i)$
for all capsule j in layer $(l+1) : S_j \leftarrow \sum_i c_{ij}\hat{u}_{j|i}$
for all capsule j in layer $(l+1) : v_j \leftarrow squash(S_j)$
for all capsule i in layer l and capsule j in layer $(l+1) : b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i}, v_j$

return v_j

Where W_{ij} is the weight.

4.3.4.3 Classification layer:

The last section of the modified capsule network is the classification section. This section receives its input of dimension [5, 16D] from the output of the digitCaps section. The main objective of this section is to predict the right labels for the dataset image. This section starts with a flatten function to represent the input of dimension [5, 16D] into a single tensor of size 80 (5*16 = 80). followed by a fully connected layer with 64 units and a ReLU activation function. In addition, to prevent overfitting, a dropout function with a probability of 50 is employed. Finally, a five-unit classification layer refers to the five labels of the dataset, with a softmax activation function (Figure 4.7).

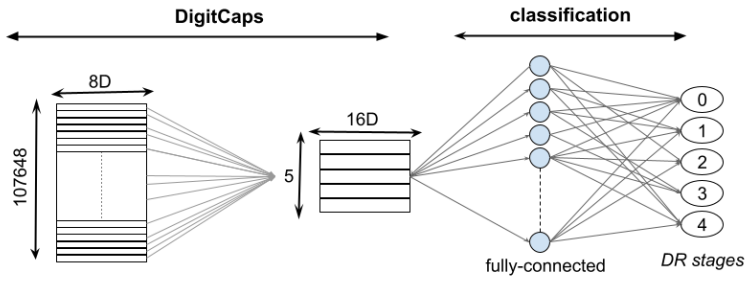

Figure 4.7: An illustration of the digitCaps layer and classification layer sections.

Table 4.2: The performance demonstrated by each model during the test process.

DR stage	Label	Image per label	train/valid set	Test set
No DR	0	1805	1512	293
Mild NPDR	1	370	313	57
Moderate NPDR	2	999	872	127
Severe NPDR	3	193	161	32
Proliferative DR	4	295	254	41
APTOS dataset	-	3662	3112	550

4.4 Experimental results and dataset

4.4.1 Dataset

We validated the performance of the proposed approach during this chapter using the APTOS [26] dataset, which is widely used in the task of detecting DR. In addition, it is a public dataset and available on the Kaggle platform. The dataset is explained in more detail in Section 2.2.

The dataset was divided as follows: 85% of images reserved for training and validation, 15% of images used for testing process. The test set images were not used in either the training or validation phases, as shown in Table 4.2.

4.4.2 Evaluation Metrics

The performance of our proposed work was evaluated based on several metrics, such as accuracy, f1-score, recall, kappa score, and precision. All the measure metrics used in the chapter are explained in more detail in the section 2.3.

4.4.3 Experimental results

To prove the effectiveness of our approach in the task of the diabetic retinopathy detection task and to demonstrate our achieved results are among the best results achieved in this field, we used several measurement metrics in addition to comparing our results with the state-of-the-art.

During the training process, our approach has shown a high accuracy score of 96.71% and an MSE error of 0.0727. Besides, our method yielded an accuracy of 85.25% and an MSE error of 0.3333 during the validation phase. During the test phase, our approach performed great results, where we achieved an accuracy score of 86.54%, an MSE error of 0.2363, besides a kappa score of 78.77%, in addition to a micro f1-score of 87% and a macro f1-score of 73%. Additionally, Table 4.3 demonstrates that the model attained a weighted precision of 86%, a weighted recall of 87%, and a macro precision and recall of 76% and 70%, respectively.

Table 4.3: The performance obtained during the test phase.

Label	Precision	Recall	F1-score	specificity
0	0.99	0.99	0.99	0.98
1	0.73	0.65	0.69	0.97
2	0.72	0.87	0.79	0.89
3	0.65	0.41	0.50	0.98
4	0.80	0.59	0.68	0.98
Accuracy	-	-	0.87	-
Macro	0.76	0.70	0.73	0.96
Weighted	0.86	0.87	0.86	0.96

In the same regard, Figure 4.8 illustrates the confusion matrix of our methodology during testing. Our method earned the highest score for the non-diabetic retinopathy stage (label 0), correctly predicting 291 out of 293 images. As for the moderate stage (label 2), it has shown good results, with a correct prediction of 111 images out of the 127 images that were expected. Subsequently, 37 out of the 57 images that were searched for in the mild stage of DR (label 1) were accurately predicted. The remainder of the prediction for each class ranges as follows: 24 for the proliferative DR (PDR) out of 41 and 13 out of 32 for the severe DR stage.



Figure 4.8: The confusion matrix during the test process of the proposed approach..

We compared our achieved results in the testing phase with several other approaches that have been working on the same APTOS dataset in order to show the performance of our approach. All those formal approaches, included in the table of comparison, are discussed in the Related Work section. It is noteworthy that we observed that our strategy yielded better results than any

other approach. The above comparison is summarized in Table 4.4.

Table 4.4: The comparison of the performance of the current study with other relevant earlier research during the test phase.

Architecture	Approach	years	Test accuracy
Dondeti et al.[83]	NASNet + t-SNE + SVM	2020	77.90%
Gangwar et al.[9]	Hybrid Inception-ResNet-v2	2021	82.18
Bodapati et al.[63]	deep neural network + gated-attention	2021	82.54%
Kassani et al.[74]	Modified Xception	2019	83.09%
Islam et al.[40]	Supervised contrastive learning	2022	84.364%
AL-ANTARY et al.[5]	Multi-scale attention network	2021	84.6%
Rao et al.[52]	InceptionResNetV2	2020	85.02%
Fan et al.[79]	Multi-scale features	2021	85.32%
Oulhadj et al.[13]	Deformable registration + deep learing	2022	85.28%
Majumder et al.[84]	Xception multitask model	2021	86%
Proposed	Capsule network + inception block + DWT	2022	**86.54**%

4.5 Conclusion

To sum up, in this chapter, we aim to overcome the challenge of DR diagnosis by proposing a new system that can assist doctors in detecting DR severity levels. Therefore, we proposed a novel architecture that is based on a modified inception block, a capsule network, and a discrete wavelet transform (DWT) combined in a way that could not be more appropriate as far as achieving the goal of this study. Specifically, the three levels of DWT decomposition enabled us to extract more information from each image than from a single image. We were able to take advantage of the three decomposed images because the new modified inception block with the three input layers enabled the model to operate on the three decomposed images. While the capsule network with classification layers at the end provided an excellent outcome due to the dynamic routing principle, the notion of utilizing a deep hybrid model to link the modified inception block to the capsule network was a major advancement toward reaching such encouraging and promising results. This novel combination, to the best of our knowledge, indeed

made us obtain one of the highest accuracies attained in this field, represented by a training accuracy score of 96.71% and a testing accuracy of 86.54%. Moreover, compared to more commonly proposed approaches on the same dataset, our approach proved itself not only performant but also accurate in fulfilling the required mission.

Chapter 5

Advanced Deep Learning Architectures for Diabetic Retinopathy Prediction.

5.1 Introduction

Ophthalmologists rely on manual methods to examine patients who suffer from DR by analyzing the retina fundus image. This manual method has many drawbacks as time-consuming, as mentioned by [5], the manual method takes over five minutes to grade retinal images by an ophthalmologist. Therefore, it is crucial to propose reliable automatic detection approaches based on deep learning for DR grading.

In the context of aiding doctors in detecting and diagnosing diabetic retinopathy (DR) and supporting their decision-making processes, numerous approaches have been proposed. These approaches have demonstrated their effectiveness in the healthcare field, achieving favorable results. Many researchers have focused on enhancing retinal fundus images to facilitate disease detection, as evidenced

by studies such as those by Fu et al.[85], Sule et al.[86], and Raj et al.[87]. Additionally, others have introduced new deep learning-based approaches to enhance the accuracy of disease detection, as demonstrated in studies by Menaouer et al.[88], Venkaiahppalaswamy et al.[89], and Alshawabkeh et al.[90]. Recent advancements in deep learning techniques, such as the Capsule Network and Vision Transformers, have been leveraged in medical image classification, as highlighted in studies by Oulhadj et al.[14], Kumari et al.[91], Kalyani et al.[92], Dai et al.[93], Gu et al.[94], and Chetoui et al.[95].

In this chapter, two architectures for diabetic retinopathy detection have been proposed based on Capsule Network and Vision Transformers, and a comparison between these studies and the state of the art has been done. Each study is described in more detail in the two subsections 5.3 and 5.4.

5.2 Materials and methods

Diagnosing diabetic retinopathy remains a difficult procedure that requires skilled ophthalmologists to examine the retina fundus image. In this section, we will introduce the materials and methods that have been utilized in order to propose a new architecture that has never been proposed before and help us produce an effective and powerful architecture for the task of DR severity level detection.

5.2.1 Power Law Transformation

The power law transformation, also referred to as the gamma transformation, serves as an enhancement technique applied to images to improve contrast and accentuate details. Several studies, including those by Tsai et al.[96], Kumar et al.[97], Solano et al.[98], and Aswathi et al.[99], have demonstrated the utility of power law transformation for image enhancement. Equation (Eq 5.1) represents

the general form of this transformation.

$$S = c * r^{\gamma} \tag{5.1}$$

S and r are the output and input pixel values, respectively. c and γ are positive constants.

5.2.2 Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image preprocessing method designed to enhance the overall contrast of an image. It also effectively enhances the visibility of details in noisy images, which is particularly important for classification tasks. CLAHE finds extensive application in image enhancement processes across diverse domains, including medical imaging, as evidenced by studies such as those by Khan et al.[49] and Mondal et al.[10], where it has demonstrated beneficial results.

5.2.3 Vision Transformer

Vision Transformer (ViT) is a novel type of transformer architecture utilized for various computer vision tasks, such as image classification and image detection. It was first introduced in the paper titled "An Image is Worth 16 $\times$ 16 words: Transformers for image recognition at scale" by Dosovitskiy et al.[100].

Unlike traditional convolutional neural networks (CNNs), ViT is based on self-attention mechanisms and processes images by dividing them into a sequence of patches, where each patch is attached to its spatial position to provide spatial information. These patches are then subject to a transformation encoding process. Ultimately, to classify the learned image, a multi-layer perceptron (MLP) is utilized. Additionally, ViT can undergo pre-training on a large dataset and

subsequent fine-tuning on smaller datasets for specific classification tasks.

5.2.4 Capsule Network

Similar to ViT, the capsule network represents a significant advancement in various image-related tasks, including image classification, segmentation, object detection, and image reconstruction. Comprising a set of capsules, which are essentially ensembles of neurons, which is based on dynamic routing between capsules. Introduced in the seminal paper by Sabour et al.[82], CapsNet addresses several challenges that conventional CNN architectures struggle with, such as the loss of significant information during pooling operations. Unlike CNN outputs, which can remain largely unaffected by minor changes in inputs, CapsNet outputs are represented by vectors containing a wealth of information, including the orientation and location of each image.

The original Capsule Network architecture, as proposed by Sabour et al.[82], consists of two main parts: the encoder and the decoder. The encoder is responsible for the classification task and comprises three subsections: convolution, PrimaryCaps, and DigitCaps layers. On the other hand, the decoder is tasked with image reconstruction.

5.2.5 Adjust the image contrast and brightness

Enhancing the intensity and appearance of each pixel in the image will improve its image contrast and brightness. To achieve it, we used the sharpening process to improve the retinal fundus images' appearance and highlight the details to make them look crisper and more refined.

The process of sharpening and enhancing image details involves applying a high-pass filter to the input image, which clarifies the edges and provides more

information. Additionally, it sharpens and defines the input image to provide more information so that the training process can use it.

5.2.6 Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module (CBAM) is an attention mechanism that can be used to enhance CNN performance. Woo et al. [101] introduced the CBAM for the first time in their paper "Cbam: Convolutional Block Attention Module.". By integrating information from both spatial and channel levels of the images and focusing on significant regions, the CBAM mechanism has the potential to significantly enhance the overall outcome. The CBAM mechanism comprises two main components: the spatial attention module and the channel attention module, as illustrated in Fig. 5.1.

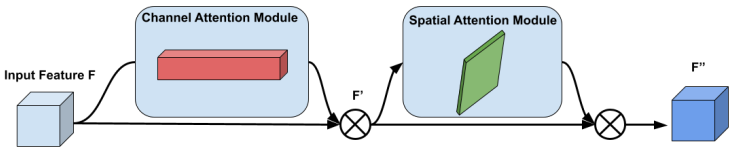

Figure 5.1: Illustration of the Convolutional Block Attention Module mechanism.

The CBAM mechanism initiates by computing the new feature map of the Channel Attention Module, denoted as $F^{'}$, through element-wise multiplication of the input features (F) with the channel attention map $M_c(F)$, as depicted in equation (5.2):

$$F^{'} = M_c(F) \otimes F \tag{5.2}$$

The CBAM generates the final feature map F'' by passing F' through the Spatial Attention Module to create the spatial feature maps $M_s(F')$. This is followed by an element-wise multiplication of $M_s(F')$ and F', as demonstrated in

equation (Eq5.3), to calculate the final feature map.

$$F'' = M_s(F') \otimes F' \tag{5.3}$$

The channel attention module takes the input feature F and processes it through both max-pooled and average-pooled features. The resulting output is then passed through a multi-layer perceptron (MLP), followed by element-wise summation multiplication and the sigmoid activation function to produce the final output attention feature maps ($M_c(F)$). Subsequently, to compute the new feature map of the Channel Attention Module F', we perform a multiplication of the input feature F and $M_c(F)$, representing the input of the Spatial Attention Module (as shown in Fig.5.2).

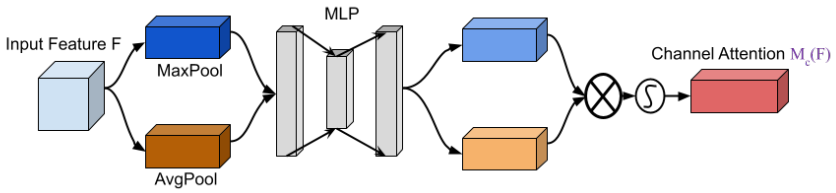

Figure 5.2: Illustration of the Spatial Attention Module architecture.

The spatial attention module takes its input from the channel attention module, denoted as F'. To generate the final feature maps F'' of the spatial attention module, average pooling and max pooling operations are applied, and their results are concatenated to create an effective feature descriptor. Subsequently, a convolutional layer is applied to compute the spatial attention feature $M_s(F')$, which is then multiplied with F' to produce the final new feature maps F'' (as illustrated in Fig.5.3).

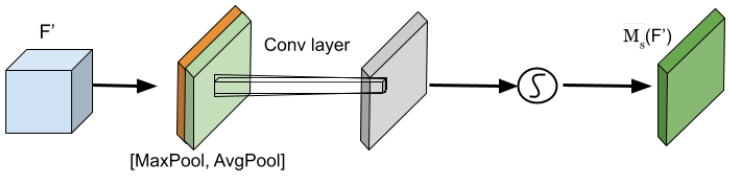

Figure 5.3: Illustration of the channel attention module architecture.

5.2.7 Squeeze and Excitation

The Squeeze and Excitation Block (SE) was designed for the purpose of improving inter-channel relationships within feature maps. This innovation was first introduced in the 2018 paper "Squeeze-and-Excitation Networks" by Jie Hu et al.[102].

The SE architecture begins with the squeeze phase, where each input channel is condensed into a single scale using global average pooling, resulting in a vector of size n that represents the number of channels. Following this, the excitation phase focuses on capturing inter-channel relationships, producing a vector of size n that represents the weights of each channel.

In the final phase of the SE architecture, the generated weights are applied to the original channels, effectively preserving important features (as depicted in Fig.5.4).

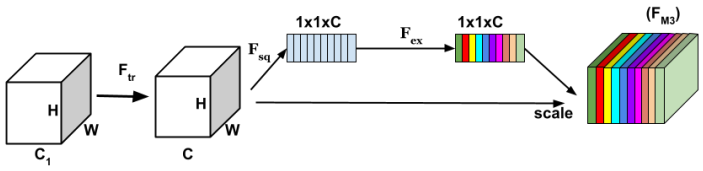

Figure 5.4: The Squeeze and Excitation Block.

5.2.8 Inception Block

The inception block was first introduced in the proposal of the inception-V1 architecture by Zegedy et al.[81], which led to their victory in the 2014 ILSVRC competition.

The principle of the inception block involves the use of multiple filters with different kernel sizes, including (3×3, 5×5, and 1×1) parallel convolution kernels, along with a Max pooling layer. These outputs from the convolution and Max pooling layers are then concatenated into a single tensor using a concatenation layer.

5.3 Diabetic Retinopathy Prediction Based On Vision Transformer and Modified Capsule Network

In this section, we propose a new approach for diagnosing the severity stage of diabetic retinopathy that can assist doctors in the task of detecting this disease. Our approach is based on a new combination of deep learning methods and a new pre-processing process. The pre-processing process consists of generating two enhanced images from a single retina fundus image to feed it to a new proposed hybrid deep learning technique. The hybrid deep learning technique is based on a fine-tuning vision transformer, a modifier capsule network, and a classification model to classify the retina fundus image into the corresponding stage. Our suggested methods' entire pipeline is shown in Figure 5.5. As we can see, the first image (Img1) generated in the pre-processing step is fed directly to the fine-tuning vision transformer, while the second image (Img2) is fed to the modified capsule network. Besides, a classification model combines the features obtained

from the two early submodels to predict the stage of DR.

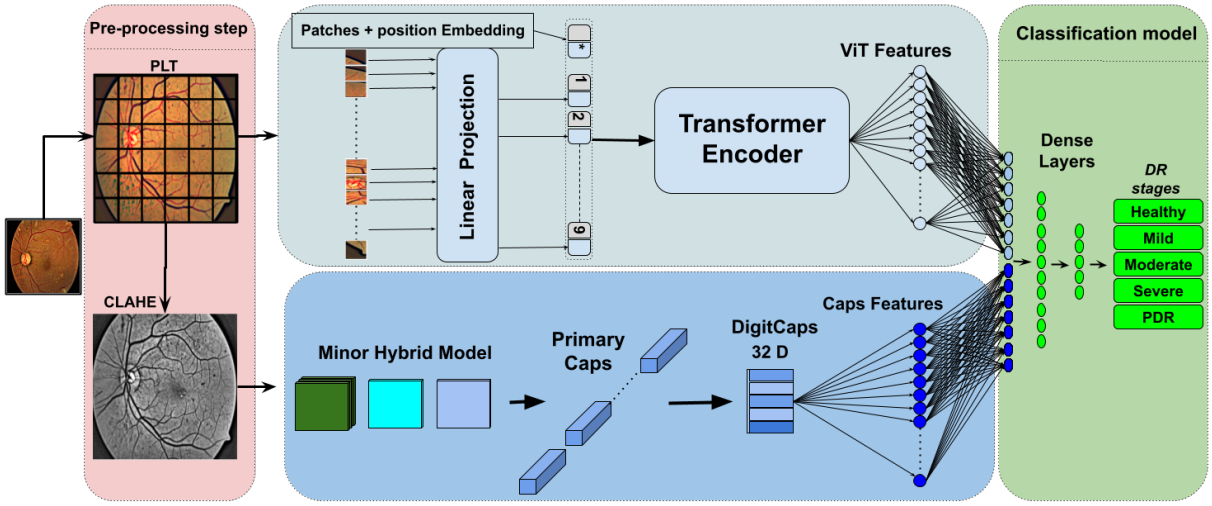

Figure 5.5: An illustration of the overall structure of the proposed approach.

5.3.0.1 Preprocessing

The pre-processing process is necessary because it directly affects the detection and classification tasks. In addition, the images in the dataset have varying sizes and types of noise, as well as being captured under various lighting conditions. Consequently, there needs to be a preprocessing stage for the image dataset. In this approach, the retina fundus images passed through a series of preprocessing operations, as illustrated below and represented in Figure 5.6.

First, the RGB image is subjected to the power law (gamma) transformation as part of the pre-processing step. This transformation enhances the contrast of retinal fundus images and provides better contrast and more details. Gamma transformation is also a helpful technique for enhancing blood vessel clarity and managing an image's overall brightness [103], which is crucial for unifying the contrast and brightness of all images. Gamma transformation is also required to provide uniform image display across all images and to preserve image appearance

and quality. The general form of this transformation is presented in the previous section (Eq1). At this stage, we generated the first image, Img1.

Second, we isolate the green channel (G) by dividing the first generated image into three two-dimensional images representing the red, green, and blue channels. Since the green channel provides more information than the other channels (red and blue), we used only the green channel for the next preprocessing step. Biswas et al.[104] state that DR may be indicated by a change in central retinal blood vessels (CRBVs), and the CRBV can be in excellent condition in the G. Moreover, the extreme contrast in the green channel makes the microaneurysms highly visible, as reported by Indumathi et al. [105]. Usually, the earliest discernible signs of diabetic retinopathy are microaneurysms.

Third, we apply contrast-limiting adaptive histogram equalization (CLAHE) to the improved green channel image that was produced in the previous stage. Which plays a crucial role in improving image contrast and resolving the noise amplification issue as noted by Tharsanee et al.[106]. We performed the CLAHE operation with a ClipLimit of 4.0, using the OpenCV library. We have now produced the second image, Img2.

Fourth, we crop the two bands of images we created at this point (Img1 and Img2). We used the cropping operation to remove and ignore the surrounding black area of the retina fundus image. Furthermore, we resize Img1 to (224, 224, 3) and Img2 to (120, 120, 1). We have decided to reduce the size of Img2, which will feed the capsule networks, because the number of parameters relating to the capsule networks grows exponentially with the size of the input image.

5.3.0.2 Fine-tuning Vision Transformer

The Vision Transformer that we used in our method starts by transforming the input image into a set of patches; in our case, we used Img1 which was generated

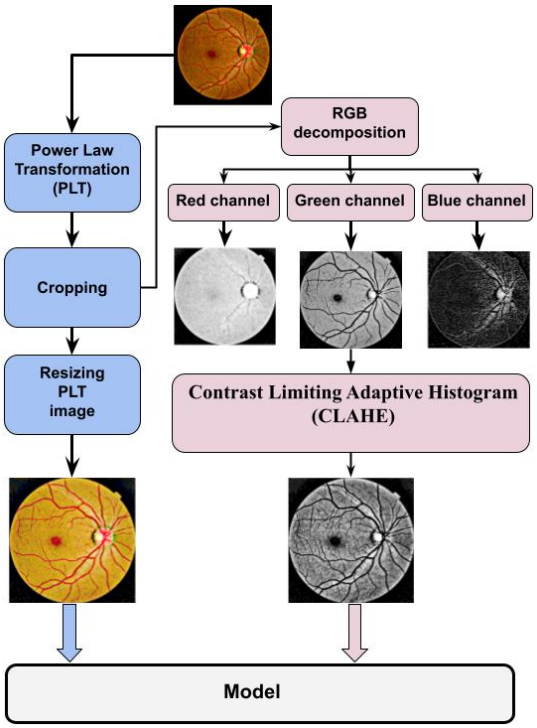

Figure 5.6: The structure of the overall proposed pre-processing step.

in the pre-processing step. During the process of creating the set of patches for each image, we take into account the relative position of each patch in the image by attaching each patch to its spatial information. After that, the transformer encoder block receives the pair (or sequence) of patches and their spatial information to apply various operations, such as the self-attention mechanism and multi-head attention. The self-attention mechanism (SA), which is regarded as the most critical component of the ViT architecture, is used to pay attention to significant parts of the input sequence while considering their relationships. The SA mechanism is based on multi-head attention to carry out their operation and perform the output.

The multi-head attention inputs are converted into three different types of representations, which are query (Q), key (K), and value (V), before being received by the multi-head attention. To produce these variables, the inputs are multi-

plied by weighted matrices. During the attention process, many operations are performed, such as the dot product between the query and the key vectors. The outcome is then scaled and passed via a SoftMax function. The entire multi-head attention system is displayed in more detail in Figure 4.3.

Our method used a pre-trained model that we fine-tuned on our dataset, rather than training the ViT model from scratch. Where the classification layer is removed and the model weights are frozen during the training process. The image of size [224, 224, 3] was split into a grid of 16x16 pixel patches and a 768 pixel embedding vector. Our ViT model consists of 12 multi-head attentions. Besides, we added a flatten layer and then a fully connected layer with 128 units and a ReLU activation function, which is the input for the fusion model.

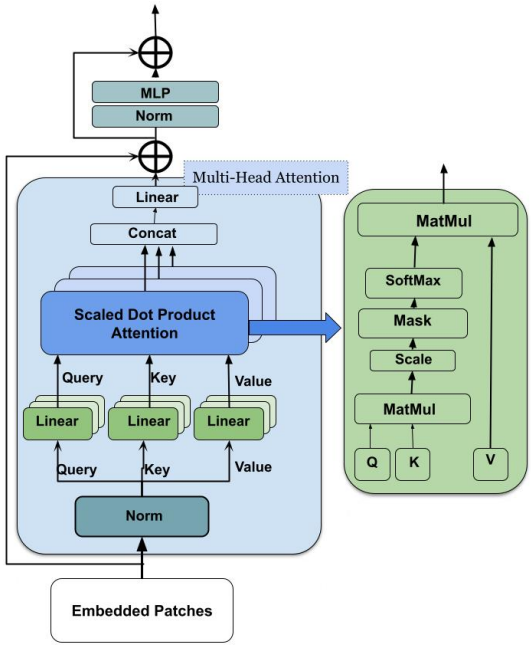

Figure 5.7: The Multi-Head Attention Mechanism.

5.3.0.3 Modified Capsule Network

The proposed modified capsule network is mainly focused on the encoder part of the capsule network, which is responsible for the classification task. Indeed, our provided modified capsule network consists of three subsections, which are: the minor hybrid model, primary caps, and DigitCaps. The image produced by CLAHE during the pre-processing stage is sent into the minor hybrid model, which is regarded as the initial component of the capsule network, with a shape input of 120x120x1. The minor hybrid model consists of three convolution layers, followed by a Max Pooling layer, and is finished by a batch normalization layer. The three convolution layers have kernel sizes of 3x3, 5x5, and 9x9, with kernel number channels of 96, 128, and 256, respectively. Furthermore, the Max Pooling layer's kernel has dimensions of 2 by 2.

The following subsequent is the PrimaryCaps, which receives information from the Minor Hybrid Model. This is based on three different techniques: convolution, reshape function, and squash function. The input tensor of this subsequent is subjected to a convolution operation with a kernel number channel, size, and stride of 256, 9x9, and 2, respectively. The next step is the reshaping function, which involves creating 32 blocks of eight dimensions out of the tensor created in the previous section. After that, a squashing function is used, and it is calculated as shown in equation (eq 2). Which consists of shrinking small vectors to zero, which represents the probability of that object being present in this vector, and shrinking large vectors to one. According to the initial version in the original paper by Sabour et al.[82], the v_i is the capsule j output vector. S_i is the weighted sum of its total input, which is determined as shown in the equation (eq 3).

$$v_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \frac{S_j}{\|S_j\|} \tag{5.4}$$

$$S_j = \sum_i c_{ij} \hat{u}_{i/j} \tag{5.5}$$

c_{ij} is the coupling coefficient. The $\hat{u}_{i/j}$ represents the prediction vector and is calculated by the following equation (eq4):

$$\hat{u}_{i/j} = W_{ij} u_i \tag{5.6}$$

where W_{ij} is the weighted.

The information from the PrimaryCaps subsection is transferred to the third subsection, digitCaps, via the dynamic routing mechanism as described below, which was first introduced in the original publication. The outcome of this section is represented by five capsules with dimensions of 16D. The five capsules represent the five levels of diabetic retinopathy in the dataset. These vectors were flattened into a dimension of [160, 1] by adding a flattened layer before feeding this capsule to the classification model.

The dynamic routing algorithm used between the two successive layers, named PrimaryCaps and DigitCaps, was published for the first time by Sabour et al.[82], as presented in Chapter 4, Subsection 4.3.4.2.

5.3.0.4 Classification model

The suggested classification model plays a critical role in classifying the retina fundus image according to the severity of diabetic retinopathy. The input of the classification model is composed of a layer that concatenates the two feature tensors produced by the ViT model and the Capsule network model, where each tensor is of size [160, 1]. Indeed, the layer that serves as the input for this subsection is of size [160*2, 1]. Next, using the ReLU activation function, four completely connected layers with 2048, 1024, 256, and 128 units are used. In

addition, a dropout layer was added to avoid overfitting with a probability of 50%. Followed by another fully connected layer with 64 units and ReLU activation. Finally, a classification layer with a softmax activation function was employed to categorize the input features. Figure 5.8 provides a more detailed presentation of the entire classification procedure.

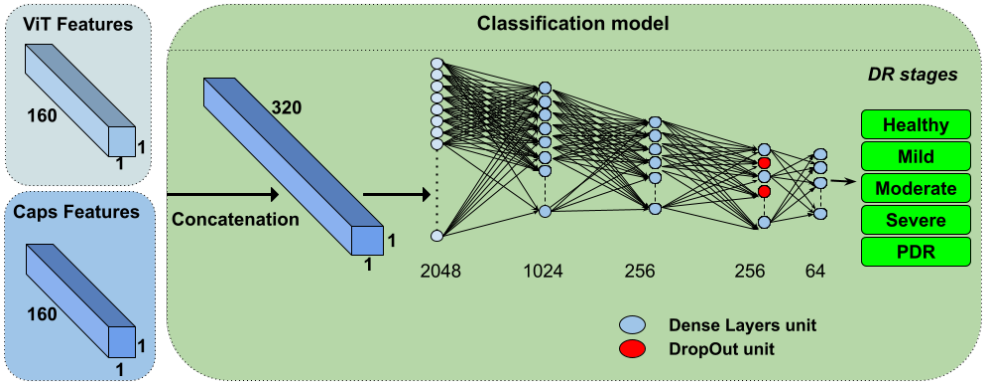

Figure 5.8: Illustration of the proposed classification model architecture.

5.4 Hybrid Deep Learning Architecture For Diabetic Retinopathy Prediction Using Medical Images

In this section, we aim to propose a new approach for diabetic retinopathy diagnosis based on a new hybrid deep-learning model. This approach starts with a pre-processing step, which consists of adjusting the image contrast and brightness in order to make sure that all retina fundus images are suitable for the classification process before feeding them to the hybrid mode. The hybrid deep-learning model is based on several deep-learning techniques to identify the DR that patients suffer from and detect the stage of DR they are experiencing.

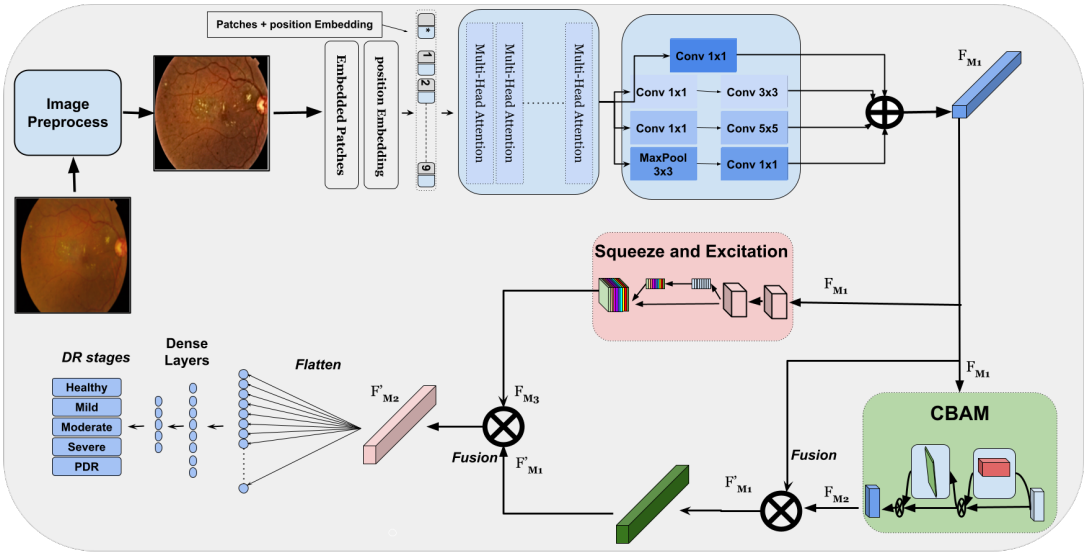

Figure 5.9: An illustration of the overall structure of the proposed approach.

5.4.1 pre-processing

The pre-processing phase of our suggested methodology includes the following techniques:

We begin with the retinal image sharpening approach and detail improvement. To accomplish this, we first used OpenCV's detailEnhance function to highlight and improve the details in the retinal fundus image. Second, use high-pass filters to enhance the brightness and contrast of the image.

Then, we remove the black space surrounding the retina in the image using the cropping operation. It enables us to ignore and neglect its impact on the training procedure.

After that, since the dataset images differ in size, we resize all the image datasets to the same size.

5.4.2 Proposed deep Hybrid model

We fed the generated enhanced image during the pre-processing step to our suggested model as input, which starts with the ViTIncep sub-model. The ViTIncep input has the shapes (H, W, C), which represent the height, width, and channel. The ViTIncep first divides the input that was received into a sequence of patches $X_t = 1, 2, ..., M$, using a patch size of p = 16, where M is the number of patches, and calculates it using $M = H * W/p2$. To preserve each patch's localization in the original image X, we indicate that each patch is attached to its spatial position. Subsequently, this series of patches is fed to the transformer encoder, which is composed of several layers of multi-head self-attention (MHSA). The purpose of the MHSA is to highlight significant parts of the input sequence and keep track of their relationships. The input of each single layer of multi-head self-attention is first projected into Query (Q), Key (K), and Value (V) by multiplying it with the weighted matrix, Fig.5.9.

The features generated by the multi-head self-attention are fed to a deep learning block that consists of four parallel units. The first unit is a convolution layer with 16 kernels of size (1,1). The second unit is made up of two convolutional layers with 16 kernels of size (1,1), followed by 32 kernels of size (3,3). Next, two successive convolution layers with 8 and 16 kernels of sizes (1,1) and (5,5), respectively, make up the third unit. The final unit consists of a convolution layer with 16 kernels of size (1,1) after a max-pooling layer with a kernel size of (3,3). The results of the four units are combined at the end of this block by concatenating all of the units using a concatenate layer. At this point, we generate a new feature map that offers and conserves all the spatial information of the retinal fundus image F_{M1}.

The Squeeze and Excitation Network (SE) block is also used to improve and enhance the features generated from the ViTIncep. The Squeeze and Excitation

Network block, as described in the Materials and Methods section, starts with the squeeze phase, which is based on average pooling to squeeze each channel to a single numeric value; next, the excitation phase, which detects the relationships and dependencies between channels; and the final phase, which consists of scaling each channel of the input to highlight the most significant feature in the input tensor F_{M3}.

We used a fusion technique to combine and concatenate the feature maps extracted from both the CBAM block and the SE block. This fusion technique helps us combine the spatial information with the channel level and the dependencies between channels. In the first stage, we used the Add operation to fuse the feature maps produced by the ViTIncep sub-module F_{M1} with the CBAM block F_{M2} to create a new feature map F'_{M1} (5.7).

$$F'_{M1} = F_{M1} + F_{M2} \tag{5.7}$$

In the second stage, we combined the F'_{M1} feature map and the feature map produced from the SE block F_{M3} to produce a new feature map F'_{M2} (Eq5.8).

$$F'_{M2} = F'_{M1} + F_{M3} \tag{5.8}$$

We fed the feature generated (F'_{M2}) using the fusion technique to a classification sub-model in order to perform the DR severity level detection task. The classification sub-model consists of several layers, starting with a flatten layer followed by a batch normalization layer. After that, there is a dense layer using the "ReLu" activation function, followed by a dropout layer to avoid overfitting, and a batch normalization layer. Afterward, two successive dense layers with 128 and 64 units and the 'ReLu' activation function were followed by a dropout layer and a batch normalization layer. In the end, we employed a dense layer with five

units for the DR severity level detection task, using Softmax as the activation function.

5.5 Experimental results and dataset

This section is reserved for pressing the used datasets, the evaluation metrics used to evaluate the two proposed approaches, and the experiment results obtained for each approach.

5.5.1 Dataset

We validated the performance of the proposed approach during this chapter using four datasets, which are the APTOS dataset, the EyePACS dataset, the Messidor-2 dataset, and the DDR dataset. These datasets are widely used in the task of detecting DR. In addition, they are public datasets and available. All Theses datasets are explained in more detail in Section 2.2.

All these datasets were divided as follows: 85% of images are reserved for training and validation, and 15% of images are used for the testing process. The test set images were not used in either the training or validation phases.

5.5.2 Evaluation Metrics

The performance of our proposed work was evaluated based on several metrics, such as accuracy, f1-score, recall, kappa score, and precision. All the measure metrics used in the chapter are explained in more detail in the section 2.3.

5.5.3 Experimental results

This section is dedicated to explaining, evaluating, and examining the performance of the two proposed approaches, employing multiple metrics across four datasets: APTOS, EyePACS, Messidor-2, and DDR.

5.5.4 Diabetic Retinopathy Prediction Based On Vision Transformer and Modified Capsule Network

We have trained, validated, and tested this approach on the four datasets (APTOS, EyePACS, Messidor-2, and DDR) for the problem of grading diabetic retinopathy using retinal fundus images.

On the APTOS dataset, our approach demonstrated a compelling training accuracy of 97.49%. During the testing phase, our approach delivered impressive results, supported by various performance metrics. Notably, we achieved an accuracy of 88.18%, an MSE error of 0.2018, and a Kappa score of 81.55%. Furthermore, the weighted precision, recall, and f1-score were all at 88%, as detailed in Table 5.1.

Evaluating the performance of our approach on the Messidor-2 dataset, we attained a notable accuracy score of 87.78%. The effectiveness of our approach was further demonstrated through other metrics, including an MSE error of 0.1793 and an f1-score of 87%, as depicted in Table 5.1.

Furthermore, when evaluating our approach using the DDR dataset, we achieved an accuracy score of 80.36% and a kappa score of 66.51%. Additional performance metrics are outlined in Table 5.1. Additionally, our approach was validated on the EyePACS dataset, where we obtained a commendable accuracy score of 78.64%, as demonstrated in Table 5.1.

Our approach demonstrated varying test accuracy scores across different datasets,

Table 5.1: The performance of the proposed approach during the test phase using Ac, Rec, Pr, Ka, and F1 denotes accuracy, recall, precision, Kappa score, and F1-score. On the four datasets used. (unit:%).

Datasets	APTOS				DDR				Messidor-2				EyePACS			
Measure metrics	Ac	Pr	Rec	F1	Ac	Pr	Rec	F1	Ac	Pr	Rec	F1	Ac	Pr	Rec	F1
Achieved values	88.18	80.00	76.00	78.00	80.36	59.00	55.00	56.00	87.78	87.00	87.00	87.00	78.64	54.00	42.00	46.00

with notable achievements of 88.18%, 87.78%, and 80.36% on the APTOS, Messidor-2, and DDR datasets, respectively. Additionally, we obtained a respectable accuracy score of 78.64% on the EyePACS dataset. while we expect that these differences in accuracy may be attributed to variations in image quality and distribution within the datasets, differences in image acquisition techniques, or the presence of various types of noise.

Furthermore, to demonstrate the effectiveness of our approach, we present the confusion matrix of the test phase results in Figure 5.10 for the four datasets. Our approach achieved the best result in classifying the healthy stage (represented by label 0) for all datasets, accurately classifying 290 out of 293 images and misclassifying 2 images on the APTOS dataset, 139 out of 147 on the Messidor-2 dataset, 590 out of 659 on the DDR dataset, and 3747 out of 3843 on the EyePACS dataset. Subsequently, for the moderate level represented by label 2, our approach correctly classified 108 out of 127 images with 13 misclassifications on the APTOS dataset. Additionally, for the mild stage (label 1), 41 images were correctly predicted out of 57, with 13 misidentified images on the APTOS dataset. Furthermore, for the PDR stage depicted by label 4, 26 images out of 41 were classified as expected, while 4 images were mislabeled. Finally, the severe stage was correctly classified with 20 images out of 32 that were sought, with the misclassification of 14 images on the APTOS dataset.

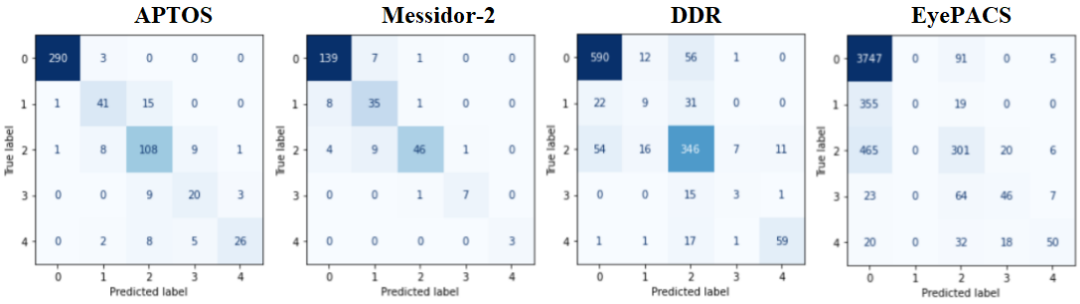

Figure 5.10: Confusion matrix of the proposed approach during the test phase across the four datasets utilized.

5.5.5 Hybrid Deep Learning Architecture For Diabetic Retinopathy Prediction Using Medical Images

We have trained, validated, and tested this approach on the two datasets (APTOS and DDR) for the problem of grading diabetic retinopathy using retinal fundus images.

For the APTOS dataset, our model was trained to assess the severity levels of diabetic retinopathy. Subsequently, we attained a high test accuracy score of 87.81%. As depicted in Table 5.2, our approach demonstrated high micro, macro, and weighted precision of 87.81%, 79.58%, and 88.36%, respectively. Additionally, we achieved high micro, macro, and weighted recall of 87.81%, 75.57%, and 87.81%. Furthermore, our approach yielded a successful F1-score of 87.81% and a high Kappa score outcome of 81.02%, accompanied by an MSE error of 0.18, as shown in Table 5.2.

Our performance on the DDR dataset also yielded impressive results, with a high accuracy score of 83.18% compared to the state of the rate of the DDR dataset. Additionally, we achieved high micro, macro, and weighted precision of 83.18%, 71.84%, and 81.62%. Furthermore, our approach demonstrated high micro, macro, and weighted recall of 83.18%, 63.87%, and 83.18%. Moreover, we attained successful F1-score and Kappa score outcomes of 83.18% and 71.55%,

Table 5.2: The results obtained for the grading of DR on the APTOS and DDR datasets in the testing phase, Ac, Ma Pr, Ma Rec, Ka, and F1 denote accuracy, macro precision, macro recall, Kappa score, and F1-score., respectively. (unit:%)

Datasets	APTOS					DDR				
Metrics	Ac	Ma Pr	Ma Rec	F1	Ka	Ac	Ma Pr	Ma Rec	F1	Ka
Values reached for multi-task	87.81	79.58	74.57	87.81	81.02	83.18	71.84	63.87	83.18	71.55

as illustrated in Table 5.2.

Additionally, as evidence of the remarkable outcomes generated by our method, we present in Figure 5.11 the confusion matrix of the results obtained during the test phase. For the APTOS dataset, our approach achieved the best outcome for label 0, representing the healthy level, correctly classifying 288 out of 293 images with precision and recall scores of 100% and 98%, respectively. Moreover, for label 2, representing the moderate level, we achieved a precision and recall score of 75% and 87%, successfully classifying 110 out of 127 images with 17 misclassifications. Furthermore, for the DDR dataset, we accurately classified 893 out of 963 images for the healthy class, with misclassification of 70 images and precision and recall scores of 86% and 93% for label 0. Additionally, for the PDR stage, represented by label 4, we correctly identified 111 out of 133 images with a misidentification of 22 images, achieving precision and recall scores of 92% and 83%.

5.6 Comparing finding results with the state-of-the-art

In the same light of demonstrating the effectiveness and performance of our approach and the outstanding results achieved, we compare our findings with those obtained by state-of-the-art methods as presented in Table5.3. All the approaches

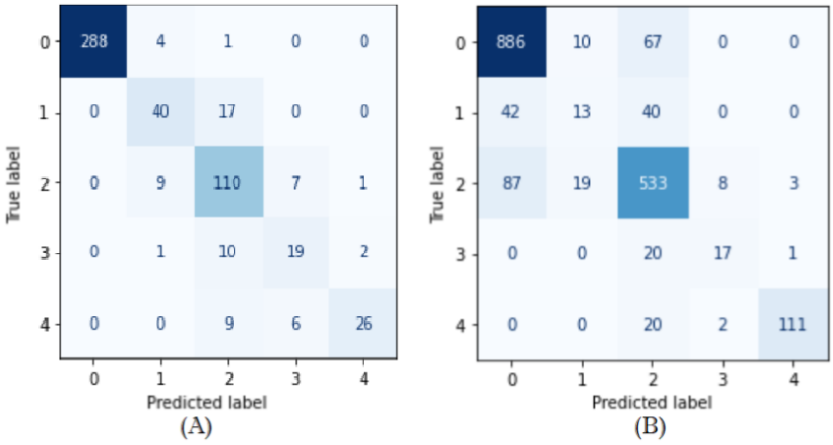

Figure 5.11: The confusion matrix of our approach in the testing phase. (A) for the APTOS dataset. (B) for the DDR dataset.

presented in Table 5.3 are proposed for the task of grading DR on the APTOS dataset during the test phase. We can state that the two presented approaches achieved the highest accuracies of 88.18% and 87.81%. Also, we achieved high precision scores of 80.00% and 79.58% and competitive metrics such as recall and f1-score.

Table 5.3: The comparison of our proposed approach with relevant earlier research (unit:%).

Dataset	Architecture	Years	Approach	Accuracy	precision	recall	F1-score	Kappa score
APTOS	Bodapati et al.[63]	2021	DNN + gated-attention mechanism	82.54	-	-	-	-
	Fan et al.[79]	2021	Multi-scale features	85.32	-	-	**85.30**	77.26
	Sugeno et al.[107]	2021	transfer learning + EfficientNet-B3	84.20	-	-	-	-
	Bodapati et al.[108]	2022	stacked convolutional auto-encoder	86.08	76.00	**82.00**	-	-
	Vijayan et al.[11]	2023	Efficientnet-B0	86.20	-	-	-	-
	Zhao et al.[109]	2022	CoTXNet	84.18	-	-	-	-
	Al-Antary and Arafa [5]	2021	extraction of features + attention	84.60	-	-	-	-
	Hu et al.[60]	2022	graph adversarial	83.50	-	-	-	-
	Islam et al. [40]	2022	Supervised contrastive learning	84.35	73.84	70.51	70.49	-
	Shaik and Cherukuri [61]	2022	Hinge Attention Networks	85.54	-	-	-	-
	Oulhadj et al.[14](ours)	2023	CapsNet + inception block + DWT	86.54	76.00	70.00	73.00	78.77
	Mondal et al.[10]	2023	ensemble deep-learning technique	86.08	76.00	**82.00**	-	-
	first approach (ours)	2023	ViT+ CapsNet+ PLT+CLAHE	**88.18**	**80.00**	76.00	78.00	**81.55**
	second approach (ours)	2023	ViT+ CBAM	**87.81**	**79.58**	75.57	87.81	**81.02**

5.7 conclusion

In this chapter, we introduced two innovative approaches for the detection of the detection of diabetic retinopathy. The first approach is called diabetic retinopathy prediction based on vision transformers and modified capsule networks, which is based on a pre-processing step to enhance the retina fundus image before feeding it to the classification step, which consists of a modifier ViT model and a modifier capsule network model to generate feature maps to feed to the classification model to detect the stage of diabetic retinopathy. The second approach is called "Hybrid Deep Learning Architecture For Diabetic Retinopathy Prediction Using Medical Images", which is based on a hybrid model that consists of VitIncep, CBAM, and SE architectures to classify the retina fundus images. Overall, these approaches represent notable progress in precise and comprehensive diabetic retinopathy detection, offering the potential for enhanced outcomes in medical image analysis as further research continues to refine and optimize these methods.

Conclusions

In this thesis, we presented various deep learning-based approaches to enhance the prediction and classification of DR, aiming to provide reliable tools to assist healthcare professionals. The proposed methodologies leveraged advanced deep learning architectures, including modified capsule networks, vision transformers, and hybrid models incorporating attention mechanisms like CBAM and SE blocks. In addition to powerful pre-processing operations such as deformable registration techniques, wavelet decomposition, power low transformation, and contrast-limiting adaptive histogram equalization. Each approach demonstrated promising results in improving the accuracy and quality of DR detection and classification, underscoring the potential of deep learning to transform medical diagnostics. In this regard, we proposed an approach based on deformable registration as a pre-processing step to eliminate the effect of the background on the training process. The processed images were then fed to an ensemble of four pre-trained models for the training task. During the testing process, we employed a majority voting technique to determine the most frequent decision among the four models. The outcomes demonstrate that our suggested approach performs better than most state-of-the-art methodologies, by achieving an accuracy score of 85.28% on the APTOS dataset. Then, we proposed a new approach for the same task based on wavelet decomposition in the pre-processing step, generating three levels of decomposition from a single image. The generated images were

fed to a newly evaluated inception block, capable of handling the three images simultaneously, and connected to a deep hybrid model. Subsequently, an evaluated capsule network model received the generated features to enhance and grade the DR stage. The outcomes demonstrate that our suggested approach performs better than most state-of-the-art methodologies, by achieving an accuracy score of 86.54% on the APTOS dataset. Moreover, we proposed a new approach based on a powerful pre-processing step that generates two enhanced images. First, the power law transformation is applied to enhance the retina fundus image, creating the first image. Then, using the green channel of this image, we apply contrast-limiting adaptive histogram equalization to generate the second one. These images are fed to a pre-trained vision transformer and a modified capsule network model, respectively. The features generated by the two models are then fused using a classification model to predict the stage of diabetic retinopathy. The results indicate that our proposed approach surpasses several state-of-the-art methods, achieving notable test accuracies of 88.18%, 87.78%, 80.36%, and 78.64% on the APTOS, Messidor-2, DDR, and EyePACS datasets, respectively. Finally, we proposed a new approach for grading DR based on a hybrid deep learning model. After enhancing the image using various techniques such as detail enhancement, sharpening, and improving contrast and brightness, we fed this image to a hybrid deep learning model. The model starts with a ViTIncep sub-model to generate feature maps, which are then fed to the CBAM and SE blocks. We used a series of fusion layers to combine features generated from the CBAM block with those from the ViTIncep, followed by further fusion with features from the SE block. These fused features are then fed to a classification sub-model for grading DR. The outcomes indicate that our proposed approach surpasses many state-of-the-art methodologies, by reaching an accuracy score of 87.81% on the APTOS dataset and a score of 83.18% on the DDR dataset. Our work utilized diverse reti-

nal fundus image datasets and applied rigorous evaluation metrics to validate the effectiveness of the proposed models. Through the integration of pre-processing techniques and innovative deep learning models, this research has contributed to the growing body of knowledge aimed at detecting the early and accurate stage of diabetic retinopathy through scalable automated solutions. Despite the significant advancements achieved, there is still room for improvement, particularly in enhancing multiclass classification performance and ensuring early-stage detection. Future research should continue to refine these, explore additional deep learning techniques, and expand the datasets used to encompass a broader spectrum of retinal images and conditions. In addition, our work can be extended to include the ability to segment diabetic retinopathy symptoms within retinal fundus images, providing additional information for ophthalmologists. Furthermore, these deep learning architectures can be enhanced to be adaptable for detecting other diseases, such as brain and skin tumors. In conclusion, this thesis demonstrated that deep learning offers a powerful and promising avenue for improving the detection of diabetic retinopathy, ultimately contributing to better patient outcomes and advancing the field of medical imaging.

References

[1] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017. 1, 2, 43

[2] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019. 1, 9, 45

[3] N. J. Mohan, R. Murugan, T. Goel, and P. Roy, "Vit-dr: Vision transformers in diabetic retinopathy grading using fundus images," in *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 167–172, IEEE, 2022. 2, 50

[4] G. U. Nneji, J. Cai, J. Deng, H. N. Monday, S. Nahar, G. T. Mgbejime, E. C. James, and S. K. Woldeyes, "A dual weighted shared capsule network for diabetic retinopathy fundus classification," in *2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, pp. 297–302, IEEE, 2021. 2, 50

[5] M. T. Al-Antary and Y. Arafa, "Multi-scale attention network for diabetic retinopathy classification," *IEEE Access*, vol. 9, pp. 54190–54200, 2021. 2, 46, 83, 85, 108

[6] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid, "Automatic severity classification of diabetic retinopathy based on densenet and convolutional block attention module," *IEEE Access*, vol. 10, pp. 38299–38308, 2022. 2, 49

[7] S. Das, K. Kharbanda, M. Suchetha, R. Raman, and E. Dhas, "Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy," *Biomedical Signal Processing and Control*, vol. 68, p. 102600, 2021. 2, 41

[8] W. Wong, F. H. Juwono, and C. Capriono, "Diabetic retinopathy detection and grading: A transfer learning approach using simultaneous parameter optimization and feature-weighted ecoc ensemble," *IEEE Access*, 2023. 2, 44

[9] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence*, pp. 679–689, Springer, 2021. 2, 45, 83

[10] S. S. Mondal, N. Mandal, K. K. Singh, A. Singh, and I. Izonin, "Edldr: An ensemble deep learning technique for detection and classification of diabetic retinopathy," *Diagnostics*, vol. 13, no. 1, p. 124, 2023. 2, 41, 87, 108

[11] M. Vijayan, "A regression-based approach to diabetic retinopathy diagnosis using efficientnet," *Diagnostics*, vol. 13, no. 4, p. 774, 2023. 2, 40, 108

[12] F. Orujov, R. Maskeliūnas, R. Damaševičius, and W. Wei, "Fuzzy based image edge detection algorithm for blood vessel detection in retinal images," *Applied Soft Computing*, vol. 94, p. 106452, 2020. 2, 40

[13] M. Oulhadj, J. Riffi, K. Chaimae, A. M. Mahraz, B. Ahmed, A. Yahyaouy, C. Fouad, A. Meriem, B. A. Idriss, and H. Tairi, "Diabetic retinopathy

prediction based on deep learning and deformable registration," *Multimedia Tools and Applications*, pp. 1–19, 2022. 3, 83

[14] M. Oulhadj, J. Riffi, C. Khodriss, A. M. Mahraz, A. Bennis, A. Yahyaouy, F. Chraibi, M. Abdellaoui, I. B. Andaloussi, and H. Tairi, "Diabetic retinopathy prediction based on wavelet decomposition and modified capsule network," *Journal of Digital Imaging*, pp. 1–13, 2023. 3, 86, 108

[15] M. Oulhadj, J. Riffi, C. Khodriss, A. M. Mahraz, A. Yahyaouy, M. Abdellaoui, I. B. Andaloussi, and H. Tairi, "Diabetic retinopathy prediction based on vision transformer and modified capsule network," *Computers in Biology and Medicine*, vol. 175, p. 108523, 2024. 4

[16] S. D. Solomon, E. Chew, E. J. Duh, L. Sobrin, J. K. Sun, B. L. VanderBeek, C. C. Wykoff, and T. W. Gardner, "Diabetic retinopathy: a position statement by the american diabetes association," *Diabetes care*, vol. 40, no. 3, pp. 412–418, 2017. 9

[17] A. R. Santos, L. Ribeiro, F. Bandello, R. Lattanzio, C. Egan, U. Frydkjaer-Olsen, J. García-Arumí, J. Gibson, J. Grauslund, S. P. Harding, *et al.*, "Functional and structural findings of neurodegeneration in early stages of diabetic retinopathy: cross-sectional analyses of baseline data of the eurocondor project," *Diabetes*, vol. 66, no. 9, pp. 2503–2510, 2017. 9, 68

[18] F. I. Himasa, M. Singhal, A. Ojha, and B. Kumar, "Prospective for diagnosis and treatment of diabetic retinopathy," *Current Pharmaceutical Design*, vol. 28, no. 7, pp. 560–569, 2022. 9

[19] A. Benzamin and C. Chakraborty, "Detection of hard exudates in retinal fundus images using deep learning," in *2018 Joint 7th International*

conference on informatics, electronics & vision (ICIEV) and 2018 2nd International conference on imaging, vision & pattern recognition (icIVPR), pp. 465–469, IEEE, 2018. 9

[20] M. J. Paranjpe and M. Kakatkar, "Review of methods for diabetic retinopathy detection and severity classification," *International Journal of Research in Engineering and Technology*, vol. 3, no. 3, pp. 619–624, 2014. 12

[21] D. J. Magliano, E. J. Boyko, *et al.*, "Idf diabetes atlas," 2022. 13

[22] D. S. Fong, L. Aiello, T. W. Gardner, G. L. King, G. Blankenship, J. D. Cavallerano, F. L. Ferris III, R. Klein, and A. D. Association, "Diabetic retinopathy," *Diabetes care*, vol. 26, no. suppl_1, pp. s99–s102, 2003. 13

[23] A. Razzak, A. Razzak, A. Kassimi, A. Mchachi, L. Benhmidoun, R. Rachid, and M. Elbelhadji, "Aspects epidemiologiques de la rétinopathie diabétique à casablanca," *Journal de la Société Marocaine d'Ophtalmologie*, no. 28, pp. 58–65, 2019. 13

[24] J. Hammoudi, N. E. H. Bouanani, Y. Bentata, H. Nouayti, A. Legssyer, A. Ziyyat, *et al.*, "Diabetic retinopathy in the eastern morocco: Different stage frequencies and associated risk factors," *Saudi Journal of Biological Sciences*, vol. 28, no. 1, pp. 775–784, 2021. 13

[25] Y. El Harrak, "Eye & diabetes in morocco: Where are we from the global screening guidelines?," *European Journal of Public Health*, vol. 33, no. Supplement_2, pp. ckad160–1149, 2023. 13

[26] APTOS, "Kaggle diabetic retinopathy detection competition," *https://www.kaggle.com/c/aptos2019-blindnessdetection/data, Accessed: 18-Mar-2020.* 32, 60, 80

[27] EyePACS, "Diabetic retinopathy detection competition," *https://www.kaggle.com/c/diabetic-retinopathy-detection/data*. 32, 43

[28] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA ophthalmology*, vol. 131, no. 3, pp. 351–357, 2013. 33

[29] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014. 33, 37

[30] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511 – 522, 2019. 33

[31] E. V. Carrera, A. González, and R. Carrera, "Automated detection of diabetic retinopathy using svm," in *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4, IEEE, 2017. 36

[32] M. Chetoui, M. A. Akhloufi, and M. Kardouchi, "Diabetic retinopathy detection using machine learning and texture features," in *2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, pp. 1–4, IEEE, 2018. 37

[33] K. M. Adal, P. G. Van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, "An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus

images," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1382–1390, 2017. 37

[34] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Classification of diabetic retinopathy and normal retinal images using cnn and svm," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, pp. 152–157, IEEE, 2019. 37

[35] M. Jahiruzzaman and A. A. Hossain, "Detection and classification of diabetic retinopathy using k-means clustering and fuzzy logic," in *2015 18th International Conference on Computer and Information Technology (ICCIT)*, pp. 534–538, IEEE, 2015. 38

[36] R. Biyani and B. Patre, "A clustering approach for exudates detection in screening of diabetic retinopathy," in *2016 International Conference on Signal and Information Processing (IConSIP)*, pp. 1–5, IEEE, 2016. 39

[37] B. Wu, W. Zhu, F. Shi, S. Zhu, and X. Chen, "Automatic detection of microaneurysms in retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 106–112, 2017. 39

[38] I. Jamal, M. U. Akram, and A. Tariq, "Retinal image preprocessing: background and noise segmentation," *Telkomnika*, vol. 10, no. 3, pp. 537–544, 2012. 40

[39] S. S. Kar and S. P. Maity, "Automatic detection of retinal lesions for screening of diabetic retinopathy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 608–618, 2017. 40

[40] M. R. Islam, L. F. Abdulrazak, M. Nahiduzzaman, M. O. F. Goni, M. S. Anower, M. Ahsan, J. Haider, and M. Kowalski, "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity

levels from fundus images," *Computers in Biology and Medicine*, p. 105602, 2022. 41, 83, 108

[41] L. K. Ramasamy, S. G. Padinjappurathu, S. Kadry, and R. Damaševičius, "Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier," *PeerJ computer science*, vol. 7, 2021. 42

[42] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016. 42

[43] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 533–540, Springer, 2017. 43

[44] H. S. Sharma, A. Singh, A. S. Chandel, P. Singh, P. Sapkal, *et al.*, "Detection of diabetic retinopathy using convolutional neural network," *Detection of Diabetic Retinopathy Using Convolutional Neural Network (May 17, 2019)*, 2019. 43

[45] A. Jain, A. Jalui, J. Jasani, Y. Lahoti, and R. Karani, "Deep learning for detection and severity classification of diabetic retinopathy," in *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, pp. 1–6, IEEE, 2019. 43

[46] R. Murugan, P. Roy, and U. Singh, "An abnormality detection of retinal fundus images by deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 33, pp. 24949–24967, 2020. 43

[47] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, 2018. 43

[48] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, *et al.*, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 185–195, 2009. 43

[49] Z. Khan, F. G. Khan, A. Khan, Z. U. Rehman, S. Shah, S. Qummar, F. Ali, and S. Pack, "Diabetic retinopathy detection using vgg-nin a deep learning architecture," *IEEE Access*, vol. 9, pp. 61408–61416, 2021. 44, 87

[50] A. Pak, A. Ziyaden, K. Tukeshev, A. Jaxylykova, and D. Abdullina, "Comparative analysis of deep learning methods of detection of diabetic retinopathy," *Cogent Engineering*, vol. 7, no. 1, p. 1805144, 2020. 44

[51] O. Dekhil, A. Naglah, M. Shaban, M. Ghazal, F. Taher, and A. Elbaz, "Deep learning based method for computer aided diagnosis of diabetic retinopathy," in *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–4, IEEE, 2019. 45, 67

[52] M. Rao, M. Zhu, and T. Wang, "Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy," *arXiv preprint arXiv:2010.11692*, 2020. 45, 83

[53] M. AATILA, M. LACHGAR, H. HRIMECH, A. KARTIT, *et al.*, "Diabetic retinopathy classification using resnet50 and vgg-16 pretrained net-

works," *International Journal of Computer Engineering and Data Science (IJCEDS)*, vol. 1, no. 1, pp. 1–7, 2021. 46

[54] M. Tariq, V. Palade, and Y. Ma, "Transfer learning based classification of diabetic retinopathy on the kaggle eyepacs dataset," in *The 3rd International Conference on Medical Imaging and Computer-Aided Diagnosis*, 2022. 46

[55] J. Ran, G. Zhang, X. Zhang, J. Xie, F. Xia, and H. Zhang, "Source-free active domain adaptation for diabetic retinopathy grading based on ultra-wide-field fundus image," *arXiv preprint arXiv:2309.10619*, 2023. 46

[56] S. Abbasi, M. Hajabdollahi, P. Khadivi, N. Karimi, R. Roshandel, S. Shirani, and S. Samavi, "Classification of diabetic retinopathy using unlabeled data and knowledge distillation," *Artificial Intelligence in Medicine*, vol. 121, p. 102176, 2021. 46

[57] E. AbdelMaksoud, S. Barakat, and M. Elmogy, "A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique," *Medical & Biological Engineering & Computing*, vol. 60, no. 7, pp. 2015–2038, 2022. 47

[58] M. M. Butt, D. A. Iskandar, S. E. Abdelhamid, G. Latif, and R. Alghazo, "Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features," *Diagnostics*, vol. 12, no. 7, p. 1607, 2022. 47

[59] S. S. Chaturvedi, K. Gupta, V. Ninawe, and P. S. Prasad, "Automated diabetic retinopathy grading using deep convolutional neural network," *arXiv preprint arXiv:2004.06334*, 2020. 47

[60] J. Hu, H. Wang, L. Wang, and Y. Lu, "Graph adversarial transfer learning

for diabetic retinopathy classification," *IEEE Access*, vol. 10, pp. 119071–119083, 2022. 48, 108

[61] N. S. Shaik and T. K. Cherukuri, "Hinge attention network: A joint model for diabetic retinopathy severity grading," *Applied Intelligence*, vol. 52, no. 13, pp. 15105–15121, 2022. 48, 108

[62] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "Canet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1483–1493, 2019. 49

[63] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9825–9839, 2021. 49, 83, 108

[64] X. Wang, Y. Fang, S. Yang, D. Zhu, M. Wang, J. Zhang, J. Zhang, J. Cheng, K.-y. Tong, and X. Han, "Clc-net: Contextual and local collaborative network for lesion segmentation in diabetic retinopathy images," *Neurocomputing*, vol. 527, pp. 100–109, 2023. 50

[65] C. O. S. Sorzano, P. Thévenaz, and M. Unser, "Elastic registration of biological images using vector-spline regularization," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 652–663, 2005. 54

[66] I. Arganda-Carreras, C. O. Sorzano, R. Marabini, J. M. Carazo, C. Ortiz-de Solorzano, and J. Kybic, "Consistent and elastic registration of histological sections using vector-spline regularization," in *International Workshop on Computer Vision Approaches to Medical Image Analysis*, pp. 85–95, Springer, 2006. 54

[67] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep transfer learning for modality classification of medical images," *Information*, vol. 8, no. 3, p. 91, 2017. 55

[68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016. 55

[69] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017. 56

[70] X. Li, X. Shen, Y. Zhou, X. Wang, and T.-Q. Li, "Classification of breast cancer histopathological images using interleaved densenet with senet (id-snet)," *PloS one*, vol. 15, no. 5, p. e0232127, 2020. 56

[71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. 56

[72] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017. 57

[73] M. Arora and M. Pandey, "Deep neural network for diabetic retinopathy detection," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 189–193, IEEE, 2019. 67

[74] S. H. Kassani, P. H. Kassani, R. Khazaeinezhad, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Diabetic retinopathy classification using a mod-

ified xception architecture," in *2019 IEEE international symposium on signal processing and information technology (ISSPIT)*, pp. 1–6, IEEE, 2019. 67, 83

[75] G. Kumar, S. K. Chatterjee, and C. Chattopadhyay, "Drdnet: Diagnosis of diabetic retinopathy using capsule network (workshop paper)," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 379–385, IEEE, 2020. 67

[76] H. Zhuang and N. Ettehadi, "Classification of diabetic retinopathy via fundus photography: Utilization of deep learning approaches to speed up disease detection," *arXiv preprint arXiv:2007.09478*, 2020. 67

[77] J. D. Bodapati, N. Veeranjaneyulu, S. N. Shareef, S. Hakak, M. Bilal, P. K. R. Maddikunta, and O. Jo, "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction," *Electronics*, vol. 9, no. 6, p. 914, 2020. 67

[78] A. K. Gangwar and V. Ravi, "Diabetic retinopathy detection using transfer learning and deep learning," in *Evolution in Computational Intelligence*, pp. 679–689, Springer, 2020. 67

[79] R. Fan, Y. Liu, and R. Zhang, "Multi-scale feature fusion with adaptive weighting for diabetic retinopathy severity classification," *Electronics*, vol. 10, no. 12, p. 1369, 2021. 67, 83, 108

[80] D. Atlas, "International diabetes federation. idf diabetes atlas, 9th edn. brussels," *Belgium: 2019. Available at: https://www.diabetesatlas.org, Accessed: 17-nov-2020*, 2019. 68

[81] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in

Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015. 70, 92

[82] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in neural information processing systems*, vol. 30, 2017. 72, 78, 88, 97, 98

[83] V. Dondeti, J. D. Bodapati, S. N. Shareef, and N. Veeranjaneyulu, "Deep convolution features in non-linear embedding space for fundus image classification.," *Rev. d'Intelligence Artif.*, vol. 34, no. 3, pp. 307–313, 2020. 83

[84] S. Majumder and N. Kehtarnavaz, "Multitasking deep learning model for detection of five stages of diabetic retinopathy," *IEEE Access*, vol. 9, pp. 123220–123230, 2021. 83

[85] J. Fu, L. Cao, S. Wei, M. Xu, Y. Song, H. Li, and Y. You, "A gan-based deep enhancer for quality enhancement of retinal images photographed by a handheld fundus camera," *Advances in Ophthalmology Practice and Research*, vol. 2, no. 3, p. 100077, 2022. 86

[86] O. O. Sule and A. E. Ezugwu, "A two-stage histogram equalization enhancement scheme for feature preservation in retinal fundus images," *Biomedical Signal Processing and Control*, vol. 80, p. 104384, 2023. 86

[87] A. Raj, N. A. Shah, and A. K. Tiwari, "A novel approach for fundus image enhancement," *Biomedical signal processing and control*, vol. 71, p. 103208, 2022. 86

[88] B. Menaouer, Z. Dermane, N. El Houda Kebir, and N. Matta, "Diabetic retinopathy classification using hybrid deep learning approach," *SN Computer Science*, vol. 3, no. 5, p. 357, 2022. 86

[89] B. Venkaiahppalaswamy, P. P. Reddy, and S. Batha, "Hybrid deep learning approaches for the detection of diabetic retinopathy using optimized wavelet based model," *Biomedical Signal Processing and Control*, vol. 79, p. 104146, 2023. 86

[90] M. Alshawabkeh, M. H. Ryalat, O. M. Dorgham, K. Alkharabsheh, M. H. Btoush, and M. Alazab, "A hybrid convolutional neural network model for detection of diabetic retinopathy," *International Journal of Computer Applications in Technology*, vol. 70, no. 3-4, pp. 179–196, 2022. 86

[91] P. Kumari and P. Saxena, "Automated diabetic retinopathy grading based on the modified capsule network architecture," *IETE Journal of Research*, pp. 1–12, 2023. 86

[92] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. N. Prasad, "Diabetic retinopathy detection and classification using capsule networks," *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2651–2664, 2023. 86

[93] W. Dai, C. Mou, J. Wu, and X. Ye, "Diabetic retinopathy detection with enhanced vision transformers: The twins-pcpvt solution," in *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pp. 403–407, IEEE, 2023. 86

[94] Z. Gu, Y. Li, Z. Wang, J. Kan, J. Shu, Q. Wang, *et al.*, "Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention," *Computational Intelligence and Neuroscience*, vol. 2023, 2023. 86

[95] M. Chetoui and M. A. Akhloufi, "Federated learning for diabetic retinopathy detection using vision transformers," *BioMedInformatics*, vol. 3, no. 4, pp. 948–961, 2023. 86

[96] C.-M. Tsai, "Adaptive local power-law transformation for color image enhancement," *Applied Mathematics & Information Sciences*, vol. 7, no. 5, p. 2019, 2013. 86

[97] M. Kumar and K. M. Singh, "Retrieval of head–neck medical images using gabor filter based on power-law transformation method and rank bhmt," *Signal, Image and Video Processing*, vol. 12, no. 5, pp. 827–833, 2018. 86

[98] A. Solano, K. N. Dietrich, M. Martínez-Sober, R. Barranquero-Cardeñosa, J. Vila-Tomás, and P. Hernández-Cámara, "Deep learning architectures for diagnosis of diabetic retinopathy," *Applied Sciences*, vol. 13, no. 7, p. 4445, 2023. 86

[99] T. Aswathi, T. Swapna, and S. Padmavathi, "Transfer learning approach for grading of diabetic retinopathy," in *Journal of Physics: Conference Series*, vol. 1767, p. 012033, IOP Publishing, 2021. 86

[100] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020. 87

[101] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018. 89

[102] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018. 91

[103] N. A. Oommen and P. Darsana, "A research insight toward the significance in extraction of retinal blood vessels from fundus images and its various

implementations," in *Cognitive Systems and Signal Processing in Image Processing*, pp. 163–201, Elsevier, 2022. 93

[104] S. Biswas, M. I. A. Khan, M. T. Hossain, A. Biswas, T. Nakai, and J. Rohdin, "Which color channel is better for diagnosing retinal diseases automatically in color fundus photographs?," *Life*, vol. 12, no. 7, p. 973, 2022. 94

[105] G. Indumathi and V. Sathananthavathi, "Microaneurysms detection for early diagnosis of diabetic retinopathy using shape and steerable gaussian features," in *Telemedicine Technologies*, pp. 57–69, Elsevier, 2019. 94

[106] R. Tharsanee, R. Soundariya, A. S. Kumar, M. Karthiga, and S. Sountharrajan, "Deep convolutional neural network–based image classification for covid-19 diagnosis," in *Data Science for COVID-19*, pp. 117–145, Elsevier, 2021. 94

[107] A. Sugeno, Y. Ishikawa, T. Ohshima, and R. Muramatsu, "Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning," *Computers in Biology and Medicine*, vol. 137, p. 104795, 2021. 108

[108] J. D. Bodapati, "Stacked convolutional auto-encoder representations with spatial attention for efficient diabetic retinopathy diagnosis," *Multimedia Tools and Applications*, vol. 81, no. 22, pp. 32033–32056, 2022. 108

[109] S. Zhao, Y. Wu, M. Tong, Y. Yao, W. Qian, and S. Qi, "Cot-xnet: contextual transformer with xception network for diabetic retinopathy grading," *Physics in Medicine & Biology*, vol. 67, no. 24, p. 245003, 2022. 108

[110] D. S. Fong, L. Aiello, T. W. Gardner, G. L. King, G. Blankenship, J. D.

Cavallerano, F. L. Ferris III, R. Klein, and A. D. Association, "Retinopathy in diabetes," *Diabetes care*, vol. 27, no. suppl_1, pp. s84–s87, 2004.

[111] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," *arXiv preprint arXiv:2003.02261*, 2020.

[112] J. I. Orlando, E. Prokofyeva, M. Del Fresno, and M. B. Blaschko, "An ensemble deep learning based approach for red lesion detection in fundus images," *Computer methods and programs in biomedicine*, vol. 153, pp. 115–127, 2018.

[113] K. Adem, "Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks," *Expert Systems with Applications*, vol. 114, pp. 289–295, 2018.

[114] M. Oulhadj, J. Riffi, C. Khodriss, A. M. Mahraz, A. Bennis, A. Yahyaouy, F. Chraibi, M. Abdellaoui, I. B. Andsaloussi, and H. Tairi, "Diabetic retinopathy prediction based on transfer learning and ensemble voting," in *International Conference on Digital Technologies and Applications*, pp. 929–937, Springer, 2023.