Université Mohamed Boudiaf –M'Sila

Département d'électronique

Faculté de technologie

OPTION : Electronique des Systèmes Embarqués

Early Detection and Classification for Diabetic Retinopathy by Deep Learning

Réalisé par :

*Lahmar Hanine

*Ziani Zohra

Dirigé par:

*Dr.Brik Youcef

Thanks

First, we thank ALLAH for finishing this dissertation

The realization of this work was possible thanks to the contribution of several people to whom we would like to express our gratitude.

First, we would like to thank Dr. Youcef BRIK, the director of this dissertation, for having guided, encouraged and advised us throughout this project.

Our thanks also go to T.BAGHRICHE for giving us his time and effort.

We also wish to thank the professors of the University of M'SILA who provided us with the necessary tools for our success in our university studies.

We also thank the members of the jury: Dr. B. ATTALLAH and Dr. M. DJERIOUI for their interest in our research by agreeing to examine our work.

We also wish to thank all the professional and administrative body of the Electronic Department at University Mohammed BOUDIAF of MSILA.

Dedication

I dedicate this thesis

To my dear parents, for all their sacrifices, their love, their tenderness,

their support and their prayers throughout my studies.

To Sabrina for her presence in my life

To my dear brothers Redouene, Saleh, Khalil, Moussa.

To my dearest sisters Akila, Dalila, Hada, Zayneb, and their little

families.

To Farha, Chourouk.

Finally to my partner Hanine for his understanding, and tolerance

during the finishing of this project.

Ziani Zohra

Dedication

I dedicate these notes

To my dear parents for all their sacrifice, their love, their tenderness,

their support and their prayers throughout my studies, their confidence

and their precious advice, and for all their presence in my life.

To my dear and only brother Houssam and his little family

Thank you for your support.

To my dear sisters Nedjma, and Souhiala for their understanding and

their encouragement

To my dearest sister Lobena and her husband Yassin and her children

Mariam, Houssam, and Amir who can be proud and find here a long result

Years of sacrifice and denial to help me move forward in life

Finally to my friend Zahra for her understanding and forgiveness in the

realization of this project.

Lahmar Hanine

ABSTRACT

Deep learning helps to facilitate and organize our daily lives and our health. Recently, medical professionals are using these techniques in the treatment of many diseases such as diabetic retinopathy because it comes with great data analysis capabilities. We propose in this dessertation a detection/classification system of diabetic retinopathy as well as its severity using deep learning models. For the evaluation of this system, we used the APTOS2019 collection from Kaggle, which contains high-resolution retinal images. Besides, a protocol similar to that used by doctors has been adopted to hierarchically judge the state and severity of the DR disease. Our experiments have been conducted on five models: VGG16, VGG19, InceptionV3, Xception, and MobileNetV2. The obtained results are very satisfactory in terms of accuracy.

Keywords: diabetic retinopathy (DR), image classification, deep learning, artificial intelligence, medical imaging.

الملخص

يساعد التعلم العميق على تسهيل وتنظيم حياتنا اليومية وصحتنا. في الآونة الأخيرة، يستخدم المهنيون الطبيون هذه التقنيات في علاج العديد من الأمراض مثل اعتلال السكري في الشبكية لأنها تتمتع بإمكانات كبيرة في تحليل البيانات. نقترح في هذه المذكرة نظام كشف / تصنيف لاعتلال السكري بالشبكية بالإضافة إلى تصنيف شدته باستخدام نماذج التعلم العميق. لتقييم هذا النظام، استخدمنا قاعدة البيانات APTOS2019 من Kaggle، والتي تحتوي على صور شبكية عالية الدقة. إلى جانب ذلك، تم اعتماد بروتوكول مشابه للبروتوكول الذي يستخدمه الأطباء للحكم الهرمي على حالة وشدة مرضDR . تم إجراء تجاربنا على خمسة نماذج من التعلم العميق و هي: : VGG16و VGG19 و InceptionV3 و Xception و MobileNetV2. النتائج التي تم الحصول عليها مرضية للغاية من حيث الدقة.

الكلمات المفتاحية: اعتلال الشبكية السكري(DR) ، تصنيف الصور، التعلم العميق، الذكاء الاصطناعي، التصوير الطبي.

RESUME

L’apprentissage profond permet de faciliter et d'organiser notre quotidien et notre santé. Récemment, les professionnels de la santé utilisent ces techniques dans le traitement de nombreuses maladies telles que la rétinopathie diabétique, car elles s'accompagnent d'excellentes capacités d'analyse de données. Nous proposons dans ce mémoire un système de détection/classification de la rétinopathie diabétique ainsi que sa sévérité à l'aide de modèles d'apprentissage profond. Pour l'évaluation de ce système, nous avons utilisé la collection APTOS2019 de Kaggle, qui contient des images rétiniennes haute résolution. Par ailleurs, un protocole similaire à celui utilisé par les médecins a été adopté pour juger hiérarchiquement de l'état et de la gravité de la maladie RD. Nos expérimentations ont été menées sur cinq modèles : VGG16, VGG19, InceptionV3, Xception et MobileNetV2. Les résultats obtenus sont très satisfaisants en termes de précision.

Mots-clés : rétinopathie diabétique (RD), classification d'images, apprentissage profond, intelligence artificielle, imagerie médicale

.

LIST OF CONTENT

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION GENERAL

Diabetic retinopathy (DR) is a progressive eye disease. It is the most common cause of visual loss in the retina. Blood vessels in the retina begin to leak and clog as a result of this. Patients become aware of DR, a silent disease, when they experience vision loss. The prevalence of DR rises as the life expectancy of diabetes individuals rises. Diabetes is one of the most pressing issues in modern medicine [Lim, 2022], More than 93 million adults are estimated to be blind due to late-diagnosed retinopathy, hence the importance of a complete and automated approach recognized by the scientific community in the early detection of such lesions.

The progression of visual impairment can be prevented or avoided if DR is identified early enough. This can be difficult, though, because the ailment usually manifests with little symptoms until it is too late to remedy. The process of detecting the disease is laborious and manual, requiring a qualified ophthalmologist to evaluate the digital photography under the mirror, but often by the time the disease is diagnosed, it is too late, and the patient has suffered irreversible losses [Samy, 2019].

Deep Learning is a branch of artificial intelligence that uses neural networks as its primary resolution approach. In recent years, it has made real progress in detecting medical images in order to diagnose disease [Samy, 2019].

An approach based on neural networks with ConvNet architecture is proposed in order to predict or classify DR diseases using medical imaging. However, we are interested on Convolutional neural networks architecture that currently gives fabulous results in the field of medical imaging [Samy, 2019].

In order to properly study and evaluate our project, this dissertation is divided into three chapters organized as following:

The first chapter presents an overview on medical imaging and its importance for automated helping systems in healthcare. Also, this chapter introduces the theoretical background of machine learning and explain the fundamental concepts associated with deep learning for medical image classification. We explain also the idea involved with deep learning for our proposed system, as well as the influence of AI in medical imaging.

The second chapter starts by giving a brief description on the used dataset. Then, a literature review is conducted for diabetic retinopathy diagnosis in the last years. A

detailed description of each part of our system is provided to give an idea of the methods, models and techniques used in this work.

Finally, chapter three evaluates and discusses many different experiments conducted on APTOS2019 dataset using five transfer learning models (VGG16, VGG19, InceptionV3, Xception and MobileNetV2). Furthermore, the preprocessing task and data augmentation are performed in order to enhance the system accuracy. The obtained results are expressed in term of accuracy, precision, recall and f1-score.

We end this dissertation by a general conclusion and some future perspectives.

CHAPTER I: BIOMEDICAL DATA AND AI

I.1. INTRODUCTION

In this Chapter, we discuss the medical imaging and its importance for automated helping systems in healthcare. Also, we introduce the theoretical background of machine learning and explain the fundamental concepts associated with deep learning for image classification. More attention will be paid on convolutional neural network and its application in medical imaging prediction.

I.2. BIOMEDICAL DATA

Due to the popularity of biomedical imaging methods, a large amount of biomedical and health informatics data is generated every day. It is notoriously difficult to get complete and correct information from medical imaging using manual (traditional) method. Therefore, it is necessary to use intelligent helping systems for better diagnosis. In fact, collecting medical data is an important first step in the process. In many cases, however, the real crux of the problem lies in trying to completely exhaust all available sources of information [Wang, 2021]. In the next section, we describe briefly different types of biomedical data used nowadays.

I.3. BIOMEDICAL DATA TYPES

Medical imaging refers to a range of techniques used to visualize specific parts of the body for clinical diagnosis and medical treatment. Medical imaging can also visualize the function of a tissue or organ. Medical imaging technology enables doctors to examine the inner structure of the skin and bones, and to make diagnoses and provide treatment. Medical imaging also helps create physiology and normal anatomy datasets for further analysis by researchers [Wang, 2021].

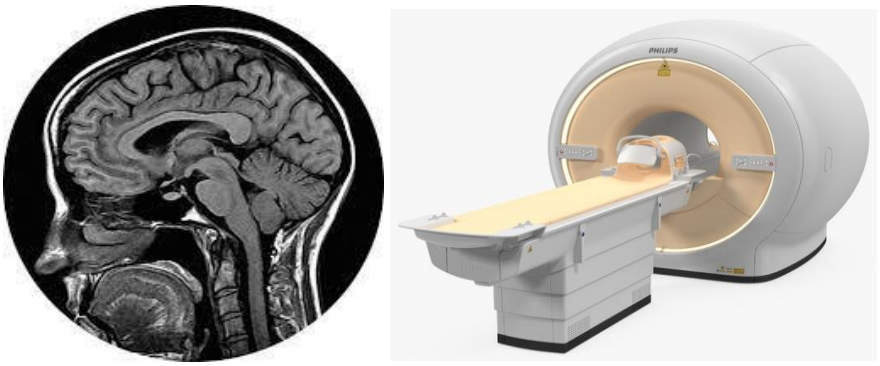
Medical imaging is part of a broader field of biological imaging that includes many different types of imaging techniques such as:

I.3.1. Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technique used to create 3D detailed anatomical images using strong magnetic fields, magnetic field gradients, and radio waves to create images of body anatomy and physiological processes. The patient should be positioned in an MRI scanner that creates a strong

magnetic field around the area of interest to be examined.as it's shown in (b) in Figure I.1

The MRI has the advantage of not requiring ionizing radiation or X-rays, which can be dangerous. Although the scanning procedure takes a long time and makes loud noises, MRI is an excellent imaging method in terms of image details. As shown in (a) in Figure I.1 [Wang, 2021].



(a) (b)

Figure I.1. Magnetic Resonance Imaging (MRI) images: a) MRI brain scan and b) MRI scanner.

I.3.2. Computed Tomography (CT)

CT is a type of imaging that uses rotating X-ray machines and computers to create cross-sectional scans of various body parts, including the head as shown in the Figure I.2 and other body components including the shoulders, spine, and heart. The CT scan is a non-invasive method of viewing the interior of the body [Wang, 2021].

(a) (b)

Figure I.2. Computer tomography (CT): a) CT brain image and b) CT scanner.

I.3.3. X-ray

X-ray waves are a form of high-energy electromagnetic radiation first discovered in 1895 by Wilhelm Rontgen. Since its invention, X-ray waves have been widely used for medical imaging because they can penetrate through the body and provide images of various sections of the body in varying colors of black and white, it's widely used for the examination of bone fractures and breaks, tooth problems.

X-ray imaging is non-invasive, painless, and quick. X-rays, but in the other side, expose individuals to radiation. As a result, it should be taken carefully. Figure I.3 shows X-ray images [Wang, 2021].

(a) (b)

Figure I.3. X-Ray image: a) X-ray scan from a body and b) X-ray scanner.

(a) (b)

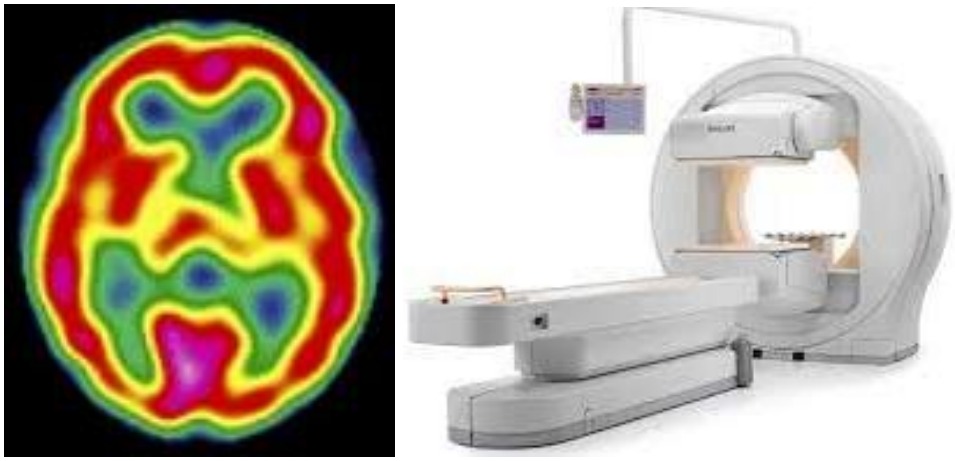
Figure I.4. PET imagery: a) PET for a healthy and depressed brain and b) PET scanner.

I.3.4. Positron Emission Tomography (PET)

Positron Emission Tomography (PET) is a type of nuclear medicine imaging that uses radioactive compounds called radiotracers to visualize changes in metabolic pathways, blood flow, regional chemical composition, absorption, and also other variables, A little volume of fluid radioactive material is injected into the body as a tracer [Wang, 2021].

I.3.5. Single-Photon Emission Computed Tomography (SPECT)

Single-Photon Emission Computed Tomography (SPECT) is another kind of nuclear medicine tomographic imaging technology that uses gamma rays, SPECT can give actual 3D information that is normally displayed as cross-sectional slices through patients and can be reformatted and altered to meet the needs of the application [Wang, 2021].



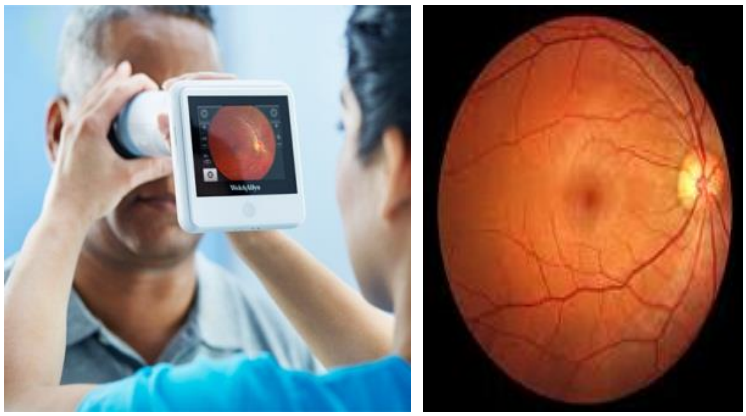
(a) (b)

Figure I.5. SPECT images: a) SPECT image from a brain healthy and b) SPECT scanner.

I.3.6. RGB data pictures

In this type, healthcare practitioners take RGB images of the organs the patient needs. Usually, these images are taken with a microscope or a high-precision camera. In this work, we focus our research on retinal fundus images, an important imaging modality for studying the internal structure of the eye. It is primarily used to identify pathological or physiological changes within the retina in a non-invasive and less time-consuming manner. These images are recorded using a fundus camera based on the

principle of a pair of monocular images. The retinal fundus images are taken under the supervision of experts to check for abnormalities within the retina as shown in Figure I.6. [Thakur.2020].



(a) (b)

Figure I.6. a) Retinal scanner and b) retinal scan.

Diabetic Retinopathy (DR) is a diabetic complication in which the blood vessels of the retina enlarge and leak fluids and blood. When DR reaches an advanced level, it might cause vision loss.

DR is responsible for 2.6 percent of blindness worldwide. Diabetes patients who have been sick with the condition for a long time are more likely to have DR. Regular retina screening is critical for diabetic patients to diagnose and treat DR at an early stage to avert blindness. The emergence of various types of lesions on a retina scan indicates the presence of DR. Microaneurysms (MA), haemorrhages (HM), and soft and hard exudates are examples of these lesions (EX).

I.3.6.1. Microaneurysms (MA):

MA is the first signs of DR and appears as little red spherical spots on the retina due to vessel wall weakening. There are sharp borders and the size is less than 125 μm. MA is divided into six categories as illustrated in Figure I.7. AOSLO reflectance and conventional fluorescein imaging were used to detect the different forms of MA.

I.3.6.2. Haemorrhages (HM):

HM occurs as bigger patches on the retina with an uneven edge that is more than 125 μm in size. Flame (superficial HM) and blot HM are the two forms of HM (deeper HM). Figure I.7 shows one sample of HM image.

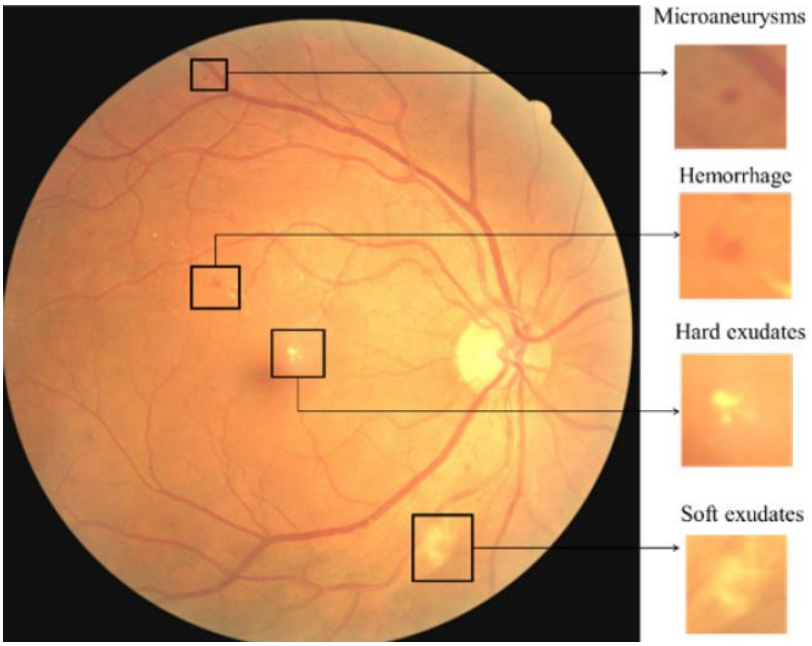


Figure I.7. The different types of HM.

a) *Hard exudates:* appear on the retina as bright-yellow spots caused by plasma leaking. They have sharp borders and are present in the outer layers of the retina.
b) *Soft exudates:* (also known as cotton wool) are white patches on the retina caused by nerve fiber swelling. The form is oval or circular [Wejdan, 2020].

There are five major DR stages:

- Non Diabetic Retinopathy (No DR).
- Mild DR.
- Moderate DR.
- Severe DR.
- Proliferative Diabetic Retinopathy (PDR).

No DR causes retinal blood vessels to be injured and the retina to swell because of glucose buildup, resulting in blood vessel leakage in the eyes and partial vision loss [Pratiksha.2021].

Mild DR: At this stage of the disease, little patches of balloon-like swelling in the retina's tiny blood vessels, known as microaneurysms, appear. These microaneurysms may cause fluid to flow into the retina [Khalifa, 2019].

Moderate DR: As the condition advances, blood vessels that nourish the retina expand and distort, and their ability to transport blood is lost. Both disorders modify the appearance of the retina and may lead to diabetic macular edema (DME) [Khalifa, 2019].

Severe DR: Many more blood vessels are clogged, depriving portions of the retina of blood supply. These regions secrete growth factors, which instruct the retina to form new blood vessels [Khalifa, 2019].

PDR is a significantly more severe form of diabetes-related eye disease. It starts with the formation of new blood vessels in the retina. This is known as neovascularization. Such fragile new-fangled vessels can bleed inside the vitreous. Some dark floaters form when they bleed a small bit. If they leak excessively, they may obscure all visibility. Such new-fangled blood arteries may produce scar tissue, which may cause the macula problem or lead to a detached retina.

Patients may not have any symptoms in the early stages of DR development, but as the patient's condition worsens, headways, spots or shady strings moving in the patient's vision (floaters), blurred vision, fluctuating vision, impaired vision, dark or purging areas appear gradually, eventually leading to loss of vision. As a result, diagnosing the DR in its early stages is necessary to avert the most terrible consequences of the later phases [Robiul, 2020].

I.4. ARTIFICIAL INTELLIGENCE

I.4.1. Definition of AI

Artificial intelligence (AI) is defined as machine intelligence as opposed to human or other living species intelligence. AI also refers to scenarios in which machines can learn and analyze in the same way that humans can, and hence assist in problem solving.

In recent years, artificial intelligence (AI) has advanced fast in terms of software algorithms, hardware implementation, and applications in a wide range of fields [Rong, 2019].

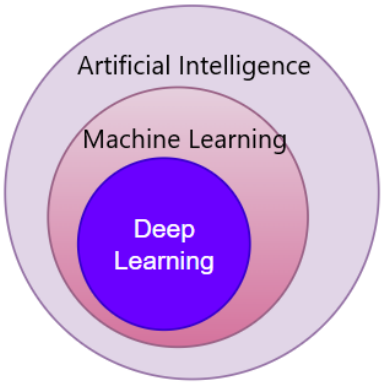


Figure I.8. AI composition.

I.4.2. History of AI

Artificial intelligence (AI) was initially described in 1950, but early models had significant flaws that precluded widespread adoption and application in medicine. Many of these constraints were overcome by the emergence of deep learning in the early 2000s. Now that AI systems are able to process complex algorithms and self-learning, we are entering a new era in medicine, where AI may be used in clinical practice to improve diagnosis accuracy and workflow efficiency through risk assessment models [Kaul, 2020].

"Expert systems" arose at the start of the 1980s as a new type of AI. The Japanese government put a lot of effort into promoting this AI. Back propagation was one of the most important advancements in neural network training throughout that decade. By the end of the 1980s, investor funding had fallen once more, ushering in AI's "second winter."

In the 90s, AI finally achieved some of its most cherished objectives. The IBM DeepBlue project, which culminated to the defeat of global chess champion Gary Kasparov in 1997, was a watershed moment in the company's history. This was partly owing to the advancement of processing power.

Finally, it found use in a wide range of fields. The Google-Alphago project, which allowed an AI-powered machine to defeat the world Go champion in 2017, was a significant milestone for deep learning applications [Bouletreau, 2019].

I.4.3. Machine Learning (ML)

I.4.3.1. Definition

Machine learning (ML) is a form of artificial intelligence that enables machines to think like humans and make decisions without the need for human interaction. It's the process of making robots learn without having to be explicitly programmed. Machine Learning's main goal is to create a computer software that can read data and learn from it [Ibrahim, 2021].

I.4.3.2. Machine learning types

As shown in Figure I.9, machine learning algorithms are separated into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [Sarker, 2021].

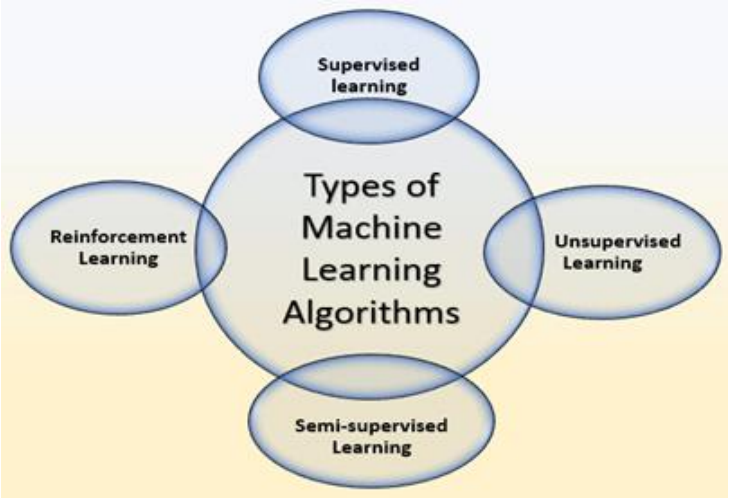


Figure I.9. Types of Machine Learning Techniques.

a) *Supervised:* The objective of supervised learning in machine learning is to train a function that translates an input to an output using sample input-output pairs [Sarker, 2021].
b) *Unsupervised:* Unsupervised learning is a data-driven method for analyzing unlabeled datasets without involving humans [Sarker, 2021].

c) *Semi-supervised:* Semi-supervised learning is a combination of the supervised and unsupervised approaches stated above, as it works with both labeled and unlabeled data [Sarker, 2021].

d) *Reinforcement:* Reinforcement learning is a form of machine learning technique that allows software agents and computers to automatically assess the most efficient behavior in a given situation or environment [Sarker, 2021].

I.4.3.3. Some Machine learning techniques

To identify the disease, different machine learning algorithms are applied. We can cite the most used of them in the following section**:**

a) *Decision Trees (DT):* The decision tree is a well-known non-parametric supervised learning approach. This technique is used in both classification and regression issues [Sarker, 2021].

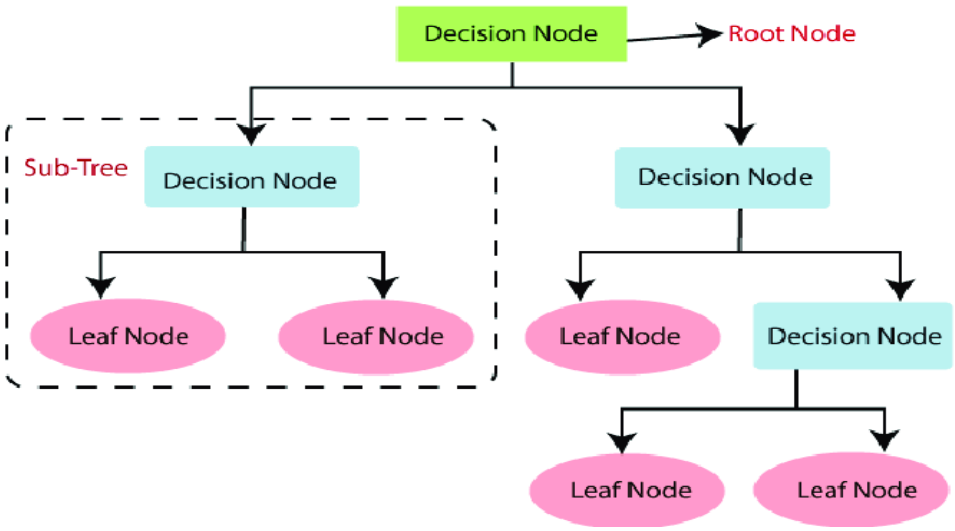


Figure I.10. Decision tree (DT) architecture.

b) *Support Vector Machine (SVM):* A support vector machine is a popular machine learning tool for classification and regression [Sarker, 2021], it is one of the most effective supervised learning methods [Stephenson, 2019].

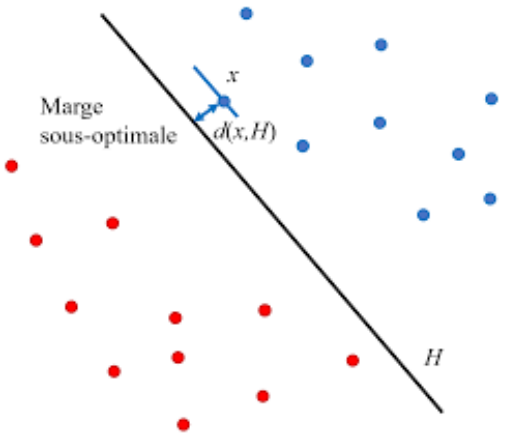


Figure I.11. SVM principle.

c) *K-nearest neighbors (KNN):* KNN is a non-generalizing or "instance-based learning" algorithm, sometimes known as a "lazy learning" technique. Rather of building a broad internal model, it keeps all instances corresponding to training data in n-dimensional space [Sarker, 2021]. The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

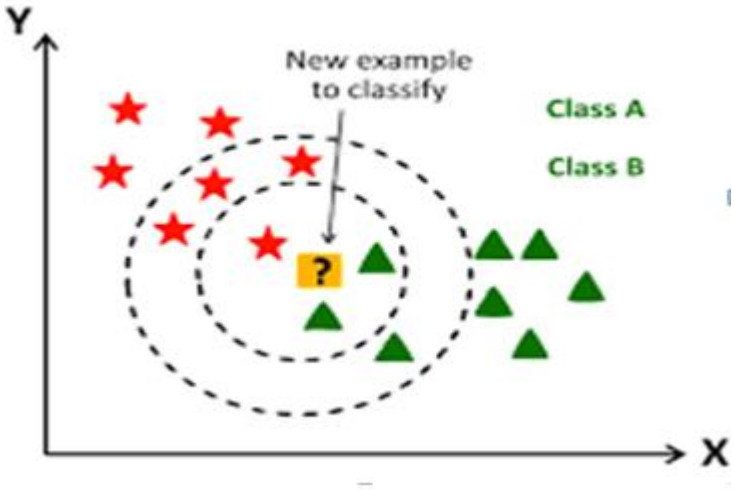


Figure I.12. KNN process.

d) *Artificial Neural Network (ANN):* The artificial neural network (ANN) is a mathematical model based on nonlinear statistical data modeling techniques in which complicated interactions between inputs and outputs emerge [Mintz, 2019]. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. ANN is trained by processing examples, each of which contains a known "input" and

"result," forming probability-weighted associations between the two, which are stored within the data structure of the net itself.

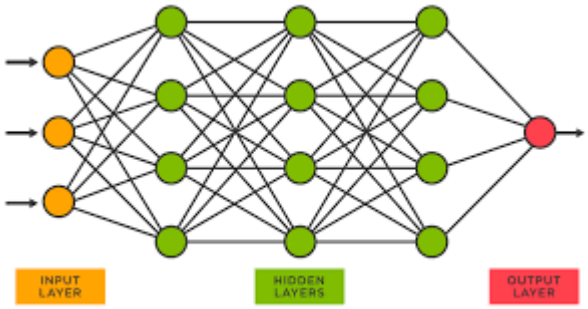


Figure I. 13. Artificial Neural Network (ANN).

I.5. DEEP LEARNING

I.5.1. definition

Deep learning is a type of ANN-based machine learning that is structured similarly to human brain processing and takes several data sets into account at the same time [Mintz, 2019].

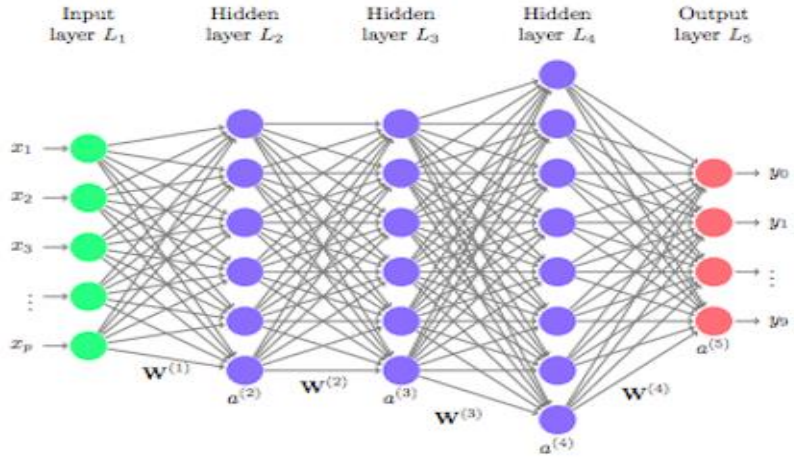


Figure I.14. Deep learning structure.

I.5.2. Some Deep Learning techniques

Several popular deep learning networks, such as the Recursive Neural Network (RNN), Convolutional Neural Network (CNN), and others, are discussed in this section [Pouyanfar, 2018].

I.5.2.1. Restricted Boltzmann Machines (RBMs):

In 1986, the Restricted Boltzmann Machine (RBM) was proposed as a generative stochastic neural network. An RBM is a Boltzmann Machine version with the requirement that the visible and hidden units form a bipartite graph. This constraint makes training methods more efficient, particularly the gradient-based contrastive divergence approach [Guo, 2016].

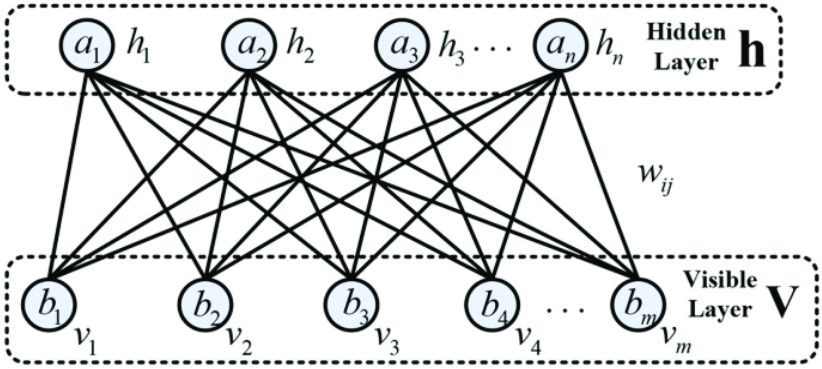


Figure I.15. RBN architecture.

In the next sections, we will expose these models and describe respectively.

a) *Deep belief network (DBN):* Hinton 21 proposed the Deep Belief Network (DBN), which was a significant development in deep learning. It's a probabilistic generative model that generates a probabilistic distribution over observable data and labels. To establish the deep network, a DBN uses an efficient layer-by-layer greedy learning technique, and then fine-tunes all of the weights together with the desired outputs [Guo, 2016].
b) *Deep Boltzmann Machines (DBMs):* Another deep learning system, the Deep Boltzmann Machine (DBM), uses layers to organize the units. The DBM contains undirected connections throughout its structure, unlike DBNs, which have an undirected graphical model at the top and a directed generative model at the bottom [Guo, 2016].
c) *Deep Energy Models (DEMs):* A more recent way to training deep structures is the Deep Energy Model (DEM). Unlike DBNs and DBMs, which have several stochastic hidden layers for efficient training and inference, the DEM only has a single layer of stochastic hidden units for fast training and inference [Guo, 2016].

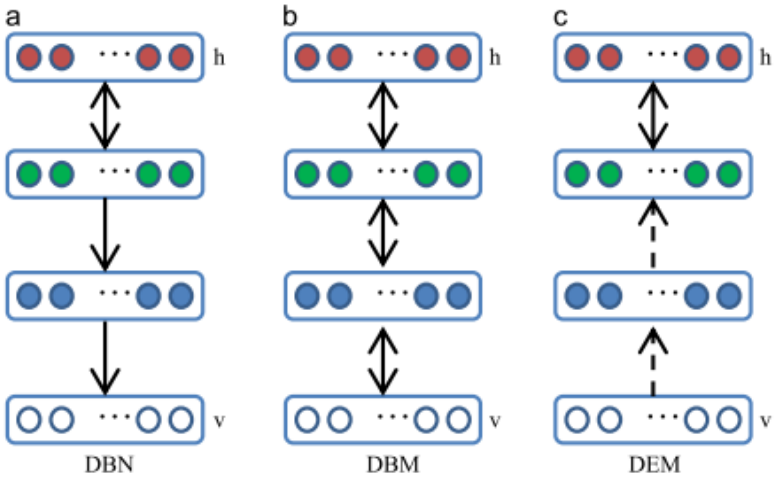


Figure I.16. RBN models.

I.5.5.2. Autoencoder:

This type of DP algorithm is a form of ANN that is used to learn effective encodings. An autoencoder is trained to rebuild its own inputs rather than training the network to predict some target value given inputs. Figure I.17 shows the general operation of an autoencoder [Guo, 2016].

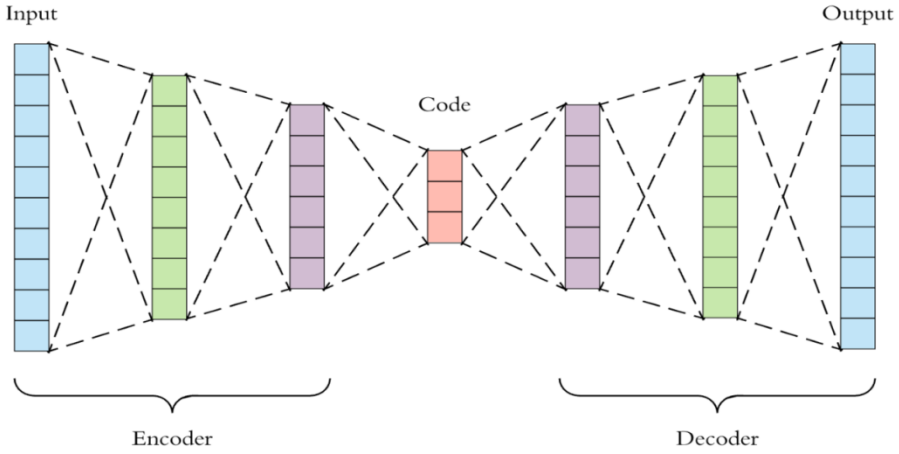


Figure I.17. Autoencoder architecture.

I.5.5.3. Convolutional Neural Network (CNN): a form of ANN that analyzes data using deep learning methods with numerous hidden layers. The interactions between layers are complicated (hence the word convolutional), and each CNN has many hidden layers [Mintz, 2019]. In the next section, we explain in more details the principle of CNN and its application for classification purpose.

I.6. Convolutional Neural Network

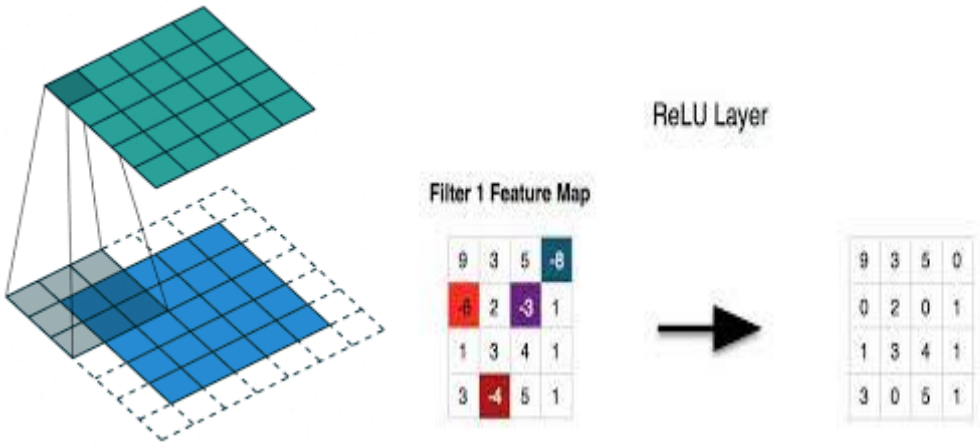
I.6.1. Principle:

CNN is a type of deep neural network that is created artificially. Image classification, segmentation, and recognition are all done with CNNs. CNNs' main tasks include classifying visual content, recognizing items supplied into it, and grouping the detected objects into clusters. CNNs rely on cross-unit connections and weights, which are then subsampled. A simple CNN design consists of one convolutional layer and a pooling layer, with fully connected layers for supervised prediction added on occasion. Hence, we describe in more detail the CNN's architecture through its different levels [Manne, 2021].

I.6.2. CNN architecture

I.6.2.1. Input layer: The image data should be in the CNN's input layer. Image data is a three-dimensional matrix that must first be flattened into a single column before being fed as an input. Each layer's output will be fed into the following layer [Manne, 2021].

I.6.2.2. Convolutional layer (convo + ReLu): The convolutional layer is where the action begins. This layer recognizes visual characteristics such as color, shape, and object elements, among others. The ReLu layer is a CNN extension layer. This layer aids in increasing the nonlinearity of the image and allows for better feature extraction. Figure I.18 shows the convolution principle as well as the ReLu function's role.



(a) (b)

Figure I.18. a) Convolution layer and b) Relu architecture.

I.6.2.3. Pooling layer: After the convolutional layer, the pooling layer lowers the spatial volume of the input image [Manne, 2021].

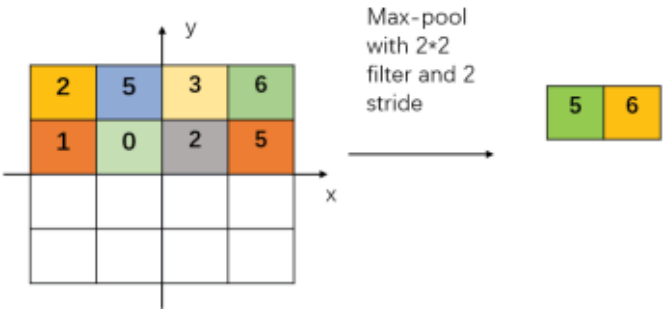


Figure I.19. The pooling process.

I.6.2.4. Fully connected layer: Weights, neurons, and biases are all found in this layer. Using a fully connected layer, neurons in one layer are connected to neurons in another layer [Manne, 2021].

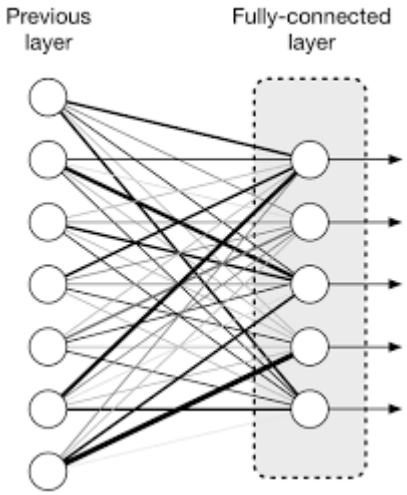


Figure I.20. Fully connected layer.

I.6.2.5. Softmax layer and output layer: These are CNN's last layers. The FC layer is followed by the Softmax layer, which is utilized for binary or multi classification. In addition, the final output label of the input image fed into the input layer will be provided by the output layer [Manne, 2021].

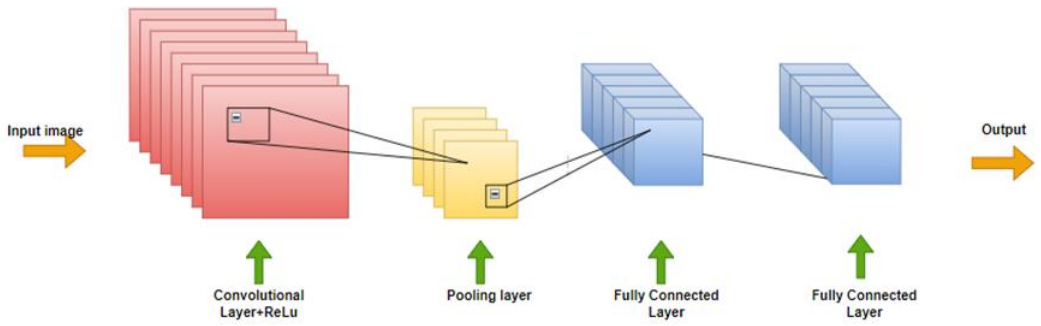


Figure I.21. Architecture of a simple CNN.

I.6.3. CNN models type

I.6.3.1. AlexNet

AlexNet is a simple structure made up of five convolutional layers stacked with pooling and/or nonlinear layers, then three fully connected layers. It employs innovative techniques such as data augmentation (DA), dropout, rectified linear units (ReLUs), local response normalization (LRN), and overlapping pooling [Wang, 2022].

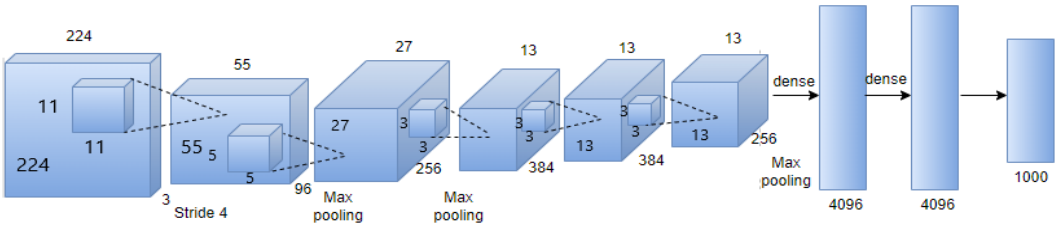


Figure I.22. AlexNet architecture.

I.6.3.2. GoogLeNet

The GoogLeNet model was proposed in a 2015 research called Going Deeper with Convolutions. The model took first place in the 2014 ILSVRC competition using a redesigned inception module. A set of parallel convolutional layers with varied filter sizes and a maximum pooling layer make up the inception module. This module's output is then concatenated. The following are the model's main features: extensive use of 1x1 convolution to reduce the number of channels; error feedback at multiple points

throughout the network; production of exceptionally deep (22 layer) models; and use of global average pooling for the models output [Adewole, 2020].

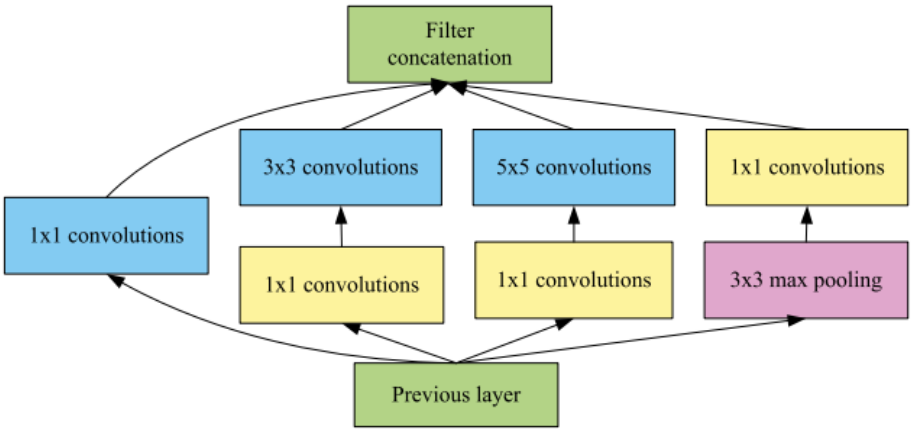


Figure I.23. Inception module with dimensionality reduction.

I.6.3.3. ResNet

In 2016, the residual network (ResNet) was proposed in publication titled Deep Residual Learning for Image Recognition. The 152-layer model suggested the usage of leftover blocks with shortcut connections. These are simple network architecture connections in which the input is kept unweighted and passed on to a deeper layer. The output of a residual block is mixed with the input of the block in a pattern of two convolutional layers with ReLU activation [Adewole, 2020].

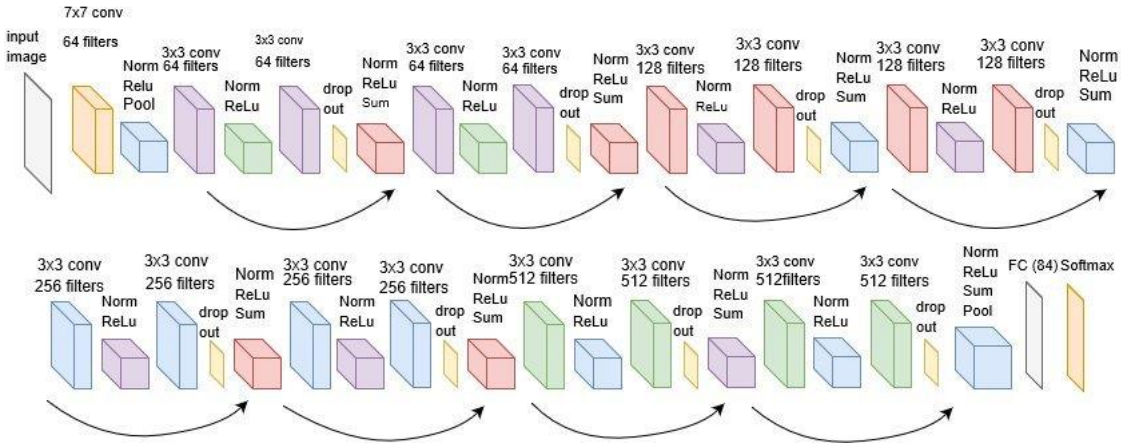


Figure I.24. RasNet architecture.

I.6.3.4. VGG-Net

The VGG-Network was proposed by the Visual Geometry Group at Oxford in a paper titled Very Deep Convolutional Networks for Large-Scale Image Recognition in 2014 [Adewole, 2020]. The utilization of a large number of small filters, specifically 3x3 and 1x1 filters with a stride of one, is one of the model's most notable features.

Most, but not all, convolutional layers are followed by 2x2 max pooling layers. Before using a max pooling layer, the network includes examples of two, three, and even four convolutional layers piled together. With the number of filters growing as the model's depth increases, various model versions have been built and tested, each named after the number of layers: VGG-16 and VGG-19 are two different types of VGG [Adewole, 2020].

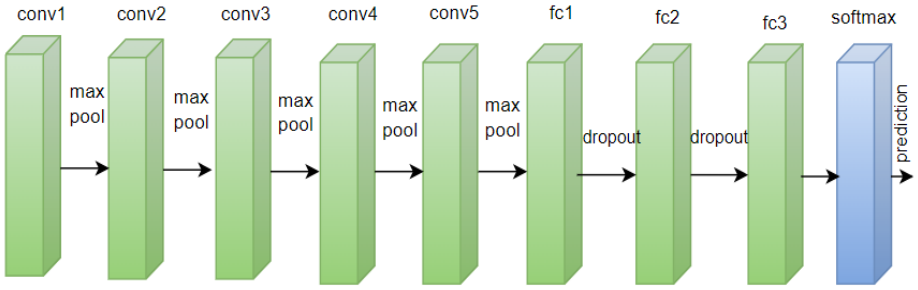


Figure I.25. VGG-Net architecture.

I.7. CNN MODEL TRAINING

I.7.1. Transfer Learning

Transfer learning is a model transformation technique for small datasets, ie. transfer the learned parameters of the model to a new model that is well pretrained on a large labeled dataset (e.g. ImageNet). Alternatively, even if the available dataset is large enough, it can be helpful to start with pretrained weights rather than randomly initialized weights. In this case, the weights of the pretrained model are fine-tuned to satisfy the new model [Wang, 2022].

I.7.2. Fine Tuning

Fine-tuning methods are one of the most popular transfer learning strategies in neural network applications. Its use is driven by the transfer of knowledge from generative to discriminative models, resulting in a high degree of generalization. The

initial pipeline consists of a pretrained network in which the last classifier layer is replaced by a randomly initialized layer [Vrbančič, 2020].

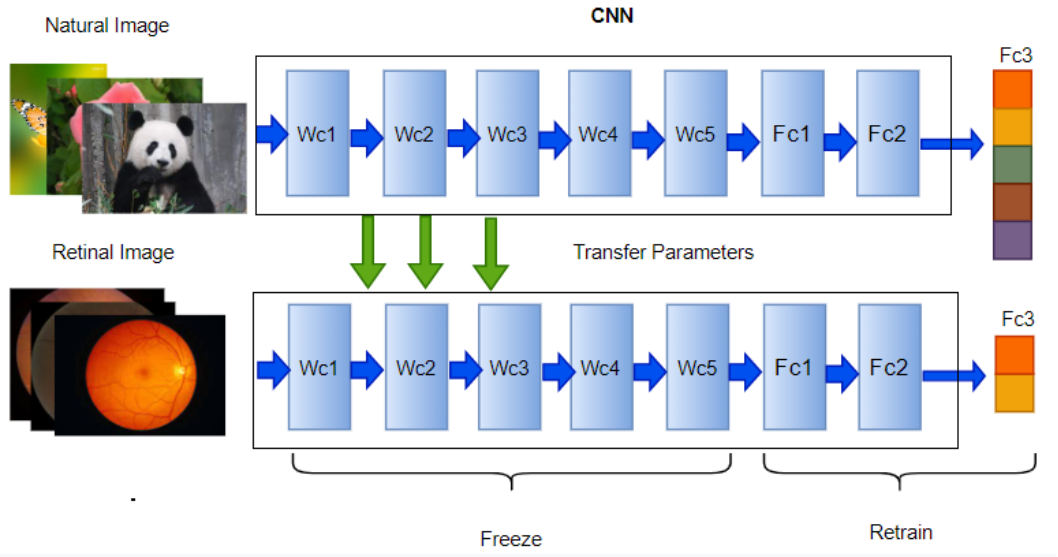


Figure I.26. The conceptual process of transfer learning technique.

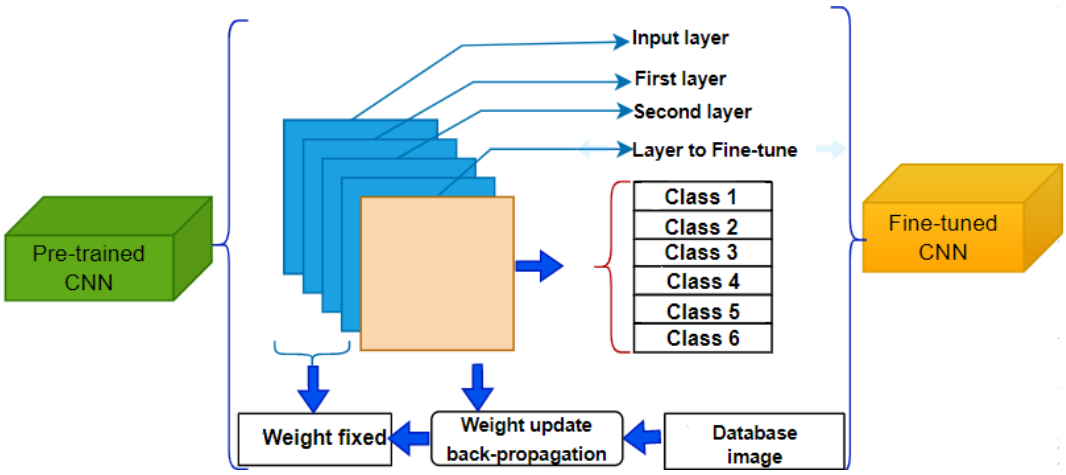


Figure I.27. The conceptual process of Fine tuning technique.

I.8. CONCLUSION

In this chapter, we presented a general description of medical imaging and we explained the different data type used in healthcare. Then, we discussed the theoretical backgrounds, the concepts and types of machine and deep learning techniques. Finally, we presented the most common CNN architectures.

In the second chapter, we describe our proposed system and we provide more detail about its different modules.

CHAPTER II: DIABETIC RETINOPATHY DIAGNOSIS

II.1. INTRODUCTION

In this chapter, we describe first the dataset used in our work. Then, we discuss the literature of diabetic retinopathy diagnosis in the last years. We also present our proposed system for diabetic retinopathy prediction. A detailed description of each part of our system is provided to give an idea of the methods and techniques used in this work.

II.2. DATA BASE

The experiments of our study are conducted on the blindness detection dataset APTOS 2019 using different deep transfer learning models. The dataset was released by the Asia Pacific Tele-ophthalmology Society in 2019. The dataset contains 3662 images, it's consists of 5 categories (No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR) [Khalifa, 2019]. Figure II.1 shows some fundus images extracted from this dataset.

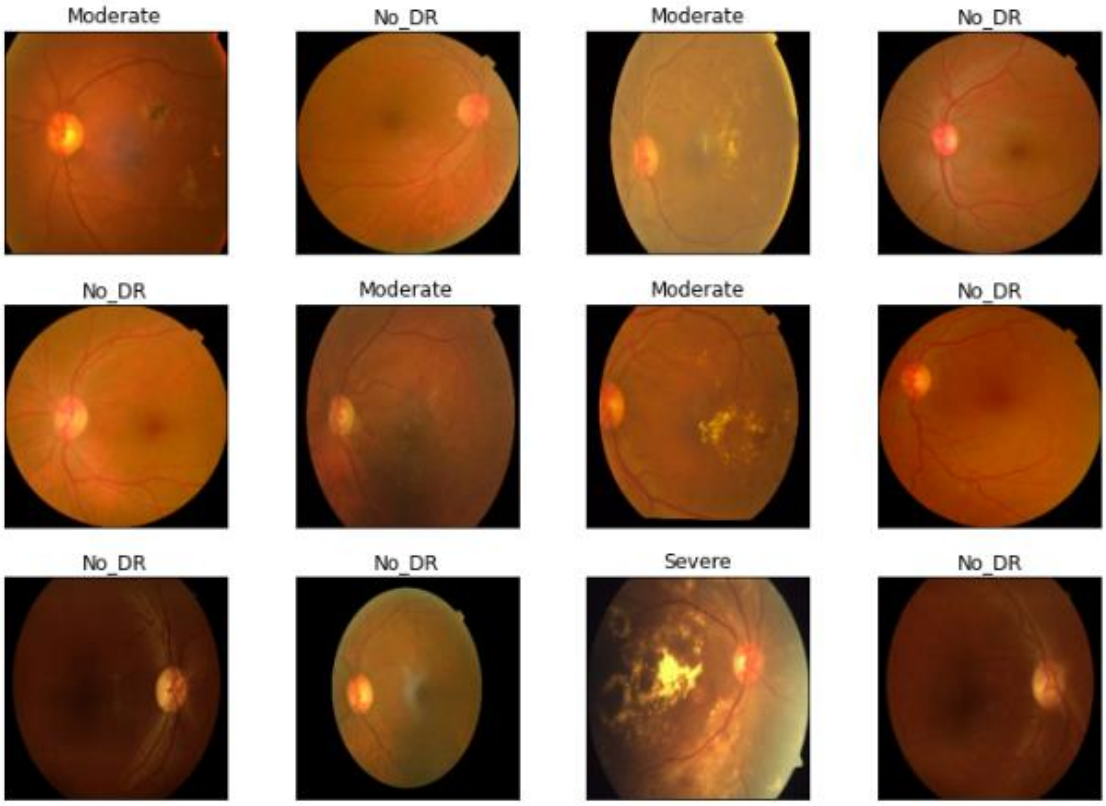


Figure II.1. Some samples from different classes of APTOS2019.

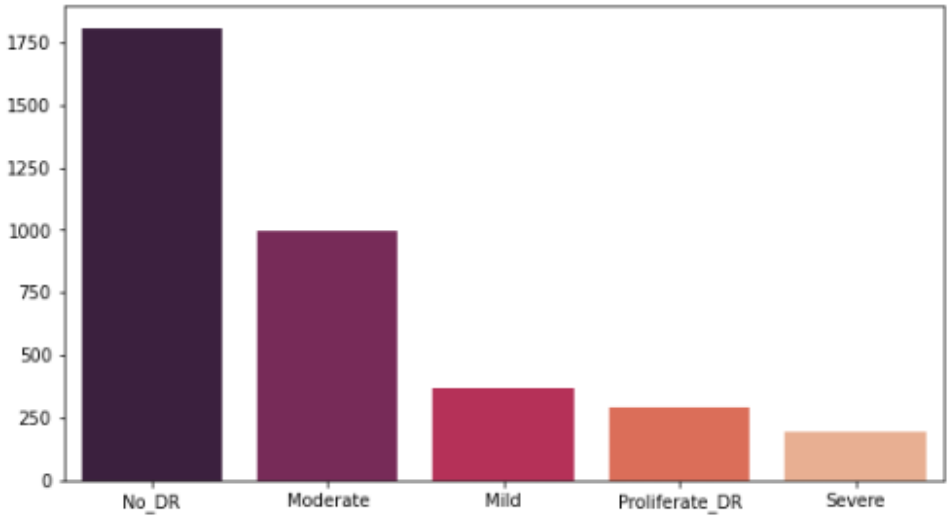


Figure II.2. Number of samples for each class in the original APTOS2019 dataset.

As we can see, Figure II.2 illustrates the composition of the dataset on different classes. Besides, Table II.1 summaries the numeral distribution of this dataset.

Table II.1. Summary of APTOS 2019 challenge dataset.

Data	**APTOS2019**
No of images	3662
No of classes	5
# of images in class 0	1805
# of images in class 1	370
# of images in class 2	999
# of images in class 3	193
# of images in class 4	295

II.3. STATE OF THE ART

Diabetic retinopathy (DR) is one of the main cause's blindness worldwide. With proper treatment and early diagnosis can prevent the development of the disease. Many researchers have relied on artificial intelligence for early detection of this disease by applying different techniques. We will see the most important studies and results in the last few years.

A group of researchers presented a new method for extracting features using Xception model for DR disease diagnosis. The proposed method is based on deep layer

aggregation, which collects Multi-level features from different convolutional layers of Xception Architecture. Later, the retrieved features are sent into a multilayer perceptron (MLP) that is trained on the DR classification intensity. Four deep-feature extracts, including InceptionV3, MobileNet, ResNet50, and the original Xception architecture, were used to evaluate the suggested approach's performance. The assembly CNN deep layers may efficiently combine deep features and improve the learning process as compared to a normal Xception architecture. Experiments show that the average Deep feature Xception extractor enhanced DR classification accuracy by 83.09 % compared to 79.59 % for the conventional Xception architecture [Kassani, 2019].

Mishra and colleagues [Mishra, 2020] developed a DenseNet model on a large dataset of 3,662 training images for automatically detecting DR stages and classifying the fundus images in high quality. They used the DR Kaggle APTOS dataset with five classes. DenseNet was then trained to extract the pertinent features. The performance of the system was 96.11% in term of Squared Weighted accuracy.

Similarly, Vinuja et al. [Vinuja, 2021] explored two models for diabetic retinopathy classification: InceptionV3 and Xception. They used also the wellknown APTOS2019 with 5 classes to train and test the two models. Data augmentation and preprocessing has been used to enhance the performance of their models. After completing preprocessing and boosting on the dataset, they evaluated their models Xception was determined to have the greatest performance with 93.10 % in term of accuracy. InceptionV3 gave only 91.90% of accuracy.

Nair et al. [Nair, 2022] made comparative analysis for severity-based deep neural network to classify the DR disease into mild, moderate, severe and proliferative stages. Pre-training architectures such as VGG16, EfficientNetB5, and ResNet50 have been used on Kaggle APTOS, which provided 76.47%, 90.2% and 97.2% in term of accuracy, respectively.

Bodapati et al. [Bodapati, 2020] evaluated a Diabetic Retinopathy (DR) recognition where they focused on the representation of retinal images. Features extracted from multiple pre-trained ConvNet models are blended using proposed multi-modal fusion module. These final representations are used to train a Deep Neural Network (DNN) used for DR identification and severity level prediction. Experimental

results were conducted on benchmark Kaggle APTOS 2019 contest dataset and showed that the model trained on proposed blended feature representations is superior to the existing methods.

Finally, two works [Vaibhavi, 2021], [Anoop, 2022] have used a convolutional neural network architecture to detect the presence or not of DR. They performed a comparison between various optimizers like Adagrad, RMSPROP Momentum and Adam. The Kaggle Blindness Detection APTOS-2019 dataset is used. Table II.2 summarizes the state of the art method using APTOS2019 dataset.

Table II.2. State-of-the-art methods for DR classification.

Reference	**# of training data**	**# of testing data**	**Method**	**Accuracy (%)**
[Kassani, 2019]	2657 images (5 classes)	662 images (5 classes)	Xception InceptionV3 MobileNet ResNet50 Modified Xception model	79.59 78.72 79.01 74.64 83.09
[Tymchenko, 2020]	3662 images (5 classes)	1928 images (5 classes)	VGG16 DenseNet121	73.26 96.11
[Vinuja, 2021]	2746 images (5 classes)	549 images (5 classes)	Inception V3 No Aug No Prep Only Aug Only Prep With Aug With Prep Xception No Aug No Prep Only Aug Only Prep With Aug With Prep	62.20 91.10 89.40 91.90 81.70 92.90 90.50 93.10
[Bodapati, 2020]	2930 images (2 classes)	732 images (2 classes)	VGG16-f1 VGG16-f2 Xception Inception ResNtV2	97.27 97.32 97.41 97.34

[Vaibhavi,2021]	2520 images (2 classes)	1080 images (2 classes)	NA	93.00
[Anoop, 2022]	2520 images (2 classes)	1080 images (2 classes)	Proposed CNN Model	94.60

Aug: Augmentation, **Prep:** Preprocessing

II.4. PROPOSED DIABETIC RETINOPATHY DIAGNOSIS SYSTEM:

The main objective of this work is to build a robust and automated system for the detection of diabetic retinopathy. This work uses deep learning methods to classify the diabetic retinopathy severity (no DR, mild, moderate, severe and diffuse). Our system is divided into three principal modules namely: preprocessing, feature extraction using transfer learning and classification.

It is worth noticing that, in our work and for the first time, we try to emulate the medical protocol applied by healthcare professionals and integrate it in our system with automatic manner. In fact, the medical protocol for DR prediction needs firstly to check whether or not a person has DR. If the patient has DR, the medical conduct is then oriented towards the severity of DR. At this level, we can divide the degree of severity into two classes and each one is also divided into two subclasses as shown in Figure II.3. Finally, we calculate the overall accuracy of our system by combining the accuracies of each level.

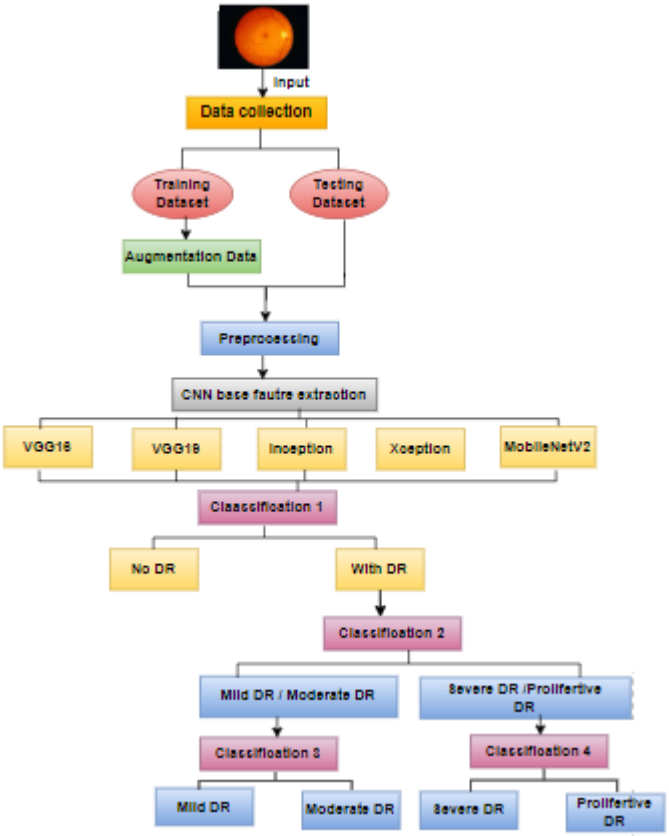


Figure II.3. The overall proposed system scheme.

II.4.1. Preprocessing

II.4.1.2. Gray scale conversion

The principle preprocessing operation when dealing with RGB images is converting color images to gray scale ones for representing each pixel value by its intensity [Kanth, 2013], the conversion is given by:

$$\mathbf{GRAYSCALE} = 0.30\,\mathrm{R} + 0.59\,\mathrm{G} + 0.11\,\mathrm{B} \qquad (\mathbf{1})$$

Where:

R: Red, **G**: Green, **B**: Blue.

The result of converting some RGB images into gray level ones is depicted in Figure II.4.b.

II.4.1.1. Gaussian filter

In order to create an enhanced dataset inspired from the original one, Gaussian filter is applied on all DR images in APTOS2019 dataset. This operation can increase the performance our system by creating different samples to train the deep learning model. The resulted images are rich in such way the classification system performs better and more accurate. Figure II.4.c shows the images obtained after applying Gaussian filter.

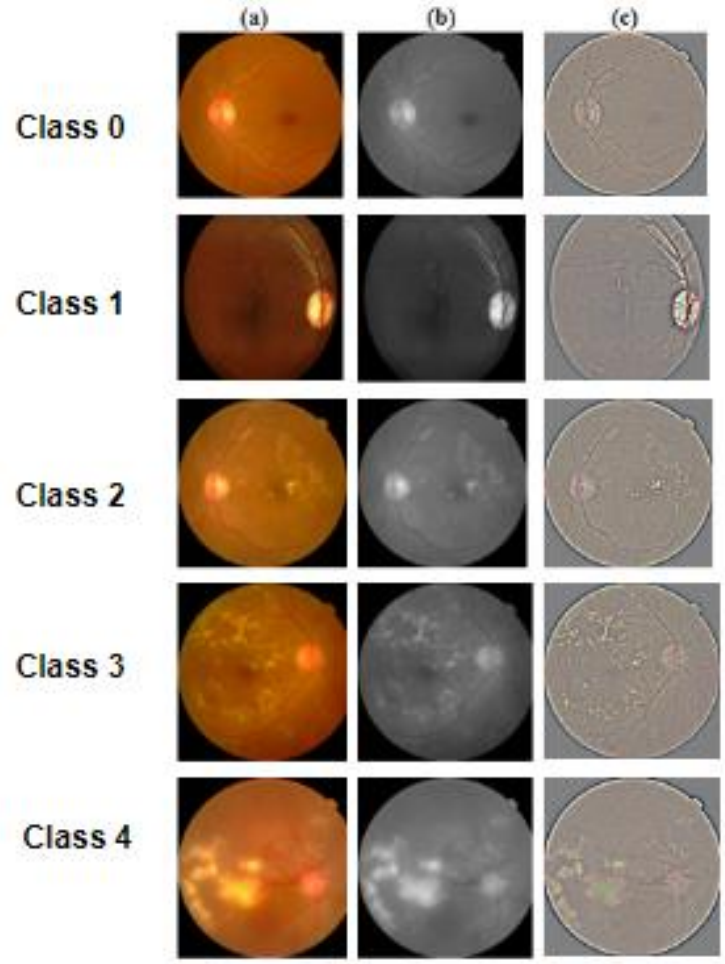


Figure II.4. Preprocessing task: a) Original images, b) Grayscale images, c) Gaussian filter images.

II.5. FEATURE EXTRACTION USING DEEP LEARNING

II.5.1. VGG16 Model

VGG16 includes 13 convolution layers, all of which employ 3x3 convolution kernels with a stride of one, and the same padding. For down sampling, max pooling

layers are utilized. Then there will be a Softmax classifier, followed by two fully connected layers with 4,096 nodes each [Robiul Islam. 2020]. Figures II.5 and II.6 present the VGG16's architecture and its summary, respectively.

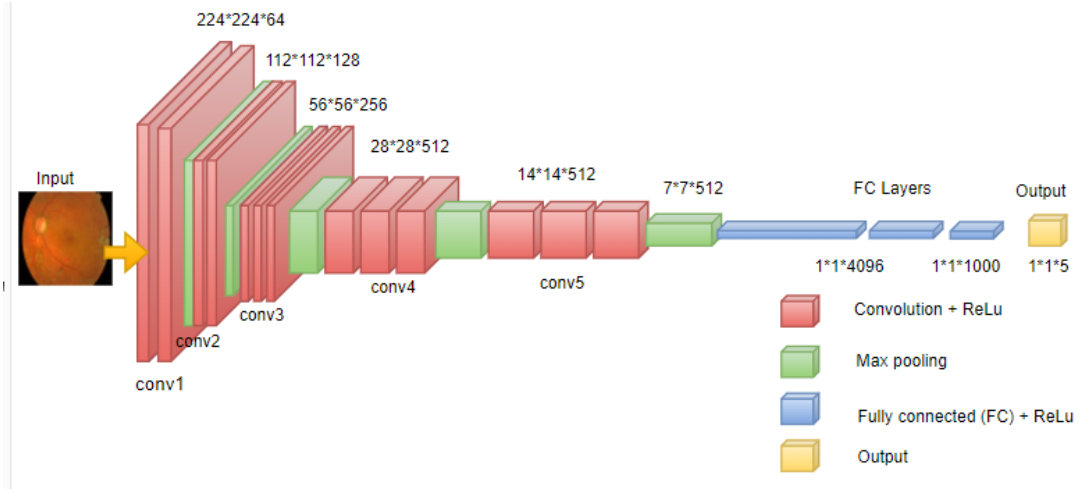


Figure II.5. The VGG16 architecture.

block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_1 (Dense)	(None, 2)	50178

Total params: 14,764,866
Trainable params: 50,178
Non-trainable params: 14,714,688

Figure II.6. Summary of the VGG16 model.

II.5.2. VGG19 Model

VGG19 model for image classification and it consists of 19 layers. Instead of using a large filter the size used in Alexnet is 11*11, a multiple of 3 by 3 Filters per layer are implemented in VGG19 Model. The VGG 19 model has three complete

connection layer, five pooling layers, sixteen Convolutional layers and Softmax layers. Two Fully connected layers are used with 4096 channels, then Another fully connected layer with 1000 label predictions, last Softmax transfer Classification in the last fully connected layer [kavitha, 2021]. Figures II.7 and II.8 present the VGG19's architecture and its summary, respectively.

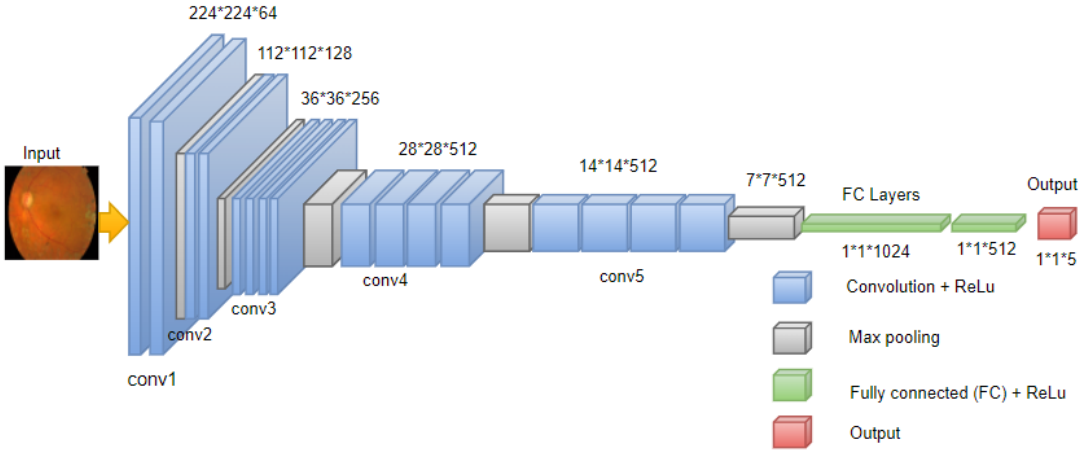


Figure II.7. VGG19 Model architecture.

```
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_conv4 (Conv2D)        (None, 28, 28, 512)       2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 14, 14, 512)       0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_conv4 (Conv2D)        (None, 14, 14, 512)       2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 7, 7, 512)         0
_________________________________________________________________
flatten_2 (Flatten)          (None, 25088)             0
_________________________________________________________________
dense_2 (Dense)              (None, 2)                 50178
=================================================================
Total params: 20,074,562
Trainable params: 50,178
Non-trainable params: 20,024,384
_________________________________________________________________
```

Figure II.8. Summary of the VGG19 model.

II.5.3. InceptionV3 Model

InceptionV3 has 48 layers and can classify 299x299 pixel input images into 1000 categories. The factorizing convolution layer is used in this architecture to reduce the number of parameters while maintaining network efficiency. This neural network's final classification layer was also modified to categorize into the desired number of classes [Kajan, 2020]. Figures II.9 and II.10 present the InceptionV3 architecture and its summary, respectively.

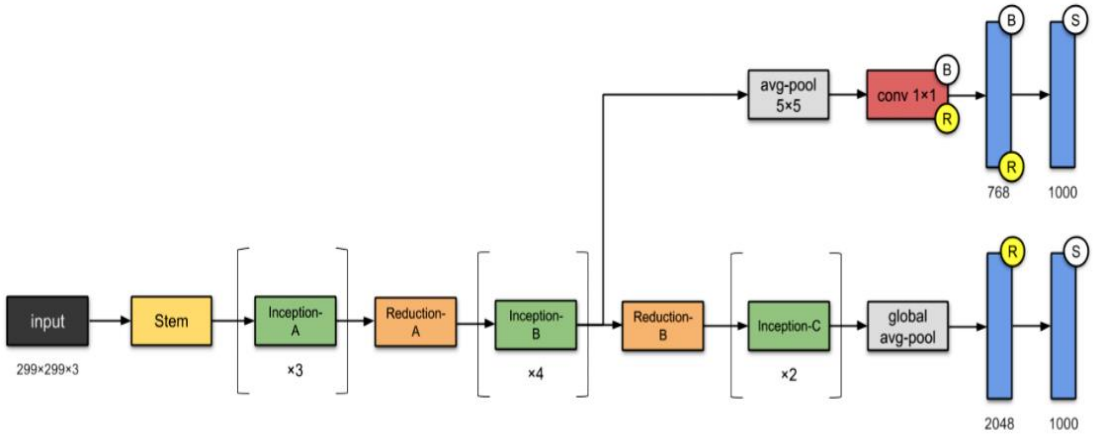


Figure II.9. InceptionV3 architecture.

batch_normalization_93 (BatchNo	(None, 5, 5, 192)	576	conv2d_93[0][0]
activation_85 (Activation)	(None, 5, 5, 320)	0	batch_normalization_85[0][0]
mixed9_1 (Concatenate)	(None, 5, 5, 768)	0	activation_87[0][0] activation_88[0][0]
concatenate_1 (Concatenate)	(None, 5, 5, 768)	0	activation_91[0][0] activation_92[0][0]
activation_93 (Activation)	(None, 5, 5, 192)	0	batch_normalization_93[0][0]
mixed10 (Concatenate)	(None, 5, 5, 2048)	0	activation_85[0][0] mixed9_1[0][0] concatenate_1[0][0] activation_93[0][0]
flatten (Flatten)	(None, 51200)	0	mixed10[0][0]
dense (Dense)	(None, 2)	102402	flatten[0][0]

Total params: 21,905,186
Trainable params: 102,402
Non-trainable params: 21,802,784

Figure II.10. Summary of the InceptionV3 model.

II.5.4. Xception Model

Xception is a CNN architecture based entirely on depthwise separable convolutional layers, as shown in Figure II.12.a, the network consists of repeated point convolutions (b) and depthwise convolutions (b). Pointwise convolution is a 1×1 convolution for changing the input dimension, while depthwise convolution is a channel-wise n×n spatial convolution. These two types of folding reduce the number of connections in Xception; making it lighter than most other CNNs. Xception is also one of the top CNNs in the ImageNet ranking. Xception uses a subset of ImageNet's 1000 categories, which contains approximately 1.3 million training images, 50,000 validation images, and 100,000 test images [Kim, 2020]. Figure II.11 demonstrates the separable convolution principle with depthwise and pointwise.

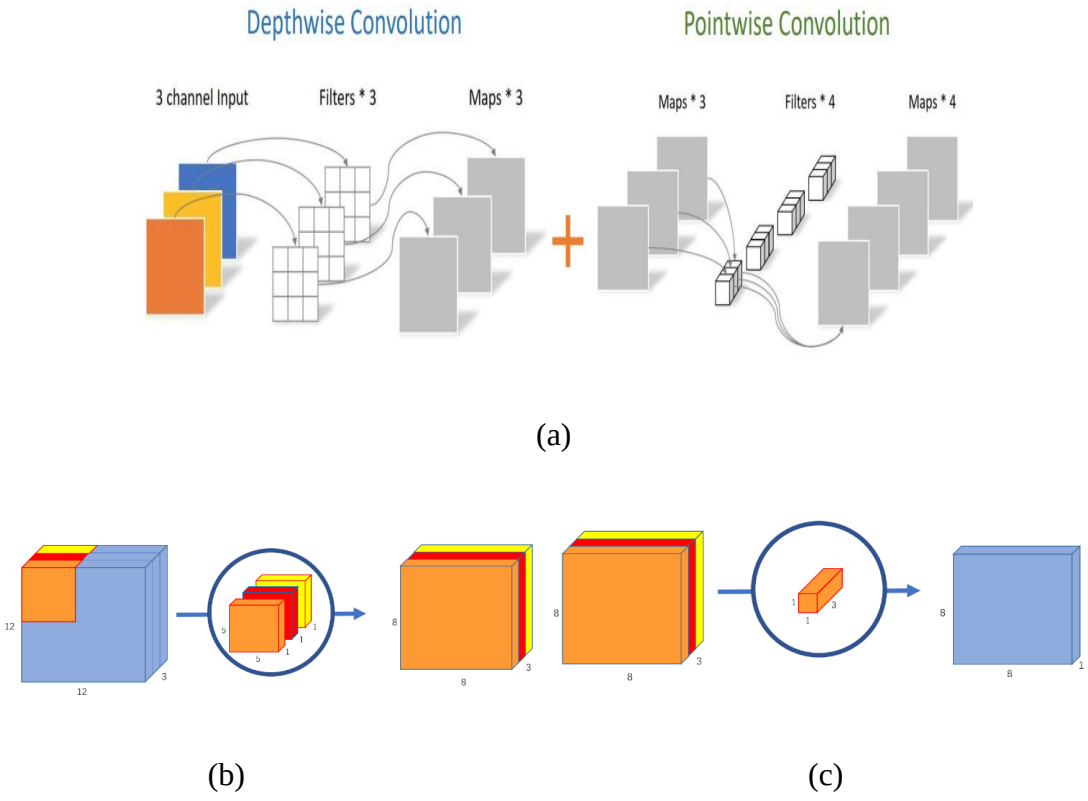


Figure II.11. a) Separable Convolution, b) Pointwise convolution, c) Depthwise convolution

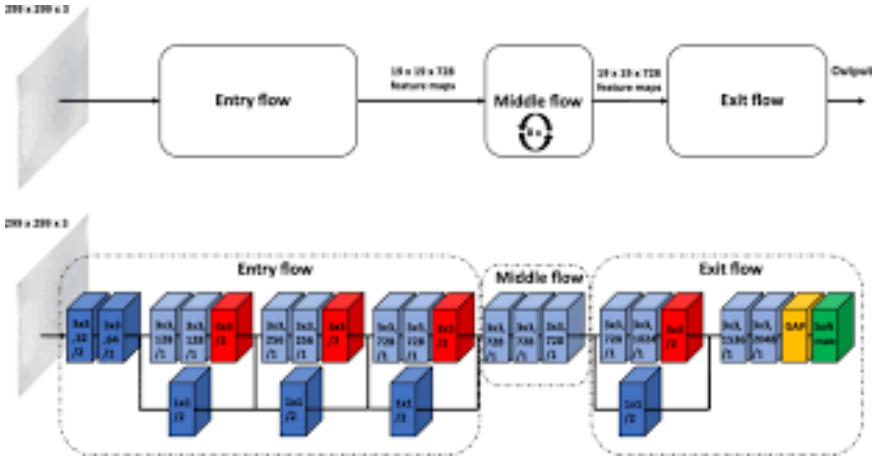


Figure II. 12. Xception architecture.

batch_normalization_97 (BatchNo	(None, 7, 7, 1024)	4096	conv2d_97[0][0]
add_11 (Add)	(None, 7, 7, 1024)	0	block13_pool[0][0] batch_normalization_97[0][0]
block14_sepconv1 (SeparableConv	(None, 7, 7, 1536)	1582080	add_11[0][0]
block14_sepconv1_bn (BatchNorma	(None, 7, 7, 1536)	6144	block14_sepconv1[0][0]
block14_sepconv1_act (Activatio	(None, 7, 7, 1536)	0	block14_sepconv1_bn[0][0]
block14_sepconv2 (SeparableConv	(None, 7, 7, 2048)	3159552	block14_sepconv1_act[0][0]
block14_sepconv2_bn (BatchNorma	(None, 7, 7, 2048)	8192	block14_sepconv2[0][0]
block14_sepconv2_act (Activatio	(None, 7, 7, 2048)	0	block14_sepconv2_bn[0][0]
flatten_3 (Flatten)	(None, 100352)	0	block14_sepconv2_act[0][0]
dense_3 (Dense)	(None, 2)	200706	flatten_3[0][0]

Total params: 21,062,186
Trainable params: 200,706
Non-trainable params: 20,861,480

Figure II.13. Summary of the Xception model.

II.5.5. MobileNetV2 Model

MobileNetV2 is a convolutional neural network architecture designed to perform well on mobile devices. It is based on an inverted residual structure, where residual connections are located between bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as non-linear sources. Overall, the MobileNetV2 architecture consists of an initial fully convolutional layer with 32 filters, followed by 19 remaining bottleneck layers [Sandler, 2018]. Figures II.14 and II.15 show the MobileNetV2's architecture and its summary, respectively.

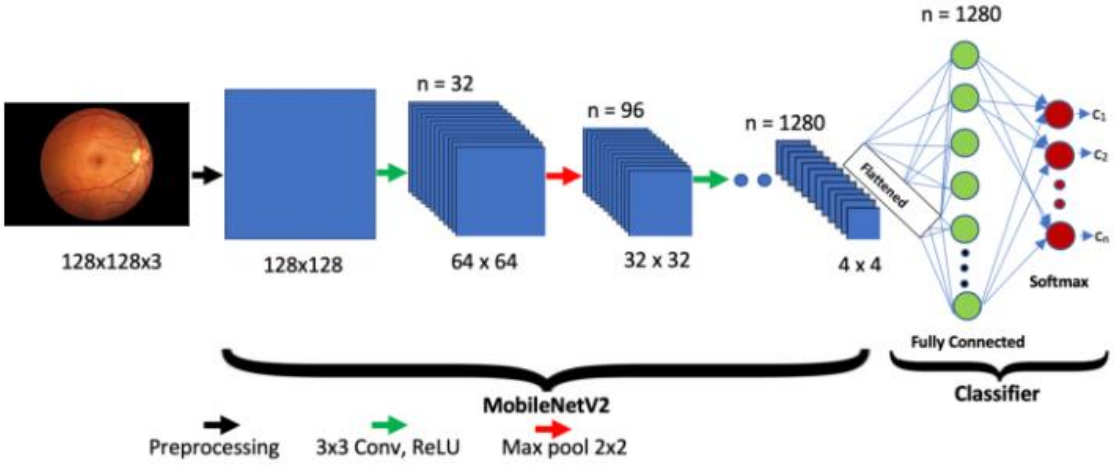


Figure II.14. MobileNetV2 architecture.

```
_________________________________________________________________
conv_pw_12_bn (BatchNormaliz (None, 7, 7, 1024)        4096
_________________________________________________________________
conv_pw_12_relu (ReLU)       (None, 7, 7, 1024)        0
_________________________________________________________________
conv_dw_13 (DepthwiseConv2D) (None, 7, 7, 1024)        9216
_________________________________________________________________
conv_dw_13_bn (BatchNormaliz (None, 7, 7, 1024)        4096
_________________________________________________________________
conv_dw_13_relu (ReLU)       (None, 7, 7, 1024)        0
_________________________________________________________________
conv_pw_13 (Conv2D)          (None, 7, 7, 1024)        1048576
_________________________________________________________________
conv_pw_13_bn (BatchNormaliz (None, 7, 7, 1024)        4096
_________________________________________________________________
conv_pw_13_relu (ReLU)       (None, 7, 7, 1024)        0
_________________________________________________________________
flatten_4 (Flatten)          (None, 50176)             0
_________________________________________________________________
dense_4 (Dense)              (None, 2)                 100354
=================================================================
Total params: 3,329,218
Trainable params: 100,354
Non-trainable params: 3,228,864
_________________________________________________________________
```

Figure II.15. MobileNetV2 summary.

Now, we make a comparison between these CNN models in term of size, Top-1, Top-5, number of parameters, depth and time-consuming using the well-known collection ImageNet. Table II.3 summarizes these statistics.

Table II.3. Comparison of the five different models on ImageNet collection.

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	# of Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
VGG16	528	71.3%	90.1%	138.4M	16	69.5	4.2
VGG19	549	71.3%	90.0%	143.7M	19	84.8	4.4
InceptionV3	92	77.9%	93.7%	23.9M	189	42.2	6.9
Xception	88	79.0%	94.5%	22.9M	81	109.4	8.1
MobileNetV2	14	71.3%	90.1%	3.5M	105	25.9	3.8

II.6. CLASSIFICATION

II.6.1. Basic Architecture

The CNN architecture consists of two main parts:

- A convolution tool is used to separate and identify various features of an image for analysis in a process called feature extraction.
- A fully connected layer that uses the output of the convolution process and predicts the class of the image based on features extracted at earlier stages [Gurucharan, 2020].

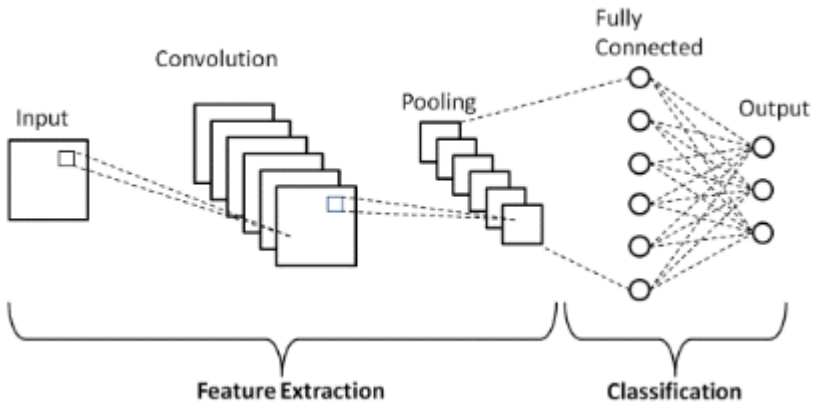


Figure II.16. CNN base architecture.

II.6.2. Classification with Softmax

Finally, one of the most important parameters of a CNN model is the activation function. They are used to learn and approximate any type of continuous and complex relationship between network variables. In short, it decides which model information should be transmitted in the forward direction and which should not be transmitted at the end of the network.

The activation function increases the nonlinearity of the network. There are several commonly used activation functions such as ReLU, Softmax, Tank and Sigmoid functions. Each of these functions has a specific purpose. For CNN models with binary classification, sigmoid and softmax functions are preferred, for multiclass classification, Softmax is usually used [Gurucharan, 2020].

The only necessary change to the network is to adjust the last layer. Our problem has 2 classes and the initial training was done with thousands of classes. It is therefore necessary to change the last layer of the model and replace it with a layer of two neurons using the Softmax classifier for multi classification [Jmour, 2018]. In this way, the convolutional neural network will be better adapted to our classification task.

II.7. CONCLUSION

In this chapter, we discussed the literature of diabetic retinopathy diagnosis systems. Also, we described the well-known APTOS2019 dataset and explained five CNN models with theirs properties and theirs structures. In the next chapter, we will report the evaluation and the results of our experiments with quantitative and qualitative discussions to highlight the strengths and weaknesses of our system.

CHAPTER III: EVALUATION & RESULTS

III.1. INTRODUCTION

In this chapter, we evaluate and discuss many different experiments conducted on APTOS 2019 dataset using five transfer learning models (VGG16, VGG19, InceptionV3, Xception and MobileNetV2). Furthermore, the preprocessing task and data augmentation are performed in order to enhance the system accuracy. The obtained results are expressed in term of accuracy, precision, recall and f1-score.

III.2. DATASET DESCRIPTION

In our work, we use the well-known APTOS2019 dataset, which contain 3662DR images. However, we should notice that we created two new dataset inspired from the original APTOS2019. The first new one consists of applying a Gaussian filter on the entire DR images. The second new one is the gray level transformation of the original one. Each dataset is divided as follows: 80% of samples for training, 10% for validation and 10% for testing. We recall also that all images from these three collections have the same size (224×224).

III.2.1. First level

As first level of our work, we make a binary classification for patients with and without retinopathy. For this purpose, we take the whole dataset and then divide it into two classes:

- **Class 0:** No DR
- **Class 1:** DR (that contains Moderate, Severe, Mild and Proliferative DR).

The following table summary the number of images in the two classes:

Table III.1. Summary of number of images for the first level.

	State	# of samples
class 0	No DR	1805
class 1	With DR	1857

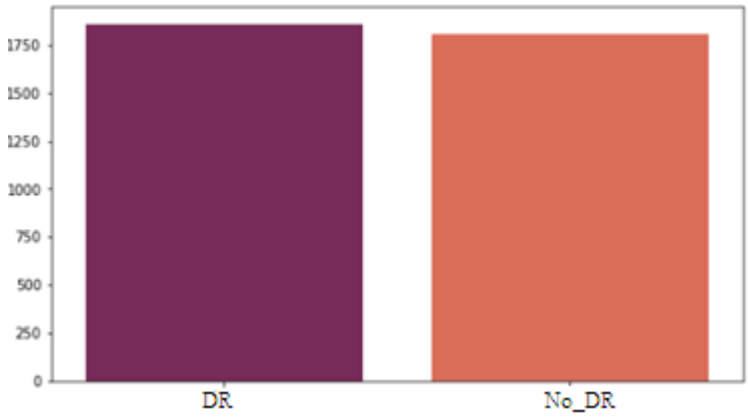


Figure III.1. Repartition of DR/No_DR images in each dataset.

III.2.2. Second Level:

In this second level, we collect the first and second categories of disease and classify them as one category, and we do the same with the remaining two categories

- **Class 1:** mild DR and moderate DR
- **Class 2:** severe DR and Proliferative DR

The next table summarizes the number of images in each class.

Table III.2. The number of images in the first sub-level.

	# of images
Class 1(Mild)	1369
Class 2 (Severe)	665

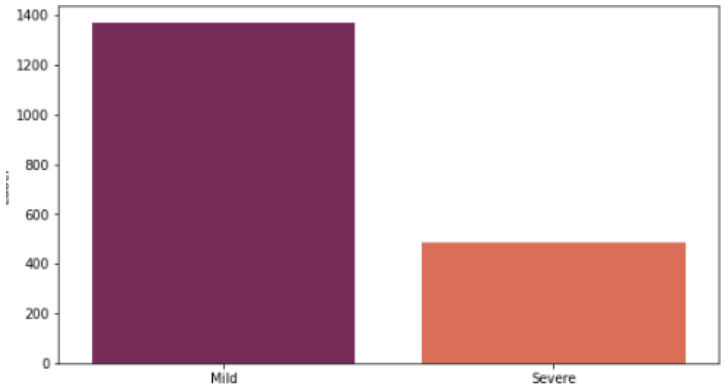
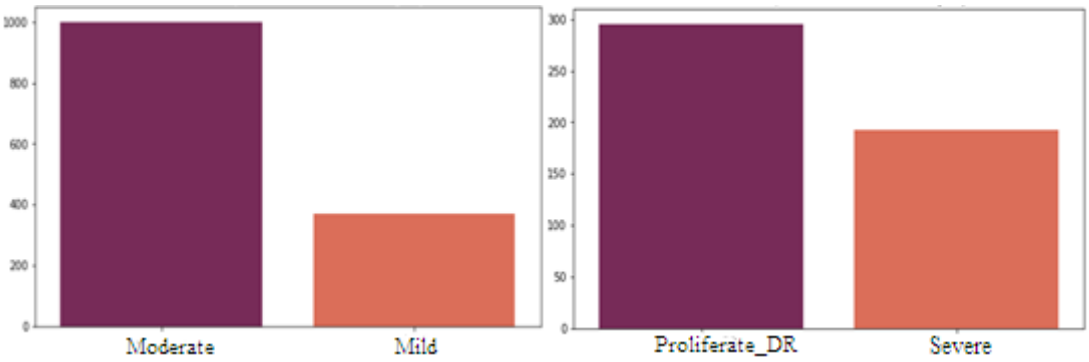


Figure III.2. The number of samples of Mild/Moderate and Severe/ Proliferative.

III.2.3. Third level:

In this binary classification level, we identify the degree of DR disease by considering all DR categories. Hence, we distinguish between the first and second category of disease, as well as between the third and the fourth, respectively.



a) b)

Figure III.3. The number of samples of each category: a) In Mild/Moderate and b) In Severe/Proliferate_DR.

III.3. EVALUATION METRICS

Based on the system we develop, we consider some metrics that respect the constraints associated with the binary classification. Using the matrix confusion, we opted for following measurements: precision, f1_score, recall and the overall accuracy.

III.3.1. Confusion matrix

The confusion matrix measures the classification accuracy of the classifier. Each column is a predicted class, and each row is the true data class. The diagonal lines indicate that their predictions matched the ground truth, the more accurate the predictions, the darker the color [San-Li, 2021].

Table III.3. Confusion matrix and its components.

Predicted \ **Actual**	**No DR**	**DR**
No DR	TN	FN
DR	FP	TP

Where:

TP = True Positive (denotes an incorrect classification of a normal subject).

TN = True Negative (means a disease subject is incorrectly detected).

FP = False Positive (denotes an incorrect classification of a normal subject as a disease subject).

FN = False Negative (means a disease subject is incorrectly detected as a normal subject).

The different metrics inspired from the confusion matrix are defined as following:

a) ***Recall:*** Recall is given by:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FN})} \qquad (\mathbf{2})$$

b) ***Precision:*** Precision is given by:

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP} + \boldsymbol{FP})} \qquad (\mathbf{3})$$

c) f1-score:

$$\boldsymbol{f1_{score}} = \frac{\boldsymbol{2TP}}{\boldsymbol{2TP} + \boldsymbol{FP} + \boldsymbol{FN}} \qquad (\mathbf{4})$$

III.3.2. Accuracy

The accuracy is defined as the correctly classified images to the total number of samples. It is given in the following equation [Shanthamalar, 2022].

$$\mathbf{Accuracy} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{TN} + \mathbf{FP} + \mathbf{FN}} \qquad (\mathbf{4})$$

A higher accuracy value means the correct classification of the samples.

III.4. RESULTS AND DISCUSSIONS

III.4.1. First level evaluation

III.4.1.1. Influence of preprocessing on the performance

After reviewing many research papers by researchers in recent years, we noticed that the Majority of them have adopted the values of these variables in order to increase the volume of Data they have. In our work, we used them as constants so that:

Rotation range = 30

Zoom range = 0.15

Width shift range = 0.2

Height shift range = 0.2

Shear range = 0.15

Horizontal flip = True.

In this experiment, we take the following model parameters: Learning Rate (L_R) = 0.01, Optimizer = Adam, Batch size = 16 and Number of epochs = 20. We recall here that a comparison of performance has been evaluated between the original data and the two new ones (Gaussian filter data and Gray level data). The following experiments show the obtained performance with different models applied on our dataset (with and without preprocessing).

c) The Original APTOS2019 dataset (without preprocessing)

Table III.4. Performance of our five models with the original APTOS2019 in first level.

Model	APTOS2019 Original						
	ACC (%)	Macro avg			Weighted avg		
		Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
VGG16	95,5	96	95	95	96	95	95
VGG19	97,25	97	97	97	97	97	97
Inception	96,5	97	96	96	97	96	96
Xception	94	95	94	94	95	95	94
MobileNetV2	90,19	91	90	90	91	90	90

d) With preprocessing

❖ The Gaussian filtered data

Table III.5. Performance of our five models with the Gaussian APTOS2019 in first level.

Model	Gaussian filtered data						
	ACC (%)	Macro avg			Weighted avg		
		Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
VGG16	94,25	94	94	94	94	94	94
VGG19	94,25	95	94	94	95	94	94
Inception	94,5	95	94	94	95	94	94
Xception	95,75	95	94	94	95	95	94
MobileNetV2	94,01	94	94	94	94	94	94

❖ The grayscale

Table III.6. Performance of our five models with the Grayscale APTOS2019 in first level.

Model	Grayscale						
	ACC (%)	Macro avg			Weighted _avg		
		Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
VGG16	94,5	95	94	94	95	95	94
VGG19	94	94	94	94	94	94	94
Inception	95,25	95	95	95	95	95	95
Xception	93	95	94	94	95	95	94
MobileNetV2	94,55	95	95	95	95	95	95

From Tables III.4, III.5 and III.6, we notice that the results were so near between the three dataset. However, the performance obtained with the original images is slightly higher than other two datasets (with preprocessing). For this reason, in the next experiments, we consider only the original APTOS2019 for our DR classification. Now, we change the parameters of each pre-trained model to get the optimal performance for each one.

III.4.1.2. The influence of VGG16 parameters on the system performance

We study in this experiment the influence of the optimizer type and the learning rate value on the VGG16 performance. Table III.7 reports the obtained results with different values of each VGG16 parameters.

From Table III.7, we notice that the optimizer RMSprop produced the highest performance with a L_R eqaul 0.01. We keep this optimizer and its value for the second experiment then we change the batch size and the epoch number. Table III.8 reports the obtained results with different values of batch size and epoch number.

As results, we obtained the best accuracy using the VGG16 model with the original APTOS2019 dataset with the following parameters: optimizer RMSprop, L_R = 0.01, batch size = 64, and number of epoch = 40.

Table III.7. The influence of the optimizer and the learning rate on the first level performance using VGG16 model.

			Macro avg			Weighted avg		
Optimizer	L_R	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
Adam	0,01	95,5	96	95	95	96	95	95
	0,001	95,5	95,75	95,5	95,5	95,75	95,5	95,5
	0,0001	94,25	94,28	94,25	94,25	94,28	94,28	94,25
	0,00001	92,5	92,52	92,5	92,5	92,52	92,5	92,5
SGD	0,01	94	94,28	94	93,99	94,28	94,25	93,99
	0,001	94,25	94,28	94,25	94,25	94,28	94,25	94,25
	0,0001	93	93,02	93	93	93,02	93	93
	0,00001	86,25	86,27	86,25	86,25	86,27	86,25	86,25
RMSprop	0,01	96	96	96	96	96	96	96
	0,001	92,75	93,32	92,75	92,73	93,32	92,75	92,73
	0,0001	95	95	95	95	95	95	95
	0,00001	93,75	93,76	93,75	93,75	93,76	93,75	93,76

Table III.8. The influence of the batch size and the number of epochs on the first level performance using VGG16 model.

			Macro avg			Weighted avg		
batch_size	#of epoch	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
16	20	96	96	96	96	96	96	96
	30	95,25	95,31	95,25	95,25	95,3	95,25	95,25
	40	93,5	93,94	93,5	93,48	93,54	93,5	93,48

	50	94,5	94,54	94,5	94,5	94,54	94,5	94,5
32	20	95,75	95,75	95,75	95,75	95,75	95,75	95,75
	30	88,5	90,1	88,5	88,38	90,1	88,5	88,38
	40	90,5	91,56	90,5	90,44	91,56	90,5	90,44
	50	95,5	95,5	95,5	95,5	95,5	95,5	95,5
64	20	94,5	94,5	94,25	94,24	94,5	94,25	94,5
	30	93,5	93,86	93,86	93,49	93,86	93,86	93,49
	40	96	96	96	96	96	96	96
	50	93	93,53	93	92,98	93,53	93	92,98

Figure III.4 shows the confusion matrix obtained using the best VGG16 parameters.

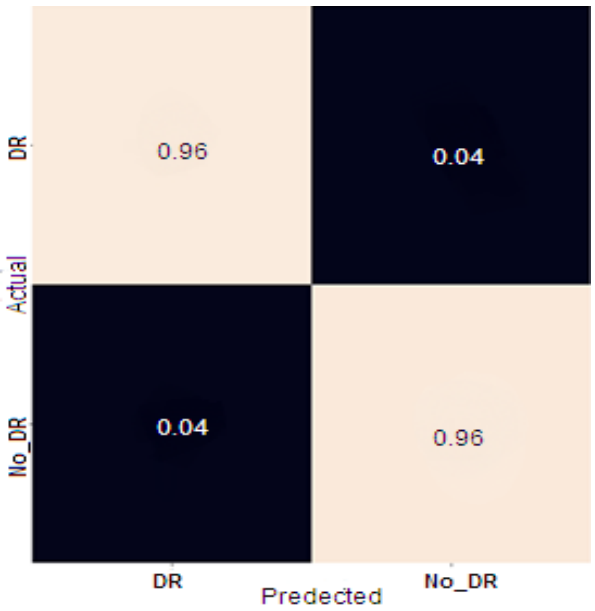


Figure III.4. Confusion matrix of the best result for the VGG16 model.

III.4.1.3. The influence of VGG19 parameters on the system performance

We study in this experiment the influence of the optimizer type and the learning rate value on the VGG19 performance. Table III.9 reports the obtained results with different values of each VGG19 parameters.

From Table III.9, we notice that the optimizer RMSprop produced the highest accuracy with a L_R = 0.001. We keep this optimizer and its value for the second experiment then we change the batch size and the epoch number. Table III.10 reports the obtained results with different values of batch size and epoch number.

As results, we obtained the best accuracy using the VGG16 model with the original APTOS2019 dataset with the following parameters: optimizer RMSprop, L_R = 0.001, batch size = 16 and number of epoch = 20.

Table III.9. The influence of the optimizer and the learning rate on the first level performance using VGG19 model.

			Macro avg			Weighted avg		
Optimizer	L_R	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
Adam	0,01	97	97	97	97	97	97	97
	0,001	97	97	97	97	97	97	97
	0,0001	96,75	96,75	96,75	96,75	96,75	96,75	96,75
	0,00001	94,25	94,28	94,25	94,25	94,28	94,25	94,25
SGD	0,01	94,25	94,5	94,25	94,24	94,5	94,25	94,24
	0,001	96,75	96,75	96,75	96,75	96,75	96,75	96,75
	0,0001	94,5	94,52	94,5	94,5	94,52	94,5	94,5
	0,00001	89,5	89,52	89,5	89,5	89,52	89,5	89,5
RMSprop	0,01	95,75	95,84	95,75	95,75	95,84	95,75	95,75
	0,001	97,25	97,25	97,25	97,25	97,25	97,25	97,25
	0,0001	96,75	96,75	96,75	96,75	96,75	96,75	96,75
	0,00001	94,75	94,75	94,75	94,75	94,75	94,75	94,75

Table III.10. The influence of the batch size and the number of epochs on the first level performance using VGG19 model.

			Macro avg			Weighted avg		
batch_size	# of epoch	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
16	20	97,25	97,25	97,25	97,25	97,25	97,25	97,25
	30	96,25	96,31	96,25	96,25	96,31	96,25	96,25
	40	96	96,07	96	96	96,07	96	96
	50	96	96,04	96	96	96,04	96	96
32	20	96	96,07	96	96	96,07	96	96
	30	96	96,07	96	96	96,07	96	96
	40	97,00	97	97	97	97	97	97
	50	97,5	97,5	97,5	97,5	97,5	97,5	97,5
64	20	96,5	96,5	96,5	96,5	96,5	96,5	96,5
	30	96,5	96,52	96,5	96,5	96,52	96,5	96,5
	40	97	97,02	97	97	97,02	97	97
	50	96,75	96,76	96,75	96,75	96,76	96,75	96,75

Figure III.5 demonstrates the confusion matrix obtained from the best VGG19 parameters.

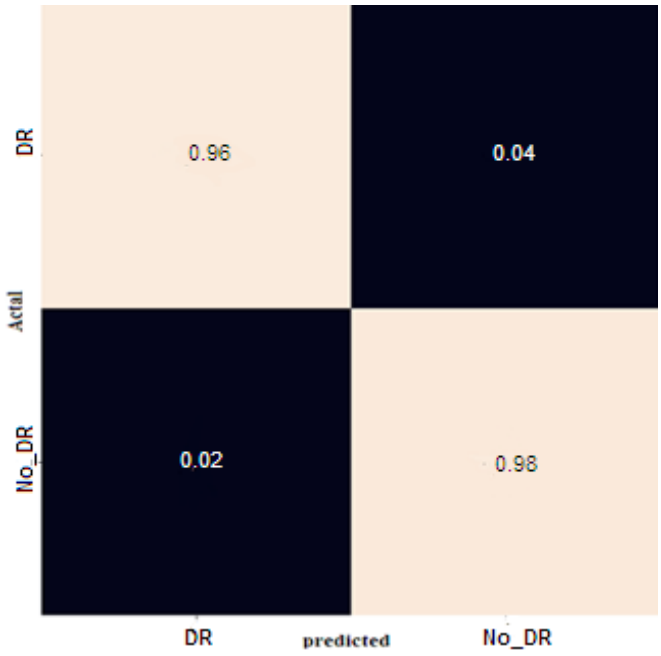


Figure III.5. Confusion matrix of the best result for the VGG19 model.

III.4.1.4. The influence of InceptionV3 parameters on the system performance

In this experiment, we study the influence of the optimizer and its learning rate value on the inceptionV3 performance. Table III.11 reports the obtained results with different values of each inceptionV3 parameters.

From Table III.11, we clearly see that the optimizer Adam produced the highest accuracy with a L_R = 0.001. We keep this optimizer and its value for the second experiment then we change the batch size and the epoch number. Table III.12 reports the obtained results with different values of batch size and epoch number.

As results, we obtained the best accuracy using the inceptionV3 model with the original APTOS2019 dataset with the following parameters: optimizer Adam, L_R = 0.001, batch size = 32 and number of epoch = 50.

Table III.11. The influence of the optimizer and the learning rate on the first level performance using InceptionV3 model.

			Macro avg			Weighted avg		
Optimizer	L_R	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
Adam	0,01	96,5	97	95	95	97	96	96
	0,001	97	97	97	97	97	97	97
	0,0001	96	96	96	96	96	96	96
	0,00001	96	96	96	96	96	96	96
SGD	0,01	96	96	96	96	96	96	96
	0,001	95,25	95,25	95,25	95,25	95,25	95,25	95,25
	0,0001	81,5	83,43	81,5	81,23	83,43	81,5	81,23
	0,00001	46,25	41,57	46,25	37,59	41,57	46,25	37,59
RMSprop	0,01	96,25	96,25	96,25	96,25	96,25	96,25	95
	0,001	96,75	96,76	96,75	96,75	96,76	96,75	96,75
	0,0001	95,25	95,44	95,25	95,24	95,44	95,25	95,25
	0,00001	96,25	96,25	96,25	96,25	96,25	96,25	96,25

Table III.12. The influence of the batch size and the number of epochs on the first level performance using InceptionV3 model.

			Macro avg			Weighted avg		
batch_size	# of epoch	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
16	20	97	97	97	97	97	97	97
	30	96,75	96,76	96,75	96,75	96,76	96,75	96,75
	40	96,75	96,76	96,75	96,75	96,76	96,75	96,75
	50	97,25	97,25	97,25	97,25	97,25	97,25	97,25
32	20	97,25	97,25	97,25	97,25	97,25	97,25	97,25
	30	96	96	96	96	96	96	96
	40	97	97	97	97	97	97	97
	50	97,5	97,52	97,5	97,5	97,52	97,5	97,5
64	20	96,25	96,26	96,25	96,25	96,26	96,25	96,25
	30	96,75	96,76	96,75	96,75	96,76	96,75	96,75
	40	96,75	96,76	96,75	96,75	96,76	96,75	96,75
	50	93,75	94,34	93,75	93,73	94,34	93,75	93,73

Figure III.6 shows the confusion matrix obtained from the best InceptionV3 parameters.

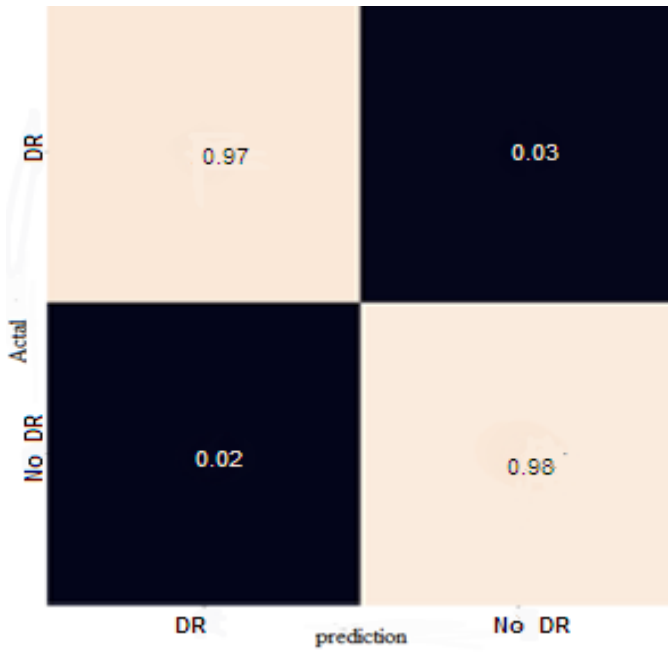


Figure III.6. Confusion matrix of the best result for the InceptionV3 model.

III.4.1.5. The influence of Xception parameters on the system performance

Now, we study the influence of the optimizer and its learning rate value on the Xception performance. Table III.13 reports the obtained results with different values of each Xception parameters.

From Table III.13, we clearly see that the optimizer Adam produced the highest accuracy with a L_R = 0.00001. We keep this optimizer and its value for the second experiment then we change the batch size and the epoch number. Table III.14 reports the obtained results with different values of batch size and epoch number.

As results, we obtained the best accuracy using the Xception model with the original APTOS2019 dataset with the following parameters: optimizer Adam, L_R = 0.00001, batch size = 16 and number of epoch = 20.

Table III.13. The influence of the optimizer and the learning rate on the first level performance using Xception model.

			Macro avg			Weighted avg		
Optimizer	L_R	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
Adam	0,01	95,5	93,53	93	92,98	93,53	93	92,98
	0,001	96	93,53	93	92,98	93,53	93	92,98
	0,0001	95,5	93,53	93	92,98	93,53	93	92,98
	0,00001	96,75	96,78	96,75	96,75	96,78	96,75	96,75
SGD	0,01	95,5	95,66	95,5	95,5	95,66	95,5	95,5
	0,001	88,75	90,28	88,75	88,64	90,28	88,75	88,64
	0,0001	61,25	74,47	61,25	55,2	74,47	61,25	55,2
	0,00001	53,5	65,51	53,5	42,34	65,51	53,5	42,34
RMSprop	0,01	92,25	92,92	92,25	92,25	92,92	92,25	92,25
	0,001	94,75	95	94,75	94,74	95	94,75	94,74
	0,0001	96	96	96	96	96	96	96
	0,00001	96	96,07	96	96	96,07	96	96

Table III.14. The influence of the batch size and the number of epochs on the first level performance using Xception model.

			Macro avg			Weighted avg		
batch_size	# of epoch	Val_ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
16	20	96,75	96,78	96,75	96,75	96,78	96,75	96,75
	30	96,25	96,31	96,25	96,25	96,31	96,25	96,25
	40	95,75	95,84	95,75	95,75	95,84	95,75	95,75
	50	96	96,12	96	96	96,12	96	96
32	20	95,25	95,34	95,25	95,25	95,34	95,25	95,25
	30	96	96,04	96	96	96,04	96	96
	40	95,75	95,84	95,75	95,75	95,84	95,75	95,75
	50	96,5	96,52	96,5	96,5	96,52	96,5	96,5
64	20	94,25	94,3	94,25	94,25	94,3	94,25	94,25
	30	94,75	94,84	94,75	94,75	94,84	94,75	94,75
	40	94,75	95,81	95,75	95,75	95,81	95,75	95,75
	50	95,75	95,81	95,75	95,75	95,81	95,75	95,75

Figure III.7 shows the confusion matrix obtained from the best Xception parameters.

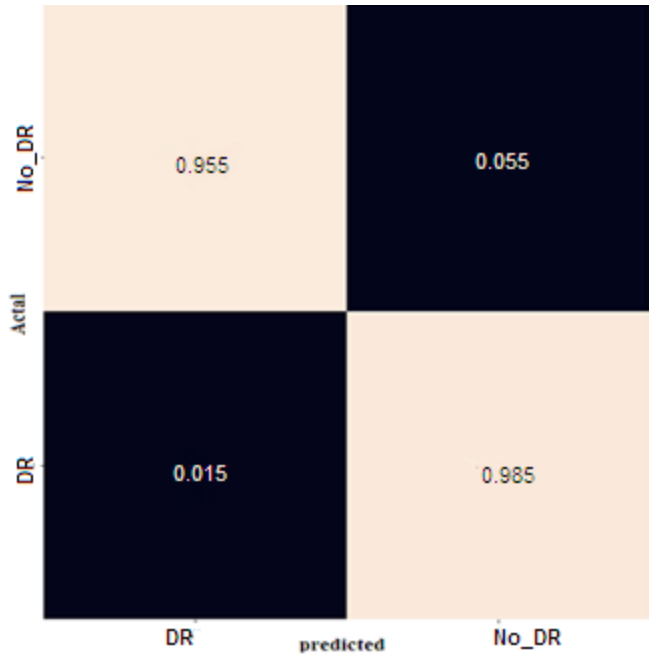


Figure III.7. Confusion matrix of the best result for the Xception model.

III.4.1.6. The influence of MobileNetV2 parameters on the system performance

Now, we study the influence of the optimizer and its learning rate value on the MobileNetV2 performance. Table III.15 reports the obtained results with different values of each MobileNetV2 parameters.

From Table III.15, we can clearly see that the RMSprop optimizer produces the highest percentage with a L_R = 0.001. We keep this optimizer and its value for the second experiment then we change the batch size and the epoch number. Table III.16 reports the obtained results with different values of batch size and epoch number.

As results, we obtained the best accuracy using the MobileNetV2 model with the original APTOS2019 dataset with the following parameters:

optimizer RMSprop, L_R = 0.001, batch size = 64 and number of epoch = 50.

Table III.15. The influence of the optimizer and the learning rate on the first level performance using MobileNetV2 model.

			Macro avg			Weighted avg		
optimizer	L_r	ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
Adam	0,01	90,19	91	90	90	91	90	90
	0,001	97,55	98	98	98	98	98	98
	0,0001	90,19	90	90	90	90	90	90
	0,00001	97,28	97	97	97	97	97	97
SGD	0,01	96,73	97	97	97	97	97	97
	0,001	96,46	96	96	96	96	96	96
	0,0001	86,1	86	86	86	86	86	86
	0,00001	71,66	73	72	71	73	72	71
RMSprop	0,01	88,56	91	89	88	91	88	88
	0,001	98,09	98	98	98	98	98	98
	0,0001	92,37	92	92	92	92	92	92
	0,00001	89,92	90	90	90	90	90	90

Table III.16. The influence of the batch size and the number of epochs on the first level performance using MobileNetV2 model.

			Macro avg			Weighted avg		
batch_size	# of epoch	ACC (%)	Precision (%)	Recall (%)	f1_score (%)	Precision (%)	Recall (%)	f1_score (%)
16	20	98,09	98	98	98	98	98	98
	30	93,73	94	94	94	94	94	94
	40	95,1	95	95	95	96	95	95
	50	98,09	98	98	98	98	98	98
32	20	91,01	91	91	91	91	91	91
	30	92,1	92	92	92	92	92	92
	40	89,92	92	90	90	92	90	90
	50	91,28	92	91	91	92	91	91
64	20	95,1	95	95	95	96	95	95
	30	97	97	97	97	97	97	97
	40	98,09	98	98	98	98	98	98
	50	98,37	98	98	98	98	98	98

Figure III.8 illustrates the confusion matrix obtained from the best MobileNetV2 parameters using the original APTOS2019 dataset.

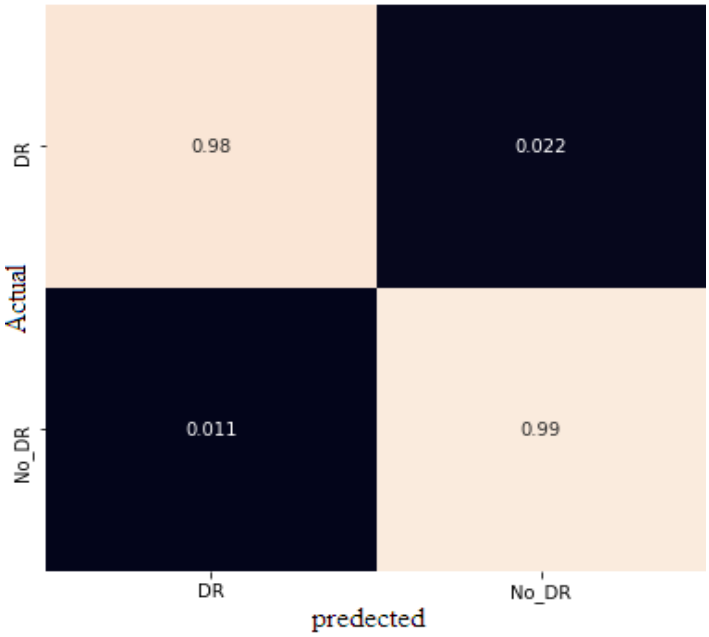


Figure III.8. Confusion matrix of the best result for the MobileNetV2 model.

III.4.2. Second level evaluation

Based on several experiments evaluated on this level, the obtained results, Table III.17 reports the best accuracy obtained with the optimal parameters of the five models. We recall here that we combined the two classes (Mild DR & Moderate DR) in one class and (Severe DR & Proliferative DR) in other class.

Table III.17. The obtained accuracies for second level using five models.

Models	Opimizers	L_R	Batch_size	# of epoch	ACC
VGG19	RMS	0,001	64	20	80,95
VGG16	Adam	0,01	16	20	76,34
InceptionV3	SGD	0,01	64	20	80,11
Xception	Adam	0,0001	32	20	79,03
MobileNetV2	**Adam**	**0,01**	**64**	**20**	**83,94**

From table III.17, the MobileNetV2 has the best result in term of accuracy, with the following parameters: Optimizer = Adam, learning rate = 0.01, batch size = 64 and number of epoch = 20. Figure III.9 illustrates the confusion matrix obtained from the best MobileNetV2 parameters using the binary classification of second level.

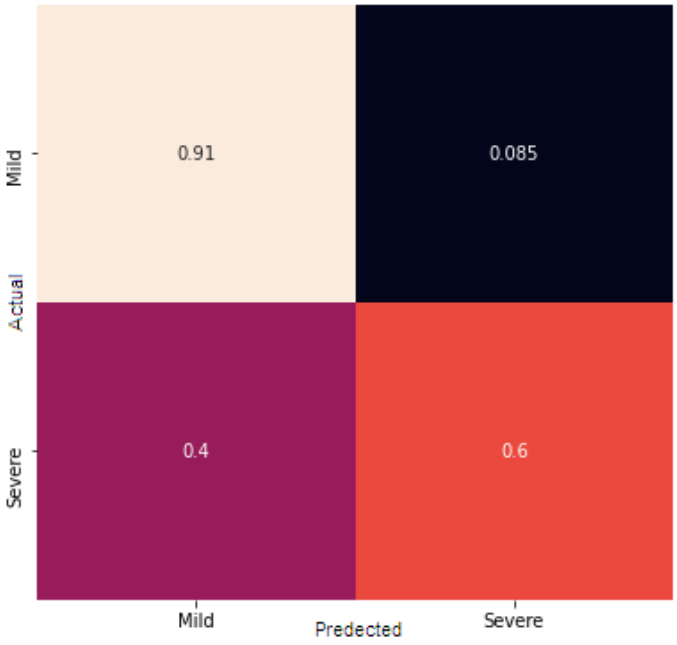


Figure III.9. Confusion matrix of the best model (MobileNetV2) using the second level.

III.4.3. Third level evaluation

a) Mild & Moderate:

At this level, we also evaluated many experiments in order to get the best accuracy for each model. Table III.18 summaries the best accuracy obtained with the updated parameters of the five models between class 1 (Mild DR) and class 2 (Moderate DR).

Table III.18. The obtained accuracies for Mild/Moderate classification using five models.

Models	Opimizers	L_R	Batch_size	# of epoch	ACC (%)
VGG19	RMS	0,001	32	40	80,29
VGG16	Adam	0,0001	32	20	82,48
InceptionV3	Adam	0,0001	16	20	79,57
Xception	RMS	0,001	32	40	83,21
MobileNetV2	**Adam**	**0,001**	**64**	**40**	**86,13**

From Table III.18, MobileNetV2 outperformed all models in terms of accuracy for classifying DR images into mild or moderate. Figure III.11 illustrates the confusion matrix obtained from the best MobileNetV2 parameters using the binary classification of Mild/Moderate classes.

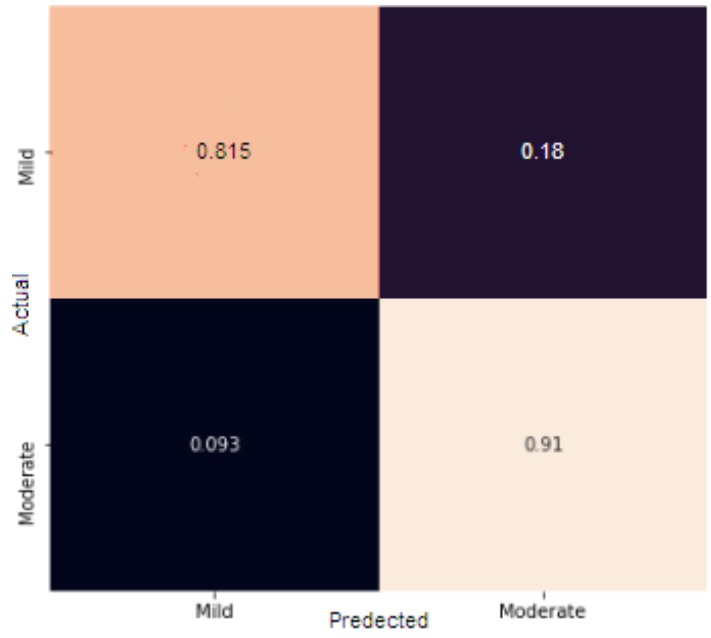


Figure III.10. Confusion matrix of the best model (MobileNetV2) in Mild/Moderate classification.

b) Severe & proliferative :

The same way in this case, we tried to classify the DR samples that belong into these two classes: Severe and Proliferative DR. Table III.19 summaries the best accuracy obtained with the five models.

Table III.19. The obtained accuracies for Severe/proliferative classification using five models.

Models	Optimizer	L_R	Batch_size	# of epoch	ACC (%)
VGG19	Adam	0,001	32	20	63,27
VGG16	Adam	0,001	32	40	65,32
InceptionV3	RMS	0,0001	64	40	65,31
Xception	RMS	0,0001	32	40	73,47
MobileNetV2	**RMS**	**0,001**	**64**	**25**	**79,59**

From table III.19, the MobileNetV2 has the best result in term of accuracy when we are staging Severe and Proliferate_DR, despite the lower classification performance. This is mainly due to their resemblance. Figure III.11 illustrates the confusion matrix obtained from the best MobileNetV2 parameters using the binary classification of Severe/Proliferate classes.

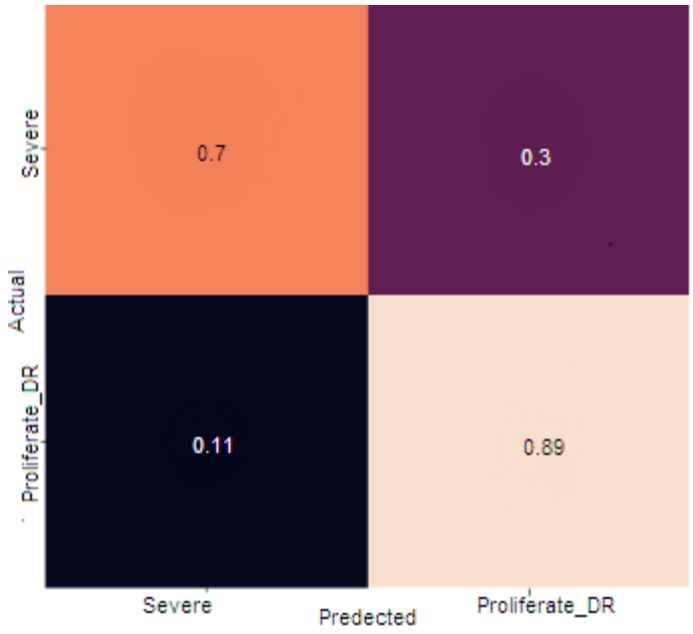


Figure III.11. Confusion matrix of the best result between Severe and Proliferate_DR.

III.5. COMPARATIVE STUDY

In this section, we make a meaningful comparison with works that used the same dataset for Diabetic retinopathy diagnosis. Table III.20 reports the comparison statistics.

Table III.20. Comparison with state-of-the-art methods.

Authors	year	Method	# of classes	# of data	Performance (%)			
					Acc.	Prec.	Recall	F1-score
Bodapati	2020	CNN with Xception	2	3662	97.41	-	-	-
Firkre	2021	CNN proposed model	2	4737	96.15	89	79	84.1
[Islam, 2022]	2022	CNN with proposed method	2	3662	98.36	98.36	98.37	-
Our work with DR identification	**2022**	**MobileNetV2**	**2**	**3662**	**98.37**	**98**	**98.89**	**98.34**

From Table III.20, we notice that our system performs very well with the MobileNetV2 model compared to other models. The performance of our proposed system presented a high performance in term of accuracy.

In this section, and in order to simulate the medical protocol evaluated in diagnosing DR, we propose a global accuracy rule reported in Eq. III-5. This mathematical rule describes the motivational diagnosis used in health care, as $\boldsymbol{\alpha}$ represents the diagnostic importance factor at each stage, which we express as follows.

$$\mathbf{Accg} = (\mathbf{Acc1} * \boldsymbol{\alpha 1}) + (\mathbf{Acc2} * \boldsymbol{\alpha 2}) + (\mathbf{Acc3} * \boldsymbol{\alpha 3}) + (\mathbf{Acc4} * \boldsymbol{\alpha 4}) \qquad (\mathbf{5})$$

Where:

Accg: Accuracy global of our system.

Acc1: Accuracy of the level 1 (No DR and With DR).

Acc2: Accuracy of the level 2 (Mild and Severe).

Acc3: Accuracy of the level 3 (Mild and Moderate).

Acc4: Accuracy of the level 3 (Severe and Proliferated).

$\alpha 1$: The importance of first decision (DR _No DR).

$\alpha 2$: The importance of second decision (Mild _ Severe).

$\alpha 3$: The importance of thirsty decision (Mild _ Moderate).

$\alpha 4$: The importance of thirsty decision (Severe _ Proliferated).

Remember here that in medicine, the first level of decision is more important to both the healthcare professional and the patient than the decisions of the other level. Each has a specific area as follows:

α1 ϵ [0.7 ; 0.5]

α2 ϵ [0.3 ; 0.1]

α3 ϵ [0.15 ; 0.1]

α4 ϵ [0.15 ; 0.1]

This table represents results for some values that can be taken byα_i.

Table III.21. The calculated global accuracy in function of α.

	Case 1	Case 2	Case 3	Case 4
α1	0,7	0,6	0,5	0,5
α2	0,1	0,2	0,3	0,2
α3	0,1	0,1	0,1	0,15
α4	0,1	0,1	0,1	0,15
Accg (%)	93,82	92,38	90,93	90,83

From Table III.21, we can notice that our system exceeds the 90% in term of accuracy whatever the values of α_i. The obtained results can make our system operational in healthcare centers with a very acceptable accuracy confidence.

III.6. CONCLUSION

The main goal of this chapter was to design a powerful system that, on one hand, identifies DR from the No_DR cases. On other hand, it identifies the severity of DR. After finding the suitable parameters of each model, we evaluated several experiments with three levels in order to simulate the real medical diagnosis.

The obtained results showed the superiority of MobileNetV2 against other models by giving a high accuracy (98.37%), which outperformed the state-of-the-art techniques.

GENERAL CONCLUSION

In this work, we have proposed an automated system for diabetic retinopathy classification using deep learning techniques. We have discussed the advantages of early diabetic retinopathy detection using the eye fundus images. The late detection of diabetic retinopathy caused by using manual protocol leads to permanent blindness. However, the automated diagnosis systems have piqued interest to get an early diagnosis in DR disease using Deep Neural Network models.

For this purpose, we have proposed a solution for overcoming the drawback of traditional diagnosis systems by using five different models including VGG16, VGG19, Xception, InceptionV3 and MobileNetV2 to extract representative features from DR images.

As dataset, we used the well-known APTOS2019 collection in three variants (the original one, the Gaussian filtered one and the gray level variant). In addition, we have used many evaluation metrics such Accuracy, Precision, Recall, and F1 score in order to assess the performance of our proposed system. This statistics enable us to analyze our models' outcomes in order to improve detection performance.

As experimental protocol, we have used hierarchical decisions that exactly emulate the medical protocol applied by healthcare professionals. In fact, the medical protocol for DR prediction needs firstly to check whether a person has DR or not. If the patient has DR, the medical conduct is then oriented towards the severity of DR. At this level, we have divide the degree of severity into two classes and each one is also divided into two subclasses in hierarchical way.

For the testing phase, we conducted series of experiments on different models with different parameters for each category. The obtained results have presented very satisfactory performances with MobileNetV2 model applied on the original data.

The future perspectives that can be proposed to improve this system would be to incorporate one of these principles:

- Enhance the DR image representation by including CLAHE technique.
- Apply fine-tuning on the Deep learning models.

REFERENCES

[Adewole, 2020] Adewole, S., Yeghyayan, M., Hyatt, D., Ehsan, L., Jablonski, J., Copland, A., ... & Brown, D. (2020, November). Deep learning methods for anatomical landmark detection in video capsule endoscopy images. In *Proceedings of the Future Technologies Conference* (pp. 426-434). Springer, Cham.

[Anoop, 2022] Anoop, B. K. (2022). Binary Classification of DR-Diabetic Retinopathy using CNN with Fundus Colour Images. *Materials Today: Proceedings*.

[Bodapati, 2020] Bodapati, J. D., Naralasetti, V., Shareef, S. N., Hakak, S., Bilal, M., Maddikunta, P. K. R., & Jo, O. (2020). Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics*, *9*(6), 914.

[Bouletreau, 2019] Bouletreau, P., Makaremi, M., Ibrahim, B., Louvrier, A., & Sigaux, N. (2019). Artificial intelligence: applications in orthognathic surgery. *Journal of stomatology, oral and maxillofacial surgery*, *120*(4), 347-354.

[Firke, 2021] Firke, S. N., & Jain, R. B. (2021, March). Convolutional Neural Network for Diabetic Retinopathy Detection. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (pp. 549-553). IEEE.

[Guo, 2016] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27-48.

[Gurucharan, 2020] Gurucharan M K, «Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network », upGrad, DEC 07 2020, https://www.upgrad.com/blog/basic-cnn-architecture/

[Ibrahim, 2021] Ibrahim, I., & Abdulazeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10-19.

[Islam, 2020] Islam, M. R., Hasan, M. A. M., & Sayeed, A. (2020, June). Transfer learning based diabetic retinopathy detection with a novel preprocessed layer. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 888-891). IEEE.

[Jmour, 2018] Jmour, N., Zayen, S., & Abdelkrim, A. (2018, March). Convolutional neural networks for image classification. In *2018 international conference on advanced systems and electric technologies (IC_ASET)* (pp. 397-402). IEEE.

[Kajan, 2020] Kajan, S., Goga, J., Lacko, K., & Pavlovičová, J. (2020, January). Detection of diabetic retinopathy using pretrained deep neural networks. In *2020 Cybernetics & Informatics (K&I)* (pp. 1-5). IEEE.

[Kassani, 2019] Kassani, S. H., Kassani, P. H., Khazaeinezhad, R., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019, December). Diabetic retinopathy classification using a modified xception architecture. In *2019 IEEE international symposium on signal processing and information technology (ISSPIT)* (pp. 1-6). IEEE.

[kavitha, 2021] Kavitha, S., Dhanapriya, B., Vignesh, G. N., & Baskaran, K. R. (2021, October). Neural Style Transfer Using VGG19 and Alexnet. In *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1-6). IEEE.

[Kaul, 2020] Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal endoscopy*, *92*(4), 807-812.

[Khalifa, 2019] Khalifa, N., Loey, M., Taha, M., & Mohamed, H. (2019). Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection. *Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH*, *27*(5), 327–332. https://doi.org/10.5455/aim.2019.27.327-332

[Kim, 2020] Kim, S. H., Kim, J. H., & Im, E. G. (2020). Malware Family Classification Model using Xception Network.

[Manne, 2021] Manne, R., & Kantheti, S. C. (2021). Application of artificial intelligence in healthcare: chances and challenges. *Current Journal of Applied Science and Technology*, *40*(6), 78-89.

[Mintz, 2019] Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy & Allied Technologies*, *28*(2), 73-81.

[Mishra, 2020] Mishra, S., Hanchate, S., & Saquib, Z. (2020, October). Diabetic retinopathy detection using deep learning. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 515-520). IEEE.

[Mushtaq, 2022] Mushtaq, G., & Siddiqui, F. (2021, February). Detection of diabetic retinopathy using deep learning methodology. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1070, No. 1, p. 012049). IOP Publishing.

[Nair, 2022] Nair, A. T., Anitha, M. L., & Kumar, A. (2022, March). Disease Grading of Diabetic Retinopathy using Deep Learning Techniques. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1019-1024). IEEE.

[Pouyanfar, 2018] Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, *51*(5), 1-36.

[Robiul, 2020] Islam, M. R., Hasan, M. A. M., & Sayeed, A. (2020, June). Transfer learning based diabetic retinopathy detection with a novel preprocessed layer. In *2020 IEEE Region 10 Symposium (TENSYMP)* (pp. 888-891). IEEE.

[Rong, 2019] Rong, G., Mendez, A., Assi, E. B., Zhao, B., & Sawan, M. (2020). Artificial intelligence in healthcare: review and prediction case studies. *Engineering*, *6*(3), 291-301.

[Samy, 2019] Samy, S. (2019). *Réseaux de neurones convolutionnels pour la détection précoce de la rétinopathie diabétique* (Doctoral dissertation, Université Mouloud Mammeri).

[San-Li, 2021] Yi, S. L., Yang, X. L., Wang, T. W., She, F. R., Xiong, X., & He, J. F. (2021). Diabetic Retinopathy Diagnosis Based on RA-EfficientNet. *Applied Sciences*, *11*(22), 11035.

[Sandler, 2018] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

[Sarker, 2021] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 1-21.

[Shetgaonka, 2021] Shetgaonkar, P., Aswale, S., Naik, S., Gaonkar, A., Gawade, S., & Mhalsekar, P. (2021, November). Diabetic Retinopathy Detection and Classification from Fundus Images Using Deep Learning. In *2021 International Conference on Technological Advancements and Innovations (ICTAI)* (pp. 133-139). IEEE.

[Shanthamalar, 2022] Shanthamalar, J. J., & Ramani, R. G. (2022). Automatic Blood Vessel Segmentation in Retinal Fundus Images Using Image Enhancement and Dynamic Gray-Level Thresholding. In *Proceedings of International Conference on Computational Intelligence and Data Engineering* (pp. 337-348). Springer, Singapore.

[Stephenson, 2019] Stephenson, N., Shane, E., Chase, J., Rowland, J., Ries, D., Justice, N., ... & Cao, R. (2019). Survey of machine learning techniques in drug discovery. *Current drug metabolism*, *20*(3), 185-193.

[Thakur.2020] Thakur, N., & Juneja, M. (2020). Pre-processing of retinal images acquired from digital fundus cameras for improved performance of diagnostic tools in biomedical engineering. *Materials Today: Proceedings*, *28*, 1525-1529.

[Tymchenko, 2020] Tymchenko, B., Marchenko, P., & Spodarets, D. (2020). Deep learning approach to diabetic retinopathy detection. *arXiv preprint arXiv:2003.02261*.

[Vinuja, 2021] Vinuja, S., TR, K. K., & Karthika, R. (2021, September). Performance Analysis of Diabetic Retinopathy Classification using CNN. In *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 823-828). IEEE.

[Vaibhavi, 2021] Vaibhavi, P. M., & Manjesh, R. (2021). Binary Classification of Diabetic Retinopathy Detection and Web Application. *International Journal of Research in Engineering, Science and Management*, *4*(7), 142-145.

[Vrbančič, 2020] Vrbančič, G., & Podgorelec, V. (2020). Transfer learning with adaptive fine-tuning. *IEEE Access*, *8*, 196197-196211.

[Wang, 2022] Wang, N., Wang, Y., & Er, M. J. (2022). Review on deep learning techniques for marine object recognition: Architectures and algorithms. *Control Engineering Practice*, *118*, 1044580.

[Wang, 2021] Wang, S., Celebi, M. E., Zhang, Y. D., Yu, X., Lu, S., Yao, X., ... & Tyukin, I. (2021). Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects. Information Fusion, 76, 376-421

[Wejdan, 2020] Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, *20*, 100377.