Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

Democratic and Popular Republic of Algeria
Ministry of Higher Education and Scientific Research
Aboubekr Belkaid University – Tlemcen –
Faculty of Science
Department of Computer Science

Graduation Project

To obtain the Master's degree

In: Computer Science
Specialization: Artificial Intelligence

On the subject:

Deep learning-based Retinopathy Diagnosis Using Fundus photography (FP) and Optical Coherence Tomography (OCT) images

Written by:

Ms BENYELLES Nour El Houda **Ms BENYELLES** Nesrine

Defended in June 2025, in front of the jury composed of:

Mr BERRABAH Sid Ahmed	MCB, University of Tlemcen	*President*
Mrs ILES Nawel	MCB, University of Tlemcen	*Examiner*
Mr HADJILA Fethallah	MCA, University of Tlemcen	*Supervisor*

Academic Year: 2024/2025

Acknowledgments

First , we thank **ALLAH** for granting us the strength, perseverance, and guidance to complete our studies and bring this thesis to fruition.

We extend our heartfelt thanks to the members of the jury **Mr. Berrabah Sid Ahmed**, President of the jury, and **Mrs. Iles Nawel**, Examiner, for evaluating our research and providing constructive feedback .

We also would like to express our sincere gratitude to our supervisor, **Mr. Hadjila Fethallah**, for his dedicated guidance, serious commitment, and valuable support throughout the course of this work.

Our deep appreciation goes to **Mr. Lazouni**, Director of the Ophthalmology Clinic, for his warm welcome and for granting us access to his clinic in a spirit of generosity and collaboration .

Finally, we thank all those who, directly or indirectly, contributed to the successful completion of this work .

Dedication

I dedicate this work to **my dear father**, who has always been my role model and source of inspiration.

To **my mother**, whose constant support and precious words carried me through the most decisive moments.

To my one and only **sister**, truly unique and irreplaceable, who has always been there for me and helped bring out the best version of myself.

To my little **brother**, the youngest treasure of our family, whose presence brings so much joy.

I also dedicate this work to my nephew **SAMI**, the sunshine of my life, and to **my sister's husband**, for his kindness.

To my professor **Mrs. SEKKAL Nawel**, for her continuous support throughout all these university years and for her unwavering generosity.

Finally, I dedicate this work to all **my dear friends**, with whom I've shared both the best and worst moments .

Nour El Houda

Dedication

I dedicate this work to my dear **father**, who has always encouraged me to aim higher, inspired me deeply, and taught me the true value of perseverance and determination.

To my beloved **mother**, whose silent understanding speaks louder than words, and whose boundless love and gentle soul have been the quiet strength behind every step I take.

To my **brother**, whose quiet support, constant care, and steady presence have meant the world to me. Thank you for always being there even when no words were needed.

To my entire **family**, your encouragement and warmth have helped me stay grounded and motivated. I'm truly blessed to be surrounded by such love.

And finally, to my soul sisters **Hadjer, Riheb, Imene, Yasmine and Samira**, thank you for being my light, my laughter, and my safe space. Your love and friendship have been a true gift throughout this journey.

Nesrine

Abstract

ABSTRACT

Diabetic retinopathy (DR) is a dangerous complication of diabetes that affects the retina and can lead to partial or complete vision loss. It progresses through five clinical stages, and the majority of patients are diagnosed at a late stage, when their vision is already severely impaired. Therefore, it is essential to diagnose DR in its early stage in order to prevent irreversible damage. With the rise of artificial intelligence, deep learning has proven to be a useful tool in medical imaging, delivering accurate and fast diagnostic support.

In this work, we explored the use of deep learning models for automatic detection and classification of diabetic retinopathy stages. We used two publicly available datasets: the APTOS 2019 dataset from Kaggle, which involves a five-class classification problem, and OLIVES, a multimodal dataset published by research institutions, used for binary classification between DR and Diabetic Macular Edema (DME).We tested different convolutional neural network (CNN) architectures , including custom-built and pre-trained models such as EfficientNet and InceptionResNetV2. To improve performance and address class imbalance,we applied training techniques like data augmentation, oversampling, and class weighting . Our results support that deep learning can effectively assist in the early diagnosis of diabetic retinopathy and enable the development of reliable and scalable screening systems in clinical practice.

Keywords:

Diabetic Retinopathy (DR), Deep Learning, Medical Imaging, Convolutional Neural Networks (CNN), EfficientNet, InceptionResNetV2, Data Augmentation, Class Imbalance.

RÉSUMÉ

La rétinopathie diabétique (RD) est une complication dangereuse du diabète qui affecte la rétine et peut entraîner une perte partielle ou totale de la vision. Elle évolue en cinq stades cliniques et la majorité des patients sont diagnostiqués à un stade tardif, lorsque leur vision est déjà gravement altérée. Il est donc essentiel de diagnostiquer la RD à un stade précoce afin d'éviter des dommages irréversibles. Avec l'essor de l'intelligence artificielle, l'apprentissage profond s'est révélé être un outil utile en imagerie médicale, offrant une aide au diagnostic précise et rapide.

Dans ce travail, nous avons exploré l'utilisation de modèles d'apprentissage profond pour la détection automatique et la classification des stades de la rétinopathie diabétique. Nous avons utilisé deux ensembles de données accessibles au public : l'ensemble de données APTOS 2019 de Kaggle, qui implique un problème de classification à cinq classes, et OLIVES, un ensemble de données multimodal publié par des institutions de recherche, utilisé pour la classification binaire entre la RD et l'œdème maculaire diabétique (OMD).Nous avons testé différentes architectures de réseaux de neurones convolutifs (CNN), y compris des modèles personnalisés et pré-entraînés tels que EfficientNet et InceptionResNetV2. Pour améliorer les performances et traiter le déséquilibre des classes, nous avons appliqué des techniques de formation telles que l'augmentation des données, le suréchantillonnage et la pondération des classes . Nos résultats montrent que l'apprentissage profond peut contribuer efficacement au diagnostic précoce de la rétinopathie diabétique et permettre le développement de systèmes de dépistage fiables et évolutifs dans la pratique clinique.

Mots-clés :
rétinopathie diabétique, apprentissage profond, imagerie médicale, réseaux de neurones convolutifs, EfficientNet, InceptionResNetV2, augmentation de données, déséquilibre des classes.

ملخص

يُعدّ اعتلال الشبكية السكري (DR) من المضاعفات الخطيرة لمرض السكري، حيث يؤثر على شبكية العين وقد يؤدي إلى فقدان جزئي أو كلي للبصر. يتطور المرض عبر خمس مراحل سريرية، وغالبًا ما يُشخَّص المرضى في مراحل متقدمة، عندما يكون البصر قد تضرر بالفعل بشكل كبير. لذلك، يُعدّ التشخيص المبكر أمرًا بالغ الأهمية لتفادي الأضرار غير القابلة للعلاج. ومع التقدم في مجال الذكاء الاصطناعي، أثبت التعلّم العميق فعاليته كأداة دقيقة وسريعة في دعم التشخيص الطبي، لاسيما في مجال التصوير الطبي.

في هذا العمل، قمنا بدراسة استخدام نماذج التعلم العميق للكشف التلقائي عن مراحل اعتلال الشبكية السكري وتصنيفها. اعتمدنا على مجموعتي بيانات متاحتين للعموم: مجموعة بيانات APTOS 2019 من منصة Kaggle، التي تتضمن تصنيفًا إلى خمس فئات، ومجموعة بياناتOLIVES متعددة الوسائط، المنشورة من قبل مؤسسات بحثية، والتي تُستخدم في التصنيف الثنائي بين اعتلال الشبكية السكري والوذمة البقعية السكرية (DME). قمنا باختبار عدة هياكل من الشبكات العصبية الالتفافية(CNN) ، بما في ذلك نماذج مصممة من الصفر وأخرى مدربة مسبقًا مثلEfficientNet و InceptionResNetV2. ولتحسين الأداء ومعالجة اختلال التوازن بين الفئات، طُبّقت استراتيجيات تدريب متعددة مثل زيادة البيانات، والإفراط في أخذ العينات، وترجيح الفئات. تؤكد نتائجنا أن التعلم العميق يمثل وسيلة فعالة للتشخيص المبكر لاعتلال الشبكية السكري، ويساهم في تطوير أنظمة فحص موثوقة وقابلة للتطبيق في الممارسة السريرية.

الكلمات الدالة :اعتلال الشبكية السكري، التعلم العميق، التصوير الطبي، الشبكات العصبية الالتفافية، EfficientNet، InceptionResNetV2،زيادة البيانات، اختلال التوازن .

Table of Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence.
ANN	Artificial Neural Network.
BERT	Bidirectional Encoder Representations from Transformers.
CNN	Convolutional Neural Network.
DL	Deep Learning.
DME	Diabetic Macular Edema.
DNN	Deep Neural Network.
DR	Diabetic Retinopathy.
FA	Fluorescein Angiography.
FNN	Feedforward Neural Networks.
GPT	Generative Pre-trained Transformer.
ML	Machine Learning.
MLP	Multilayer Perceptron.
MSE	Mean Squared Error.
NLP	Natural Language Processing.
NPDR	Non-Proliferative Diabetic Retinopathy.
OCT	Optical Coherence Tomography.
PDR	Proliferative Diabetic Retinopathy.
RL	Reinforcement Learning.
RNN	Recurrent Neural Network.
SSL	Semi Supervised Learning.
VEGF	Vascular Endothelial Growth Factor.

General Introduction

Deep learning, computer vision, and medical diagnosis are dominant areas of research in the past several years that lie at the intersection of medicine and artificial intelligence. Deep learning provides powerful tools to automatically extract helpful patterns from huge and complicated sets of data, computer vision enables visual data to be interpreted automatically with accuracy and consistency beyond human analysis . In medical diagnosis, the conjunction of these technologies is a paradigm shift, introducing new means of supporting clinical decision making processes, particularly when utilized in medical imaging.

The reasoning behind the employment of deep learning and computer vision in medical diagnosis is grounded in the growing requirement for faster, more accurate, and more consistent diagnostic procedures. Traditional image based diagnostics , as rudimentary are generally extremely reliant on specialist interpretation, which can be vulnerable to variation and scarcity, especially in environments of low resource. Computer aided analysis aims to supplement human expertise, reduce diagnostic error, and improve access to quality health care services .These technologies enable early and precise detection of pathological features, quantitative assessment of disease progression , and predictive modeling of patient outcomes.

Diabetic Retinopathy (DR) is a sight threatening retinal condition that is responsible for one of the leading causes of visual impairment and blindness in diabetic patients . In this study, we try to develop an automated diagnostic system based on color fundus images to classify diabetic retinopathy into five standardized clinical grades(No DR,Mild,Moderate, Severe, Proliferative DR) .This is the central task of our dataset that presents several challenges, The first among these is not only the need to achieve high overall classification accuracy, but also consistent performance across all the individual classes. This is further complicated by the severe class imbalance , where some classes are greatly more prevalent than others . In answer to this, our objective are Improving per class accuracy, especially for the underrepresented stages,and employing good methods of addressing class imbalance.

We also examine a second dataset which unites two dissimilar imaging modalities: color fundus photography and Optical Coherence Tomography (OCT). The two modalities present complementary visual images of the retina, Fundus photographs give a generalized surface view of the retina , and OCT cross-sectional images centered on the retinal layers and macula. However, this multimodal configuration raises new challenges, particularly due to differences in image resolution and dimensions.

In order to address these challenges, in this work, we proposed a series of deep learning techniques applicable to each of the data sets. For the first data set we used three models:

- A model from-scratch, with two dense layers and two dropout layers.
- A transfer learning model with EfficientNetB0, and another on Inception-ResNetV2.both developed employing the identical two-layer dense network architecture and two dropout layers .

In the second data set , we used two recent state of the art transfer learning architectures:

- EfficientNetB2 .
- InceptionResNetV2 .

again with two densely connected layers and two dropout layers.

This Master thesis is divided into three chapters :

- The first chapter lays down the basic principles of deep learning and provides important strategies pertinent to medical image analysis.
- The second chapter presents a medical summary of diabetic retinopathy, outlining its causation, clinical phases, and related risk factors.
- The third chapter outlines the deployment of our deep learning models, the data used, experimental setup, and a thorough analysis of the achieved results.

Chapter 1

Deep learning

1 Introduction

Deep Learning (DL) has been among the best fields of artificial intelligence in the past few years. It is a subset of machine learning involving multi layer neural networks to learn automatically complex patterns from vast datasets. This feature has achieved monumental success in many fields .

With the thrust of computer power and access to big data, deep learning has become the need of hours to solve real world issues with high accuracy . This chapter provides a detailed overview of deep learning, starting from the fundamental concepts of neural networks to more advanced techniques and architectures . By the end of this chapter , we will have a sound understanding of the principles of deep learning and its pivotal role in modern artificial intelligence.

2 Artificial Intelligence

The process of Artificial Intelligence (AI) art is one of mimicking human intelligence. that requires the development of ”intelligent agents”, or more accurately systems that can perceive their environment, analyze information, and act on those perceptions to maximize their chances of successfully achieving their predefined goals [1] . AI enables machines to perform functions such as learning, problem solving, decision making that would otherwise require cognitive reasoning by the human mind [2] .

In other words, the goal of artificial intelligence is to develop techniques by which computers can simulate human behavior, and reasoning to improve their behavioral interaction with, and adaptability to complex environments [3].

3 Machine learning

Machine Learning (ML) is a subset of artificial intelligence that systematically applies algorithms to recognize and combine patterns in data, to enable machines to learn and improve their decision making capabilities in a variety of tasks without human intervention [3][4].

ML models make use of an appropriate combination of mathematical functions or rule based systems, being trained on input data sets for potentially achieving high degrees of accuracy while classifying or predicting [5].

4 Types of machine learning

There are three types of Machine Learning, depending on how these models learn from data :

4.1 Supervised learning

The algorithm learns by analyzing a training data set, which contains some data and labels that indicate that this data has the underlying patterns .

In the training step, these learned techniques represent the patterns in the form of a mathematical description, generally called a model. In the inference step, this model is used to make predictions or information about new data that has never been seen before[6].

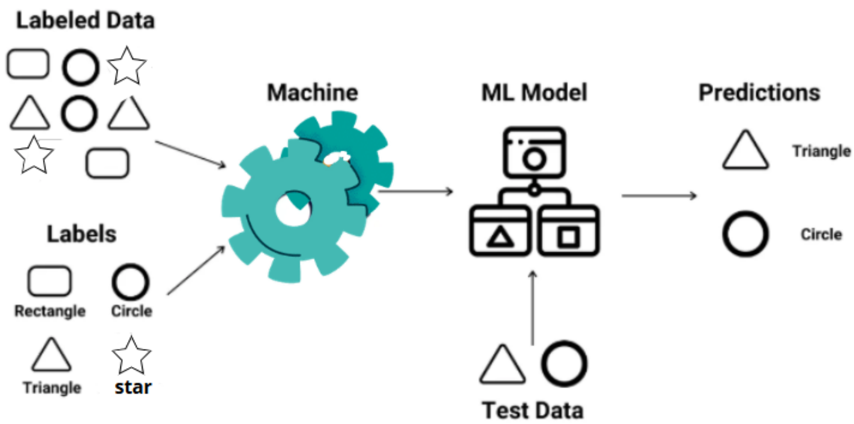


Figure 1.1 – Supervised Learning Diagram

4.2 Unsupervised learning

Unsupervised Learning is a machine learning approach that automatically identifies patterns in datasets where the data points have no labels or structure(Figure 1.2).

Unlike supervised learning, this method provides an artificial intelligence system with input data without corresponding output labels [7]. In this paradigm, the system itself detects specific patterns and structures within the data without any prior specification of background or labeling [8].

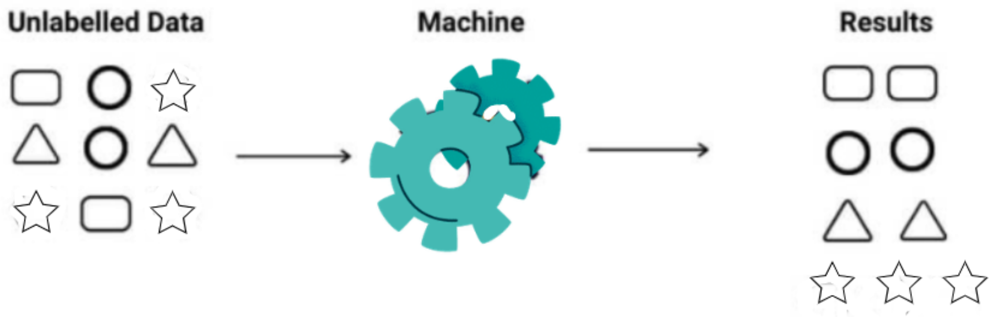


Figure 1.2 – Unsupervised Learning Diagram

4.3 Semi-supervised learning

Semi Supervised Learning (SSL) uses labeled and unlabeled data at the same time to improve the performance of machine learning models. SSL relies on assumptions such as data continuity and class separation. A few of the commonly employed techniques include generative models, consistency regularization, graph based methods, and self labeling. These techniques are particularly useful when labeling by humans is expensive or impractical [9].

4.4 Reinforcement learning

In simple terms, Reinforcement Learning (RL) involves an agent learning to make sequential decisions through interaction with the environment. Instead of receiving direct input-output pairs, the agent is presented by a description of its current state, some goal, choices of actions that it could take, and constraints on those actions[8]. Since the agent will try different strategies through trial and error, it will learn to maximize the cumulative reward.

The core of RL revolves around the construction of a behavior policy which can link states or situations to actions for the purpose of maximizing long term rewards. Such approaches are very useful in dynamic environments where learning can only be done through experience [10].

5 Foundations of Deep Learning

DL is a Subset of machine learning in which computers learn by example and discover higher level concepts in data. Deep learning is different from traditional algorithms where humans have to provide the necessary features and rules, it enables machines to learn on their own by learning to combine certain features that are then trained in multiple layers to form new concepts [11].

Essentially, deep learning is rooted in Deep Neural Network (DNN), where the data is computed via multiple layers of artificial neurons. The layers pull out and convert features from raw data in a series of linear and non linear transformations, learning abstract representations incrementally at each level, simulating the functioning of the human cerebral cortex [5].

5.1 Basic concept of Deep Learning

5.1.1 Artificial neural networks

Artificial Neural Network (ANN), are computing models that were inspired from the biological neural networks of the human brain. Their architectural designs allow for processing information mimic how biological neurons transmit and process information [12].

An ANN consists of interconnected processing units, also known as neurons, which hence work together for the purposes of recognizing patterns, processing inputs, and making predictions [13].

ANN models consist of three primary layers:

- **Input layer :** to receive raw data.
- **hidden layer :** in which the data is processed through interconnected neurons using different mathematical transformations in order to extract features and patterns.[14]

- **Output layer:** which produces the final prediction or classification result .

This architecture is illustrated in (Figure 1.3).

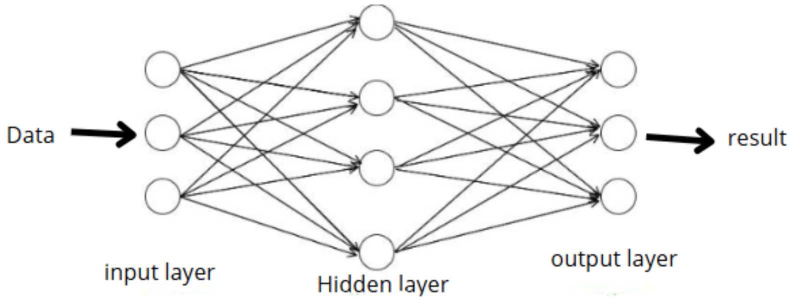


Figure 1.3 – ANN Architecture

5.1.2 Hyperparameters in Deep Learning

The performance of Deep Learning (DL) models is highly reliant on the hyperparameter setting particular to the dataset. Therefore, Hyperparameter Optimization is a key step while designing DL models, and have a dramatic effect on accuracy, generalization, and other performance measures [15].

Major Hyperparameters of Deep Learning are :

- **Learning Rate:** This hyperparameter determines the optimizer step size in each training iteration. Slow convergence is caused by a low learning rate, whereas divergence and instability are caused by a high learning rate[15].
- **Epochs:** Number of complete passes through the training set.
- **Batch Size:** Is how many examples go through the network before parameters get updated.
- **Number of Hidden Layers:** Determines the depth of the network.
- **Activation Functions:** Activation functions such as ReLU, Sigmoid, and Tanh introduce non-linearity to the network so that the network is capable of learning intricate decision boundaries and structures [15].
- **Optimizer:** Optimizers such as Adam, that update the weights during training in order to reduce the loss function.
- **Dropout Rate:** This is a method of regularization that randomly shuts off some neurons during training to prevent overfitting.

5.1.3 Loss Functions

A loss function is one of the significant components in deep learning to quantify how good a prediction by a model is compared to the real. it computes the difference between the predicted and actual outcomes.Good predictions imply low loss, while high loss implies big prediction errors.

Trained models attempt to minimize the loss by adjusting their internal parameters. It is trained when the loss is minimized below a given value .

The top most Loss Functions are :

- **Mean Squared Error (MSE) :** Mean Squared Error (MSE) is the mean squared difference between the target output and the predictions.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **Cross-Entropy Loss :** It estimates the difference between true class labels and predicted probabilities. The predicted probabilities must lie in the range of 0 and 1, thus it is ideally appropriate for classification problems .

$$\text{Cross-Entropy} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

6 Neural Network Architectures

6.1 Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a specific type of supervised neural network that employs the backpropagation algorithm to learn adjusting its weights to improve performance. It is one of the feedforward neural networks, where information flows just in one direction ,from the input layer to the output layer with no feedback connections [16].

An MLP comprises multiple layers of neurons that are interconnected:

- **Input Layer:**It receives the raw input and transmits signals to the hidden layers, There is one neuron for each input feature of the problem.
- **Hidden Layers:** They perform complex transformations of the data , and consist of neurons activated by a nonlinear function to enable the network to learn nonlinear relationships within the data.

- **Output Layer:** It gives the network's end prediction. The Number of neurons within this layer is one neuron per class in classification.

All neurons in a layer of an MLP are fully connected to neurons in the next layer, this Connections are denoted by weights, which are changed during training to improve prediction performance .

An MLP is trained through supervised learning, where the network successively updates its weights in order to minimize the difference between its output and true target values [17] . This is achieved in a two-stage process known as backpropagation, where forward propagation and backward propagation take place.

In the forward propagation phase, the input data is propagated layer by layer into the network. Each neuron takes the input from the layer before it, weights it with corresponding weights, applies a bias and applies an activation function to produce an output. This is done all the way until the final output layer, where the network outputs a prediction . To determine the error, a loss function is used to find the difference between the predicted and actual value .

Following calculation of the error, the process of backward propagation begins. This is the process in which error is propagated back through the network to update model parameters (bias and weights). This is done via gradient descent, an algorithm used to move weights in a direction towards minimization of the loss.

This forward propagation, calculation of error, and backpropagation cycle is repeated multiple times (epochs) until the model is at the optimal solution level where the error is minimized .

6.2 Convolutional Neural Networks (CNN)

Convolutional Neural Network (CNN) is a class of deep learning models specifically created to handle grid structured data like images. CNNs enabled computer vision through the tremendous boost in the ability of machines to recognize and classify objects with great accuracy .

The most significant strength of CNNs is their ability to automatically learn and extract relevant features from images. They achieve this by increasingly extracting low-level features (e.g. edges and textures) in the early layers and increasingly combining them to create more complex, high level features (e.g. shapes and objects) in the deeper layers[19]. This hierarchical feature extraction renders CNNs highly efficient and effective at handling large scale image datasets.

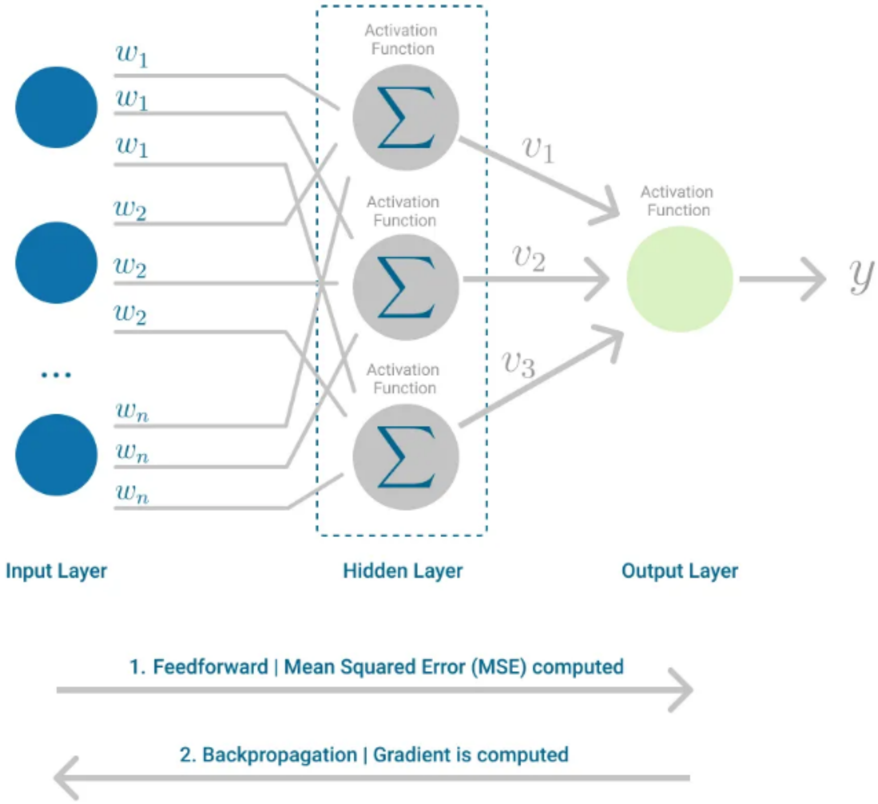


Figure 1.4 – The process of MLP
[18]

6.2.1 CNN Layers and Architecture :

A CNN is made up of a sequence of layers that transform the input image into an understandable representation step by step [20] . The following is a detailed explanation of each type of layer and its role in the network.

- **Convolutional Layer:**

  The convolutional layer is the foundation of CNNs. that employs convolution operations to remove important features from an input image. A convolution operation involves sliding a small matrix, referred to as a filter (or kernel), over the input data to compute a weighted sum at each location, and filters detect particular patterns such as edges, texture, and shape in the image[21].

  Each filter generates a feature map that indicates the locations where the particular patterns occur, with multiple filters. the network learns to detect multiple patterns at different locations in the image. To ensure that the entire image is processed, filters are slid over all portions of the image.

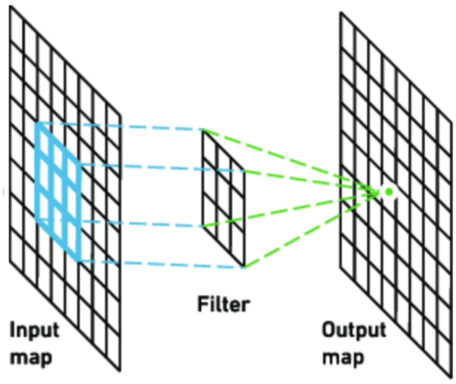


Figure 1.5 – Convolution operation
[22]

- **Pooling Layer:** Pooling layers decrease spatial dimensions of the feature maps reducing parameters and computation efficiently. They make the CNN simpler as it is less sensitive to small variations or distortions in the input image.

  Most popular pooling techniques are:

  — Max Pooling: Maximizes max value over small patch (e.g., 2×2 or 3×3), keeping only the strongest features.

  — Average Pooling: It averages values in a region, thus denoising.

- **Fully Connected Layer (Dense Layer):** The feature maps are flattened to a one dimensional vector following feature extraction and fed into one or more fully connected layers (dense layers). They are just like an ordinary feedforward neural network and are utilized to make final predictions. Every neuron is connected to every neuron in the preceding layer[23].

  The features extracted are integrated to create high level representations for classification, and a final Softmax layer is used for multi-class classification, which generates probability scores for each class.

  For example , in an image classification task, the final fully connected layer can generate probability scores for different classes (e.g., Mild DR, Moderate DR, Severe DR), and the class having the highest probability is the predicted class like shown in (Figure 1.6) .

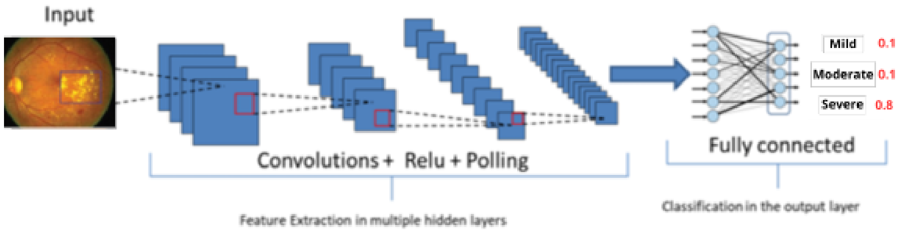


Figure 1.6 – Convolutional neural network

6.3 Recurrent Neural Networks (RNN)

A Recurrent Neural Network (RNN) is a deep learning algorithm especially applied to manage sequential data, such as time series, and text. In contrast to Feedforward Neural Networks (FNN), which process data in a single direction (input to output), RNNs have recurrent connections where data can be saved and reused at different times [24].

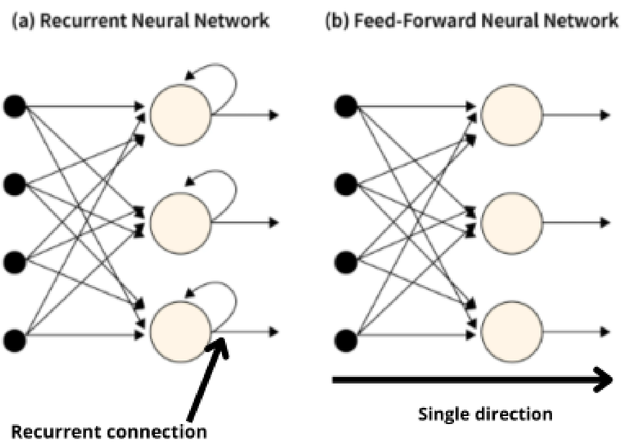


Figure 1.7 – The difference between RNN and FNN

The fundamental idea of RNNs is that they have internal memory (hidden state) to remember past inputs. This makes them ideal for understanding context and sequence related tasks. • **The Mechanism of RNNs:**

1. **Accepting an input:** The network accepts a new data point. Such as a word in a sentence.
2. **Updating the hidden state:** The new input is combined with the last hidden state that was remembered, in order to get the updated internal memory of the network[25].
3. **Generating an output:** Depending on the task, the hidden state can be further transformed to generate an output at each time.

4. **Propagating information:** The new hidden state is then fed to the next time step, enabling the network to have memory over several steps in the sequence [25].

6.4 Transformers

Transformers are a robust type of deep neural networks (DNNs) developed to address some serious limitations of earlier models used for sequential data, such as recurrent neural networks (RNNs) . Traditional sequence-to-sequence (seq2seq) models process input one step at a time, resulting in slow training and incapable of learning long distance relations among items in a sequence effectively, They also focus on short term dependencies [26].

To address such problems, Transformers introduce a process called self-attention(multi-head self-attention). With this process, the model becomes capable of inspecting the full input sequence as a whole at once and extracting what is most important for a given position . As a result, Transformers become more capable of capturing long-distance dependencies and learn much faster compared to RNNs since the sequence is handled in parallel, not sequentially [26].

The original Transformer model is made up of an encoder and a decoder, both consisting of multiple identical layers stacked on top of each other.each encoder layer contains two main components:

A multi-head self-attention mechanism, which allows the model to view different parts of the input sequence from multiple perspectives.

A feed-forward neural network (FFN) applied to each position separately.

To make it easier for the model to learn more smoothly and avoid complications, each sub-layer is wrapped in a residual connection and Layer Normalization follows it.

The decoder layers follow the same pattern but have one added part , it a cross-attention layer to allow the decoder to pay attention to the relevant areas of the encoder output. furthermore, the decoder self-attention is masked so that the model cannot see ahead to the future positions when generating output, something that is necessary for text generation applications [27].

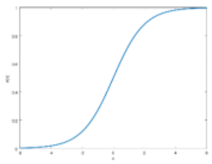
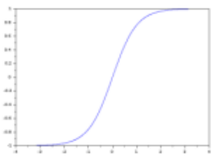
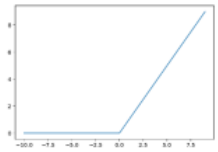
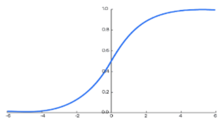
7 Learning algorithms and techniques

7.1 Activation functions

An activation function determines whether a neuron in a neural network should fire or not based on the input. It determines how important the input of the neuron is to make the prediction using a mathematical operation.

Its primary role is to convert the weighted summed inputs of a neuron to an output value that is propagated on to the next layer within the network or utilized as the final output [28].

The most important activation functions are shown in the following table:

Name	**Graph**	**Function**
Sigmoïde		$f(x) = \frac{1}{1+e^{-x}}$
Tanh		$f(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$
ReLU		$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$
Softmax		$f(x) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$

7.2 Optimizers

Optimization algorithms in deep learning are methods used to minimize the loss function, a measure of how much the model's predictions diverge from actual outcomes. Optimization methods iteratively adjust the model's parameters ,such as weights and biases in an effort to improve performance[29].

7.2.1 Gradient Descent

Gradient Descent is a first-order iterative optimization algorithm used to find the local minimum or maximum of a function. In ML and DL, it is used broadly to minimize the loss function.

Gradient Descent will converge to local minima and not the global minimum. Local minima are points where the loss is lower than points around it but not necessarily the point of lowest position.

How Gradient Descent Functions :

1. **Begin with Starting Parameters:**It begins with initial random values of the model's parameters (like weights).

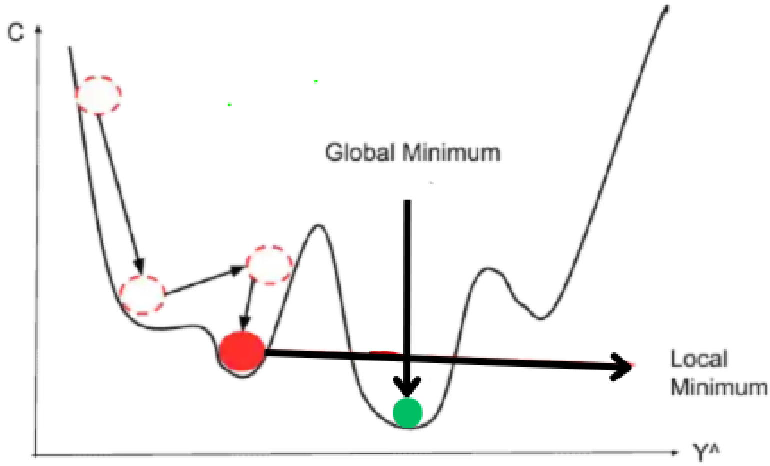


Figure 1.8 – The difference between local and global minima

2. **Compute the Gradient:** It calculates the gradient (slope) of the loss function to determine the direction of steepest ascent.
3. **Update Parameters:** The parameters are updated by moving in the opposite direction to the gradient (in order to decrease the loss). The size of this move is controlled by the learning rate.but If the learning rate is very low, the algorithm will take small steps and converge slowly, and If it's too high, it may overshoot the minimum or fail to converge(Figure 1.9).
4. **Repeat:** These are repeated until the loss is minimized to a point or until a specific number of iterations are obtained.

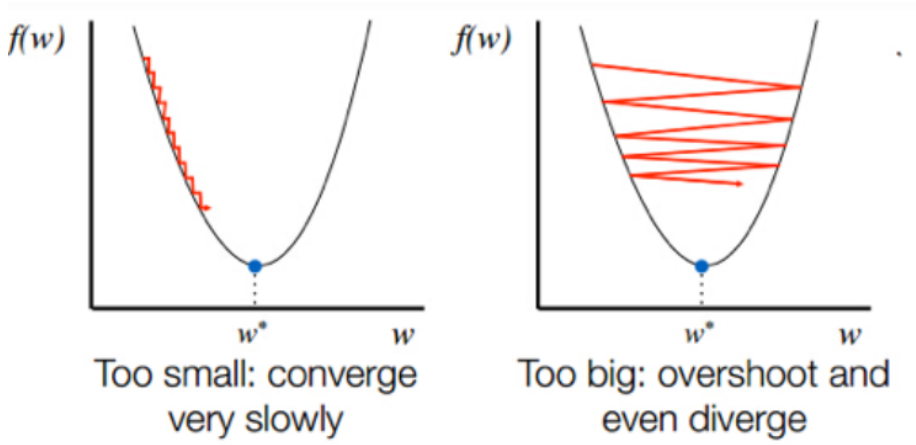


Figure 1.9 – The impact of choosing wrong Learning rate
[30]

7.2.2 Adam

The Adam optimizer (Adaptive Moment Estimation) is an iterative algorithm for optimization that is used to reduce the loss function while training neural networks.

It is popular because it can dynamically change the learning rates of each parameter, which makes it very suitable for large datasets and complicated models.

Adam uses adaptive scaling of learning rates through first and second moment gradients estimates. The key steps in the Adam process are:

- **Initialization:** Two zero-initialized moving averages: The first moment keeps track of the average gradient (how steep the slope becomes).
  The second moment keeps track of the uncentered variance of the gradients (how rapidly the slope is changing).
- **Gradient Computation:** In each iteration, the optimizer computes the gradient of the loss function with respect to the model parameters.
- **Moment Updates:** The first moment (mean gradient) is updated from values that have preceded it.
  The second aspect (variance of gradients) is revised to indicate how the gradients vary with time.
- **Bias Correction:** As the moving averages are being initialized, Adam employs bias correction to ensure the estimates are as accurate as possible, especially in the initial stages of training.
- **Parameter Update:** Adam uses the updated moment estimates to update all parameters of a model in order to improve its performance.
  This iterative procedure is repeated until the model achieves its optimal or near-optimal state.

7.2.3 Regularization

Regularization is a method in deep learning and machine learning that avoids overfitting and enhances the generalization capacity of a model. Regularization adds a penalty term to the loss function during model training to keep it away from being overly close to the training data, including noise and non generalizable information.

A poorly regularized model might be too complex and hence learn not just the general trends in the data but also their anomalies and noise. This leads to excessive memorization of training data with poor performance on unknown data, which characterizes the phenomenon of overfitting[31]. As shown in (Figure 1.10):

- **Regularization Techniques**

1. **Dropout:**

   Dropout is a regularization method applied to deep neural networks. For every iteration during training, a random set of the neurons are temporarily

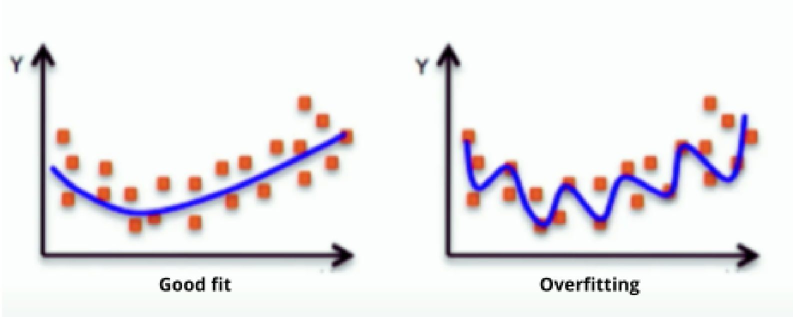


Figure 1.10 – Overfitting

deactivated (all their connections too), so the network is not over dependent on the neurons[31].

2. **Batch Normalization:**

   Batch Normalization standardizes the input of every layer by subtracting mini-batch mean and dividing by mini-batch standard deviation .this technique Regularizes training by reducing the dependence on weight initialization.

3. **Data Augmentation:**

   Data augmentation is a regularization technique that is commonly used in computer vision and natural language processing (NLP). It does generate new data from old data by applying random transformations such as rotations, flips, and translations

4. **Early Stopping:**

   Early stopping is a regularization method that monitors validation set performance and stops training when a degradation is observed.it Avoids wasting time training an overly complex model.

8 Evaluation Metrics

Evaluation metrics are quantitative measurements for gauging the performance and correctness of a machine learning or statistical model. They provide an unbiased comparison between different models and help during hyperparameter tuning to achieve optimal output.

1. **Confusion Matrix:** A confusion matrix is an N×N matrix, where N is the number of predicted classes. It is heavily employed in classification tasks, offering a high level analysis of the model prediction mistakes.It is particularly convenient to compute precision, recall, specificity, accuracy.

Definitions Related to the Confusion Matrix :

- **True Positive (TP):** The instance is positive, and the model has predicted it as positive accurately.
- **True Negative (TN):** The instance is negative, and the model has predicted it as negative accurately.
- **False Positive (FP):** The sample is negative but the model incorrectly reports it as positive.
- **False Negative (FN):** The sample is positive but the model incorrectly reports it as negative.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Figure 1.11 – Confusion matrix

2. **Accuracy :** The Accuracy estimates the number of correctly predicted samples out of the total number of samples:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of samples}} \tag{1.1}$$

3. **Precision and Recall:**

- **Precision:** Proportion of positively predicted instances actually correct out of total instances predicted positive.

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1.2}$$

- **Recall (Sensitivity):** Proportion of actual positives predicted correctly by the model.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1.3}$$

4. **F1-Score :** The F1-score is the harmonic mean of recall and precision, of use in cases where class distributions are severely imbalanced.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{1.4}$$

9 Advanced techniques in deep learning

Over time deep learning has seen a substantial increase in its capabilities, resulting in newer approaches that improve both efficiency and performance of the resulting model. All these techniques aim at tackling several difficulties and challenges like overfitting, interpretability, computational cost, etc. This part explores some of the essential advanced techniques such as transfer learning, fine tuning, and ensemble learning.

9.1 Transfer learning

Transfer learning is an advanced deep learning technique. Training a model from scratch typically requires large datasets, which are not always available. This is where transfer learning comes in hand. The term ”transfer” refers to the idea that knowledge learned in one domain can be applied to another, reducing the need for extensive new training data.

This methode allows to create powerful models of DL using small datasets by building on models pre-trained on large datasets [32]. This is especially useful in areas such as medical imaging where collecting large amounts of labeled data can be difficult and expensive. It also significantly reduces the duration of training by using pre-trained models, as training from scratch can take days or weeks [33].

9.1.1 Notable Architectures

EfficientNet: optimizes accuracy and efficiency, using depth, width, and resolution via a uniform compound coefficient. It achieves greater performance at fewer parameters than older models, This makes that be ideal for tasks with some limited resources [34].

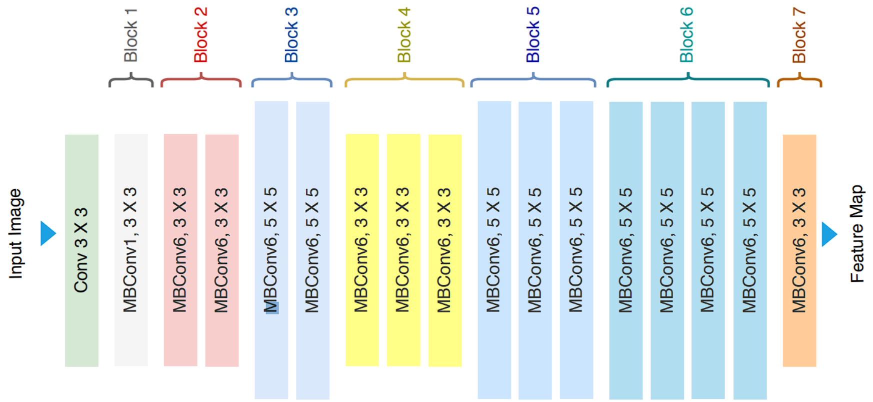


Figure 1.12 – EfficientNet Architecture
[35]

ResNet: uses skip connections, or "residuals" to allow gradients to flow through the network more easily during training. This does help to prevent that gradient problem in deep networks. ResNet is broadly used for its good performance. The neural network excels in image recognition tasks [36].

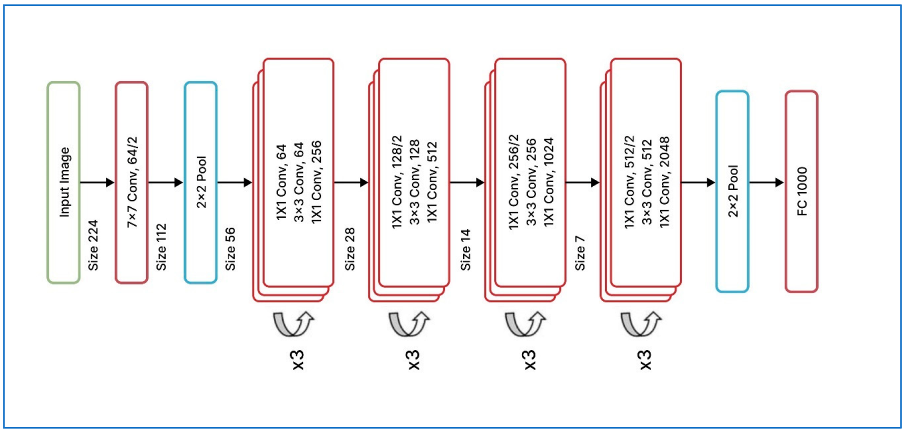


Figure 1.13 – ResNet50 architecture
[37]

InceptionNet : uses multiple convolution filters of different sizes in parallel within the same layer. This allows the model to capture features at various scales, making it efficient and accurate for image classification tasks. It was designed to optimize both computational efficiency and performance [38].

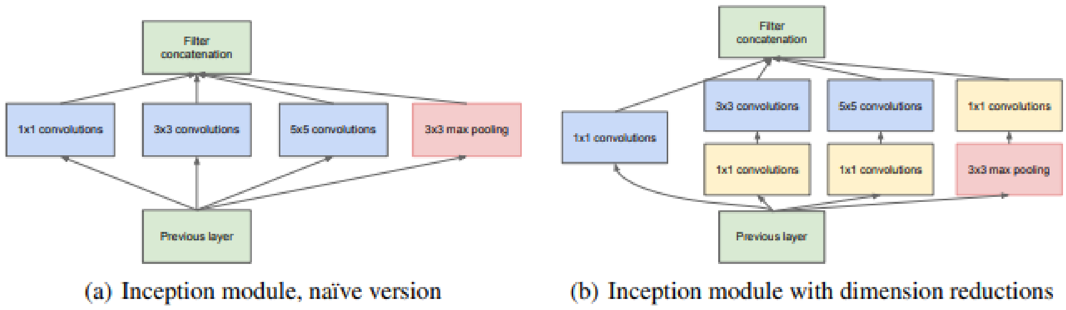


Figure 1.14 – InceptionNet architecture
[39]

Inception-ResNet : combines the strengths of Inception modules with ResNet's residual connections. This hybrid model improves learning speed and accuracy by enabling deeper architectures [40].

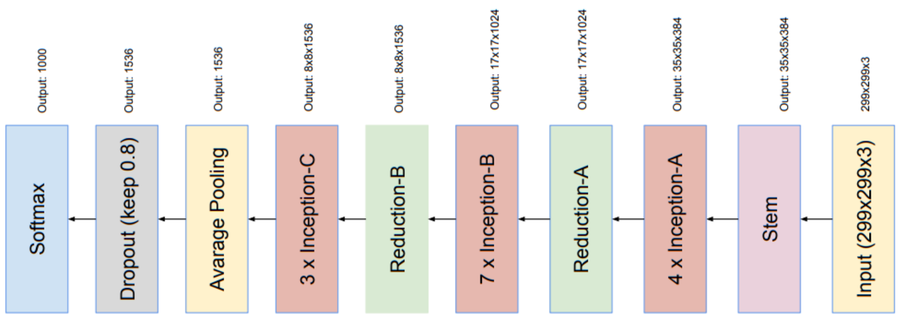


Figure 1.15 – Inception-ResNet General Architecture
[40]

9.2 Fine tuning

Another key aspect of transfer learning is fine-tuning, which involves adjusting the pre-trained model to better fit the new task. Instead of simply reusing the learned features as they are, fine-tuning allows us to update certain layers of the network—typically the last few layers—by continuing the training process with a smaller learning rate. This helps the model adapt more specifically to the target dataset while still benefiting from the knowledge gained in the original training [41].

9.3 Ensemble Learning

Ensemble learning is a technique within machine learning that combines multiple models so as to improve prediction accuracy as well as generalization , such methods reduce errors that are related to a high variance, high bias, or a low accuracy.

Parallel and sequential methods are the two main categories. Parallel ensembles, such as bagging as well as its popular extension, random forests, train multiple base models independently and then fully aggregate their predictions. This certain approach helps diversify the models and stabilize performance [42].

Ensemble learning can also be categorized into homogeneous and heterogeneous ensembles.Homogeneous ensembles use one same type of model for each base learner, while heterogeneous ensembles integrate diverse algorithms so as to benefit from these diverse strengths. Ensemble learning uses these techniques, and improves machine learning models robustness and accuracy, which makes it a widely used real world approach [43].

10 Deep Learning applications

10.1 Computer Vision

With the coming to the fore of neural networks in the last decade, computer vision research has grown exponentially, it involves teaching machines to interpret and understand visual information from the world, much like humans do [44].

Deep neural networks (DNNs), especially convolutional neural networks (CNNs), enable machines to identify, label, and perform more complex image processing tasks, such as facial recognition, driverless cars, and medical imaging [45].

10.2 Natural Language Processing

Natural Language Processing (NLP) is a branch of Artificial Intelligence that focuses on understanding and communicating human language. They encompass corpus analysis, information retrieval , machine translation , semantic analysis, text classification, and more. Those applications depend on how efficiently and quickly systems can process large volumes of text, and they have a direct impact on the advancement of information technology [46].

NLP has been transformed by deep learning and transformer architectures such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT) have pioneered this revolution. They can follow instructions, write essays, summarize long pieces of writing, and even translate between languages with high accuracy. Unlike previous methods, which depended heavily on manually defined rules or surface-level algorithms, deep learning systems have the ability to detect complex language structures from extremely large corpora

[47].

10.3 Deep Learning in Healthcare and Medicine

At present, a massive volume of biomedical data is available which is continuously altering the processes of disease diagnosis and treatment within the healthcare industry. Given the success of deep learning in various fields, it has naturally found promising applications in the medical domain [48].

Deep Learning is commonly applied in medical imaging and involves the use of X-rays, CT scans, and MRIs for diagnosing and detecting cancer, cardiovascular disease, and a whole host of other problems pertaining to neurology [49]. CNNs are particularly well-suited for image analysis tasks due to their ability to extract hierarchical features from visual data [50].

11 Advantages of Deep Learning

Deep learning has far and away been one of the most effective techniques in artificial intelligence, enabling machines to perform complex tasks with record accuracy.

However, Having knowledge of the strengths of deep learning is important to being able to effectively use it for real world tasks.

- **Handling Large and Complex Data :** Deep learning algorithms are best placed to handle extremely big datasets with complex structures, which are computationally intensive for ordinary machine learning algorithms to handle. This makes deep learning highly viable in big data applications.
- **Handling Non-Linear Relationships :** Deep learning models can fit very non-linear and complex relationships in data. This enables them to predict more accurately than typical machine learning models with linear assumptions.
- **Generalization Ability :** Deep learning models can generalize well to new situations by learning abstract and hierarchical representations of data. This makes them highly versatile and fit for a variety of real world applications.

12 Challenges and Future Perspectives of Deep Learning

12.1 Computational Complexity and Data Cost

It is well known that deep learning models can have extraordinary effectiveness, but often at high computational cost. Generally speaking, these models require large amounts of data and powerful computers to train on, which can be a significant barrier to entry. For example, training modern architectures is often reliant on expensive hardware such as clusters of GPUs or TPUs [51].

Furthermore, the expense goes beyond computation alone because the data itself needs to be carefully selected and tagged, which is a costly and time-consuming process [52].

These challenges become worse by the increasing size and complexity of modern models, which impacts environmental problems. This aspect led to the advent of the concept of Green AI, which focuses on developing models and methods that achieve high performance with reduced energy consumption and environmental impact. Researchers are currently discovering strategies such as model optimization, learning how to reduce training time and architectural design more effectively. Such efforts aims to balance the compromise between the effectiveness of the computer and the model performance, which makes it easier to learn about access to depth and more environmentally friendly, especially in resource related context [53].

12.2 Model Explainability and Interpretability

Interpretability and explainability are generally important as they provide insight into decisions made by deep learning algorithms. This is especially important in certain areas such as medicine. In this area, decisions made can have direct consequences on people's lives. Therefore, understanding how algorithms work in deep learning tasks can help ensure that the decisions made by these algorithms are correct and that errors are minimized.

Explainability refers to the ability to describe internal mechanisms and decision-making processes in human-understandable terms, and to describe the model's decision process. In contrast, according to Doshi-Velez and Kim interpretability is the ability of a model to provide explanations or present its reasoning in a way that is understandable to humans [54].

Various techniques have been developed for improving model explainability and interpretability. Methods like Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) offer comprehension into how features impact the model predictions. In computer vision, visualization techniques such as that of Grad-CAM help to highlight regions in an image that can be most relevant to a model's decision. Notwithstanding such advances, achieving any complete comprehension of complex deep learning models remains quite challenging. Active research, in addition to that, is dedicated to developing methods that can offer deeper understandings while maintaining performance [54].

12.3 Ethics and Bias in AI

As deep learning becomes increasingly prevalent in the medical field, it's crucial to consider the ethical implications—particularly how bias can influence outcomes. These systems have the potential to perform incredible tasks, such as analyzing images, comprehending language, and predicting health results [55].

However, due of unconscious biases in their training or construction, they are still capable of making mistakes, sometimes very catastrophic ones, despite their immense potential [55].

These biases aren't always easy to spot, but they can lead to unfair or even harmful decisions. Usually, bias in AI models falls into three main categories: data bias, algorithm bias, and engineer bias. Understanding where these come from is a key step in building safer and more trustworthy AI.

12.3.1 Data bias

Data bias is a frequent source of error in machine learning models, often caused by poorly selected or representative data. There are several types of data bias, including cognitive bias, selection bias and reporting bias. To reduce data bias, techniques such as under-sampling and over-sampling are used to balance the data and reduce the influence of these biases on the results [56].

12.3.2 Algorithm Bias

Data and the choice of machine learning models (estimators) that may not fit specific tasks can potentially cause algorithmic bias. Multiple comparative studies in different domains like agriculture, healthcare, business, etc. Show how algorithm performance varies by context, with ensemble or neural networks often yielding better accuracy and less bias. Additionally, various bias mitigation techniques have been proposed, such as adversarial debiasing, fairness constraints, and variational autoencoders, to ensure more equitable predictions across sensitive attributes [56].

12.3.3 Engineer Bias

When training or evaluating models, machine learning engineers can unintentionally introduce bias , often due to cognitive biases such as confirmation bias. As a result, they interpret data in ways that confirm their initial assumptions , and to reduce such biases, strategies like teamwork, challenging assumptions, and applying rigorous statistical methods are recommended. More objective and reliable assessments are also ensured by transparency, third-party evaluation, and comparing with baseline models [56].

12.4 Future Developments in Deep Learning

Here are some forward-looking perspectives on the future of deep learning proposed by Gary Marcus (researcher in cognitive science and artificial intelligence) and François Chollet (the creator of the deep learning library Keras) :

1.Models that resemble computer programs: Future deep learning models may incorporate features such as loops, conditional branches, memory mechanisms, and data structures such as lists and graphs. This will enable them to move beyond pattern recognition and in the direction of abstraction and reasoning [57].

2. Automated architecture design (AutoML): Future systems will no longer need to be designed with human network architecture and hyperparameter optimization. They will automatically learn best architectures via techniques such as reinforcement learning and evolutionary algorithms [57].

3. Recurring use of modular subroutines: We expect a move towards constructing models out of learned reusable components on previous tasks. These can be geometric blocks like convolutional layers or algorithmic blocks like reasoning routines, and this will lead to more efficient and more generalized learning [57].

4. The limits of deep learning for abstraction and generalization: Deep systems now are founded upon big amounts of data and can't generalize to anything outside of what they've seen. They're great at interpolation but weak at extrapolation , thus less capable with novel situations or abstract concepts [58].

5. The need for hybrid models combining symbolic and neural systems: Future AI will be supported by the integration of deep learning with symbolic systems. This combination might enable models to reason logically, to manipulate variables, and to utilize structured knowledge—abilities that neural networks currently do not possess [58].

6. In the direction of human-like unsupervised learning: There is a push to develop systems that, like children, can set their own goals, explore, and learn from unstructured experiences—without any explicit supervision or labeled data [58].

13 Conclusion

This chapter has provided for a deeper overview into deep learning, highlighting its role as a key subfield from of artificial intelligence. We examined its basic few concepts, key architectures, and algorithms that drive learning in the field. A real emphasis was placed at on how deep learning enables machines to process complex data, learn representations, in addition to perform tasks such as image recognition and language understanding with outstanding accuracy.

In the next chapter, we will focus on one of the many diseases for which deep learning has played a significant role in diagnosis: diabetic retinopathy. We will explore the nature of this condition and its prevalence, laying the foundation for acknowledging how deep learning is being applied in several medical diagnoses,including this one.

Chapter 2

Diabetic Retinopathy

1 Introduction

Diabetic retinopathy (DR) is one of the common complications of diabetes that affects the eyes, It may lead to vision problems and blindness if early diagnosis and treatment are not done.

DR progresses slowly and may be asymptomatic at first. It takes on many stages , starting with early changes in the retina through to severe forms such as irregular development of new blood vessels . Another condition, Diabetic Macular Edema (DME), which occurs along with DR, causes swelling of the central part of the retina and affects vision at any level of DR.

It is imperative to know diabetic retinopathy to be able to develop tools to diagnose and cure it. In this chapter, we present the most essential characteristics of DR, including its types, cause, and how it is usually diagnosed.

2 Definition and General Overview

2.1 Definition of diabetic retinopathy

Diabetes mellitus is a chronic metabolic disorder with a profound impact on multiple organs, including the eyes. One of the most common and vision-threatening ocular complications of diabetes mellitus is diabetic retinopathy (DR), which affects the retina, the light sensitive tissue layer at the back of the eye that receives visual information and transfers it to the brain [59].

DR is asymptomatic at its initial stage, and routine screening is required .it is typically diagnosed by ophthalmologists on fundus examination, where the eye pupil is expanded and retina examined with specialized equipment [60]. Clinical diagnosis relies on the presence of typical retinal lesions such as microaneurysms, hemorrhages, hard and soft exudates, cotton wool spots, and intraretinal abnormalities [61].

Globally, DR occurs in 30–40 % of individuals with diabetes [62] . It was the fifth avoidable cause of blindness and moderate to severe visual impairment globally between 1990 and 2010 [63]. It is also the principal cause of vision impairment among adults aged 20–74 years, impacting not just quality of life but also socioeconomic status,especially working-age adults with early-onset diabetes and proliferative disease, who have a higher risk of unemployment [64].

2.2 Diabetic retina

Diabetic retinopathy (DR) is primarily caused by chronic hyperglycemia-induced damage to the retinal microvasculature. The earliest clinically detectable lesions i.e., microaneurysms, intraretinal hemorrhages, and hard exudates are hallmarks of such microvascular damage [65]. These lesions result from breakdown of the blood retina barrier, microcapillary occlusion, and increased vascular permeability, leading to ischemia and retinal edema [66] .

Historically, pathophysiological understanding of DR has focused on vascular abnormalities. This is appropriate, in as much as the early ophthalmoscopic findings .e.g., venous beading, microaneurysms, and lipid exudates are direct reflections of capillary damage, as confirmed by histopathologic studies [65].

As observed in Figure 2.1, chronic hyperglycemia leads to leakage from damaged capillaries, resulting in retinal edema, hard exudate deposition (yellow spots), hemorrhages, and cotton wool spots classic signs of tissue ischemia. These changes distort the retinal architecture and result in blurred or distorted vision [66].

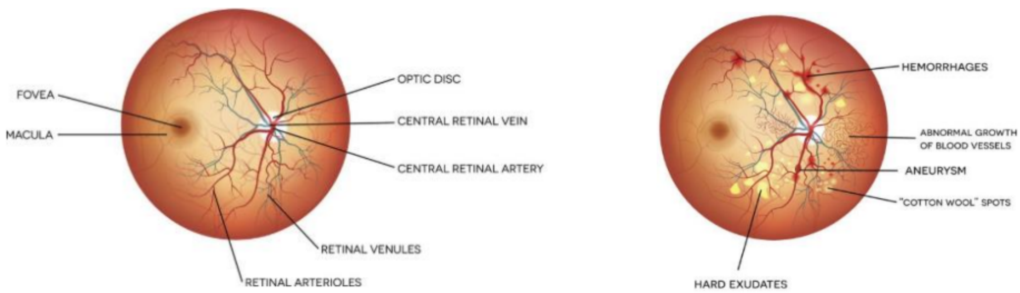


Figure 2.1 – Normal Retina vs Diabetic Retina
[67]

Interestingly, DR also evolves unnoticed over a period of many years in patients with unrecognized or inadequately controlled diabetes. Most individuals are unaware of the condition until after there has been widespread retinal involvement, emphasizing the need for early screening and stringent glycemic control to exclude or delay the onset of vision-threatening stages of DR [68].

3 Risk Factors

The clinically most significant risk factors for the course and development of diabetic retinopathy (DR) are poor glucose control, hypertension, and the duration of diabetes and genetics [67].

3.1 Duration of Diabetes

The longer the duration of diabetes, the stronger the risk factor for DR development and progression. In the Wisconsin Epidemiologic Study of Diabetic Retinopathy , the incidence of any retinopathy was 8 % at 3 years, 25 % at 5 years, 60 % at 10 years, and 80 % at 15 years in young-onset diabetic patients. The incidence of Proliferative Diabetic Retinopathy (PDR) was 0 % at 3 years and increased to 25 % at 15 years [59].

More prolonged follow-up cohort studies have confirmed that nearly all individuals with type 1 diabetes will ultimately develop some degree of DR if the disease persists long enough [67].

3.2 Genetic Susceptibility

Genetic susceptibility is a significant determinant of variation in DR expression. There are individuals who develop severe DR in the context of good glycemic control, whereas others do not develop DR in the presence of short disease duration and poor control. Familial clusters studies and large clinical trials like the Diabetes Control and Complications Trial have proven a heritable predisposition to DR in type 1 and type 2 diabetes, after adjusting for conventional risk factors [67].

DR is believed to be a polygenic disorder. Certain gene polymorphisms have been associated with susceptibility to DR [69].

3.3 Hyperglycemia

Hyperglycemia is the most important modifiable risk factor in DR progression. The trials demonstrated that tight glycemic control reduced the incidence and progression of DR in type 1 diabetic patients. The U.K. Prospective Diabetes Study also demonstrated that tighter control of glycemia in type 2 diabetes reduced the risk of retinopathy [59].

Chronic hyperglycemia leads to advanced glycation end-product production, inducing oxidative stress and damage to retinal cells. Increased HbA1c has a close association with DR progression . More recently, studies have also shown glycemic variability, postprandial glucose spikes to be another contributing factor in DR progression in type 2 diabetes [70]. Therefore, both management of long-term glycemic control and normalization of daily glucose fluctuation are required to prevent DR.

3.4 Hypertension

Hypertension is primarily responsible both for the development and evolution of DR and DME. In a randomized controlled trial involving 1,148 type 2 diabetic patients with hypertension, those who received tighter control of blood pressure (150/85 mmHg) compared to less intensive control (180/105 mmHg) were 34 % less likely to have progression of DR and 47 % less likely to lose visual acuity [59].

4 Classification of Diabetic Retinopathy

Diabetic retinopathy (DR) is classified into different grades according to the severity of retinal damage observed during fundus examination. Grading ranges from the early changes limited to the retina's microvasculature to advanced stages with new abnormal vessel growth.

DR progresses through Non-Proliferative Diabetic Retinopathy (NPDR) stages, varying from mild to severe, and may further progress to proliferative diabetic retinopathy (PDR). At the same time, diabetic macular edema (DME), a condition of swelling in the macula due to fluid accumulation, may occur at any stage and is the most common cause of visual impairment [59].

4.1 Non-Proliferative Diabetic Retinopathy (NPDR)

Non-proliferative diabetic retinopathy is the early stage of DR. In NPDR, damage is restricted to the retinal microvasculature and does not extend beyond the inner limiting membrane of the retina [68].

The hallmark findings of NPDR are Microaneurysms (focal capillary dilatation),Dot and blot hemorrhages,Retinal edema,Hard exudates,Cotton-wool spots, and Capillary non-perfusion or occlusion [70].

NPDR evolution is typically divided into three sub-stages:

- **Mild NPDR :** Characterized mainly by microaneurysms, the first clinically detectable signs. Microaneurysms with a rise in number and size and hemorrhages and hard exudates and infrequent cotton-wool spots also occur [68].
- **Moderate NPDR :** At this stage, there is an increase in microaneurysm and hemorrhages size and number and venous beading or rosary-like dilatation in a few quadrants of the retina [68].

- **Severe NPDR:** This is diagnosed based on the ”4-2-1 rule,” and it encompasses the presence of one or more of the following symptoms [71]:
  — Severe intraretinal hemorrhages and microaneurysms in all four quadrants.
  — Venous beading in two or more quadrants.
  — Clear intra-retinal microvascular abnormalities (IRMA) in at least one quadrant.

  There is extensive capillary closure at this stage, which makes it susceptible to retinal ischemia and advancement to PDR.

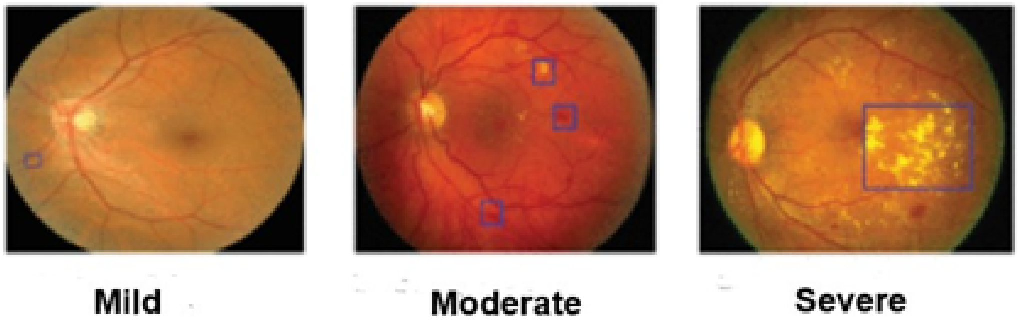


Figure 2.2 – NPDR stages

NPDR is an expression of progressive vascular endothelial injury, secondary to chronic hyperglycemia, which in turn leads to leakage (producing edema), vascular occlusion, and retinal tissue ischemia [70].

4.2 Proliferative Diabetic Retinopathy (PDR)

Proliferative diabetic retinopathy (PDR) is the most advanced stage of diabetic retinopathy. It is marked by neovascularization (the formation of new abnormal blood vessels on the surface of the retina or optic disc), initiated by extensive non perfusion and ischemia of the retinal capillaries [68].

These newly formed vessels are often accompanied by a fibro-glial scaffold, creating extra-retinal fibrovascular proliferation [70].

PDR is typically categorized into three progressive stages :

- **Early PDR:** Presence of neovascularization without significant vitreous hemorrhage [68].
- **High-risk PDR:** Neovascularization on or near the optic disc associated with preretinal or vitreous hemorrhage [68].

- **Advanced PDR:** Involves severe complications such as extensive vitreous hemorrhage, tractional retinal detachment, and neovascular glaucoma [68] .

The contraction of the fibrovascular membrane can exert traction on the retina, resulting in tractional retinal detachment, often accompanied by retinal tears. These late-stage complications define what is known as advanced diabetic eye disease [71].

Moreover, disruption of the blood retinal barrier, together with leakage of inflammatory cytokines and plasma proteins, contributes to the formation of hard exudates visible under fundoscopy. As the disease advances, vasoconstriction and capillary occlusions lead to tortuous vessels and worsening ischemia. Cotton wool spots , representing localized nerve fiber infarctions, may also appear at this stage [70].

In summary, PDR is driven by severe retinal hypoxia , which leads to Pathological neovascularization both within and beyond the retina, Vitreous hemorrhage, Tractional retinal detachment, and severe vision loss if left untreated [69].

4.3 Diabetic Macular Edema (DME)

Diabetic macular edema (DME) is a leading cause of blindness in diabetic individuals. It is characterized by retinal thickening due to accumulation of fluid within the macula, the foveal component of the retina that is essential for sharp central vision [71].

DME can occur at any stage of diabetic retinopathy, both non-proliferative (NPDR) and proliferative (PDR) [63]. It is a result of breakdown of the blood-retinal barrier, leading to leakage of plasma components, including lipids and fluid, into surrounding retinal tissue.

Historically, the majority of interest in diabetic retinopathy has focused on vascular abnormalities, such as intraretinal hemorrhages, microaneurysms, venous beading, and hard exudates. Such features, as observed by ophthalmoscopic findings, are indicative of retinal capillary damage, as supported by histopathologic findings [65]. Over time, this capillary dysfunction and subsequent retinal ischemia are central factors in the pathogenesis of macular edema.

5 Medical Imaging Techniques

5.1 Fundus Photography

Funduscopy is an ophthalmological examination intended to study the structures of the eye behind the lens, and more particularly the retina, optic disc, macula, and blood vessels. It is often performed using a fundus camera, which is a device made of a low power microscope and a digital or film based camera. The optical system resembles that of an indirect ophthalmoscope, enabling magnified and wide angle views (30° to 50°, and up to 140° with auxiliary lenses).

During imaging, light is directed through a series of lenses and reflected onto the retina through the pupil, while the returning light forms a detailed image on a digital sensor or film. Fundus imaging permits clinicians to screen for retinal diseases such as diabetic retinopathy, even in its initial stages. Though digital photography may have slightly lower image quality than traditional film, it is highly efficient, offers instant results, and creates a permanent visual record, making it a preferred method for large scale screening and monitoring [72].

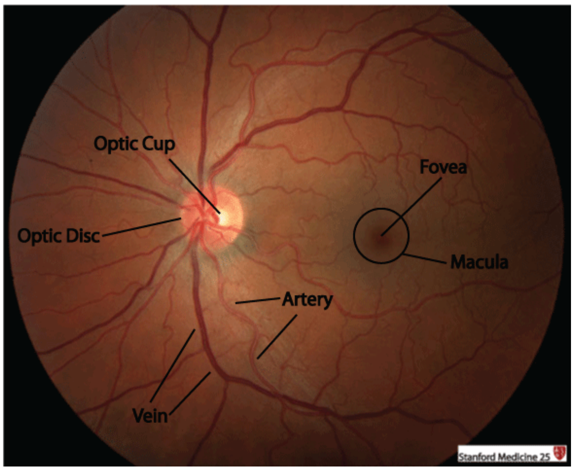


Figure 2.3 – Normal fundus

5.2 Optical Coherence Tomography

Optical coherence tomography (OCT) is a non invasive imaging tests that allows ophtalmologists to examine the retina in cross sectional layers by measuring the depth and thickness. This technique helps in diagnosing multiple eye deseases , including diabetic retinopathy [73].

It is used as a reference standard for diagnosing diabetic macular edema (DME), because fundus photographs alone can miss certain abnormalities (leading to false negatives) or mistakenly detect abnormalities that are not actually present (resulting in false positives) [74].

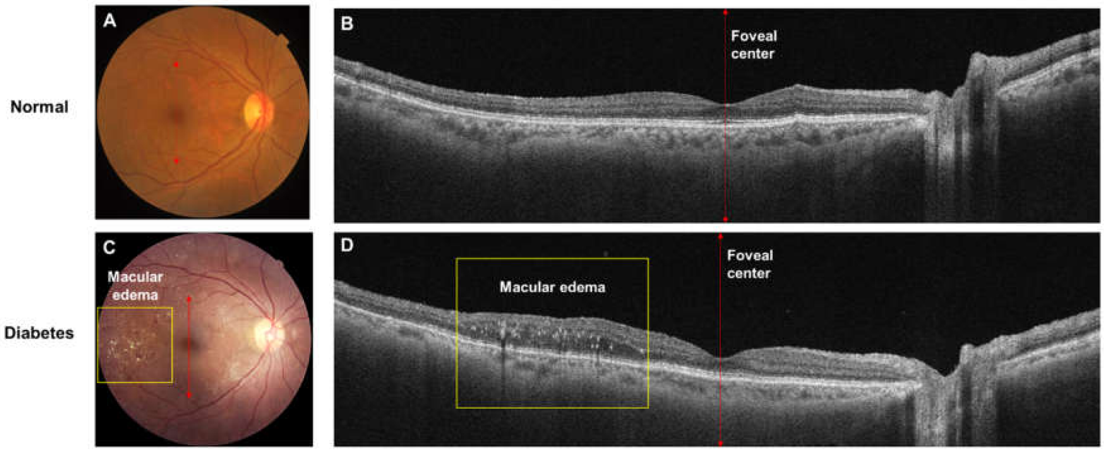


Figure 2.4 – OCT and Fundus images show a normal retina for a healthy patient and retinal damage for a diabetic patient
[75]

5.3 Fluorescein Angiography

Fluorescein Angiography (FA) is a diagnostic imaging test that assesses retinal and choroidal circulation by introducing a fluorescent dye into the circulatory system. FA has been the mainstay of ophthalmology in the identification of vascular lesions in diseases such as diabetic retinopathy and age related macular degeneration. Invasive though it is, FA provides high-resolution dynamic circulation imaging. Recent advancements have witnessed the introduction of widefield imaging and artificial intelligence based analysis. However, newer non invasive tools like OCT angiography are now emerging as viable substitutes [75].

6 Medical Treatement

6.1 Controlling Diabetes

In order to lower the risk of developing diabetic retinopathy for a patient or help slow its progression, it's important to keep his blood sugar, blood pressure, and cholesterol under control. Making healthy lifestyle choices, such as eating well, staying active, and avoiding smoking, can make a big difference. In some cases, medication may also be necessary to keep these levels in check and protect his eyes over the long term [76].

6.2 Intra-vitreal injections of anti-vascular endothelial growth factor(VEGF) agents

In more advanced stages of the disease, substances are injected into the vitreous body to reduce macular edema, prevent or slow the progression to proliferative retinopathy, and improve or stabilize vision. These anti-Vascular Endothelial Growth Factor (VEGF) work by blocking a protein that stimulates abnormal blood vessel growth and fluid leakage in the retina [77].

6.3 Intravitreal corticosteroid injections

Intravitreal corticosteroid injections can be administered in liquid form, with a faster but shorter lasting effect ,or as a biodegradable implant that gradually releases the medication over several months [78], they are mainly used to reduce macular edema.

In diabetic retinopathy, diabetic macular edema (DME) can occur when damaged blood vessels leak fluid, Corticosteroids help reduce inflammation and reduce this leakage, which can stabilize or improve vision [79].

6.4 Laser treatment

This procedure uses laser light to stop the growth of abnormal new blood vessels, helping to halt neovascularization and prevent further vision loss [80]. A study using a rabbit model of retinal ischemia showed that photocoagulation helped reduce VEGF levels, vascular leakage, and abnormal blood vessel growth caused by ischemia [81].

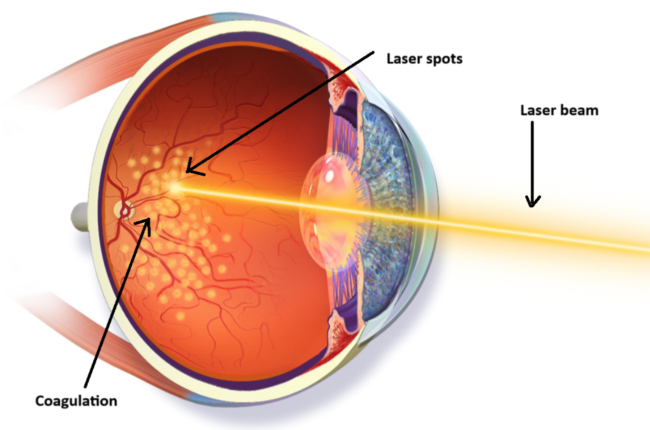


Figure 2.5 – Laser Photocoagulation

6.5 Surgical treatment

It is a type of eye surgery used to manage retinal diseases, known as vitrectomy, involves removing the vitreous humor(wich is a clear gel like substance that fills the inside of the eye. It helps the eye keep its round shape by gently pressing against the retina) from the eye and replace it with a clear solution [82]. This procedure is typically used to treat severe complications of diabetic retinopathy, such as vitreous hemorrhage or retinal detachment [83].

7 Advantages of AI for diagnostic assistance

- Faster, Cheaper DR Diagnosis :Automated methods for detecting diabetic retinopathy (DR) provide a quicker, more reliable, and more affordable alternative to traditional manual diagnosis. Manual methods can be slow, require a lot of effort from specialists, and sometimes lead to interpretation errors [84].

- High Diagnostic Accuracy :CNNs have shown impressive results in identifying and classifying DR signs in retinal images. underlining the strong potential of deep learning tools in helping doctors make better and faster decisions in real world clinical settings [85].

- Early Detection and Grading :DL models can identify early signs of DR like microaneurysms, this makes early intervention possible.A study developed a CNN that achieved a sensitivity of 98% and specificity above 94% in early stage detection [86].

- Reduction of Screening Costs and Improved Accessibility: Using a deep learning based automated tool to assess diabetic retinopathy from color fundus images can help reduce the risk of misdiagnosis and improve the overall diagnostic workflow. Such tools also offer significant advantages, including lower screening costs, better access to healthcare services, and the possibility of initiating treatment earlier [87].

- Integration with Mobile and Real-Time Applications: DL models can be integrated with electronic health records and mobile applications, thereby enhancing the practicality and reach of DR screening programs which helps facilitate point of care diagnostics and continuous monitoring [88].

8 Conclusion

In the present chapter, we have given a clear and comprehensive overview of diabetic retinopathy (DR), one of the main causes of preventable blindness around the world. We have illustrated its pathophysiological mechanisms, risk factors, clinical presentation, and classification based on disease severity. Understanding these elements plays an important role in detecting the disease at an early stage and taking appropriate measures in time.

In the following chapter, we will present the models we developed to detect diabetic retinopathy using deep learning techniques. We give a brief overview of the dataset, the models architecture, and the methods used. Additionally, we perform a comparison between the different models in terms of performance and robustness. This is followed by a detailed analysis of the results and conclusions.

Chapter 3

Implementation and Evaluation Results

1 Introduction

This chapter presents the evaluation of the experimental results gathered during training and evaluation of different deep learning models for diabetic retinopathy classification. The evaluation is conducted on two independent datasets: the APTOS 2019 Blindness Detection dataset and the OLIVES dataset. Our prime objective in this chapter is to compare the performance of the proposed models : CNN, EfficientNetB0, and InceptionResNetV2 for the APTOS 2019 dataset, and EfficientNetB2 and InceptionResNetV2 for the OLIVES dataset ,identify their positives and negatives, and study the impact of varying training techniques, i.e., data augmentation and oversampling.

Throughout the chapter, comparisons are made between models for each of the datasets, revealing which architecture worked best under the same experimental conditions. This evaluation provides insight into what deep learning techniques perform well on multiclass and binary diabetic retinopathy classification tasks.

2 Dataset Description

2.1 APTOS 2019 Blindness Detection

The first data used under this study were retrieved from Kaggle, having been part of the APTOS 2019 Blindness Detection challenge[89]. It comprises a total of 3662 retinal fundus images collected using fundus photography. The images originated from various clinical centers at multiple time intervals using different camera models, resulting in extreme variability when it comes to imaging resolution, lighting, as well as total image quality.

Consequently, some of the images may include noise, blurring, or exposure inhomogeneities, which are the kind of complexities one would expect to encounter in real clinical data [90].

All images have been annotated by a trained medical specialist using a standardized five grade diabetic retinopathy severity scale:

- **0**: No DR .
- **1**: Mild .
- **2**: Moderate .
- **3**: Severe .

- **4**: Proliferative DR .

The Figures 3.1 indicate a clear class imbalance with higher numbers of early stage instances , as represented in the below figure:

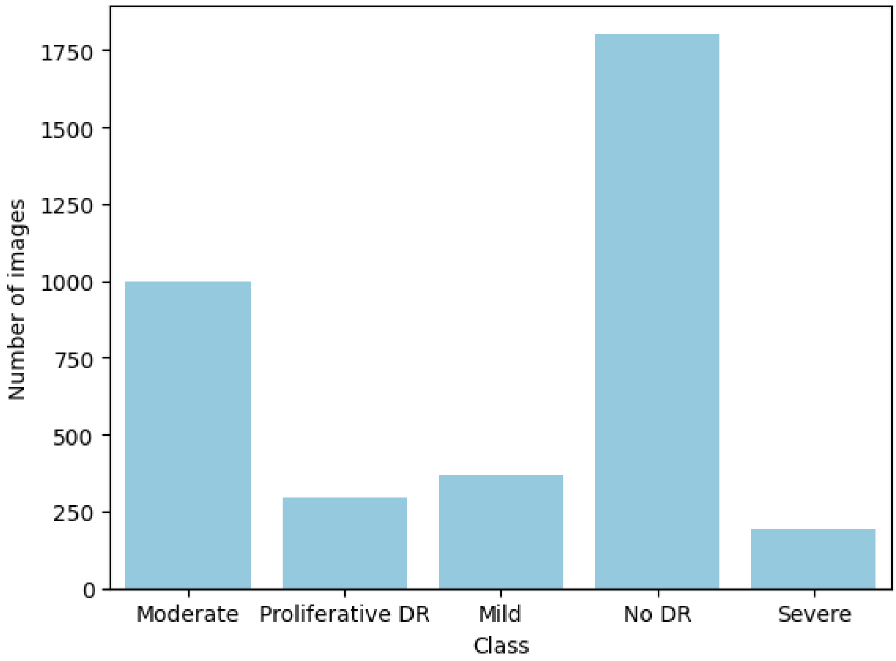


Figure 3.1 – Aptos 2019 Dataset Distribution

2.2 Ophthalmic Labels for Investigating Visual Eye Semantics (OLIVES)

The second dataset used is the OLIVES (Ophthalmic Labels for Investigating Visual Eye Semantics) dataset , made available in the public domain on GitHub [91]. Unlike the vast majority of available datasets that consist of single modality data or address the disease classification task only. OLIVES is a detailed, multimodal representation of the state of the eye over time, making it particularly well adapted to machine learning application in ophthalmology.

OLIVES consists of 1268 near infrared fundus images ,All corresponding to at least 49 OCT scans,16 vectorized biomarkers derived by clinical interpretation,4 clinical labels, And a diagnosis label for the presence of either Diabetic Retinopathy or Diabetic Macular Edema .

These data were collected from 96 eyes, with mean follow up of 66 weeks, and 7 treatment injections per eye for clinical trials. The dataset is meant to facilitate investigation of relationships between different data modalities fundus images, OCT scans, biomarkers, and clinical labels during treatment[92]. This dataset also exhibits a class imbalance between the two diagnostic categories (DR and DME), with one class significantly more represented than the other, as illustrated in the following figure:

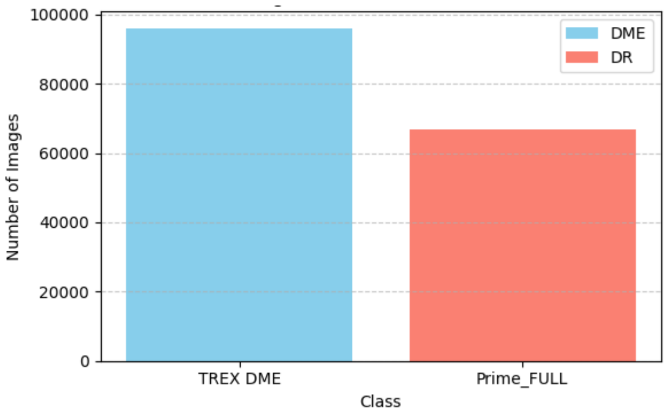


Figure 3.2 – OLIVES Dataset Distribution

3 Data Preparation and Augmentation

3.1 Image Preprocessing

All images, before training, were preprocessed in the standard manner to make them compatible with convolutional neural network architectures. so in our work , each image was resized to 224×224 pixels,to make an input size commonly used in computer vision models. Normalization was also carried out by dividing each pixel value by 255, normalizing pixel intensities to the [0, 1] range. It makes the input data consistent and improves training stability and efficiency, irrespective of the model architecture used .

3.2 Data Augmentation

To make the model stronger and prevent overfitting, we applied several data augmentation techniques . These operations produce synthetic versions of the original images but keep their labels, This enables the model to learn more about generalizing to new cases .

The transformations applied are mentioned in the table below:

Technique	**Description**
`horizontal_flip=True`	Horizontally flips the image to simulate a mirrored view.
`width_shift_range=0.1`	Shifts the image horizontally by up to 10% of its width.
`height_shift_range=0.1`	Shifts the image vertically by up to 10% of its height.
`brightness_range=[0.8, 1.2]`	Adjusts image brightness between 80% and 120% to simulate lighting variations.
`fill_mode='nearest'`	Fills in empty pixels from shifting using the nearest pixel values.

Table 3.1 – Summary of data augmentation techniques applied

These operations were applied only on the training set, but the images of validation were simply normalized and not subjected to geometric transformation.

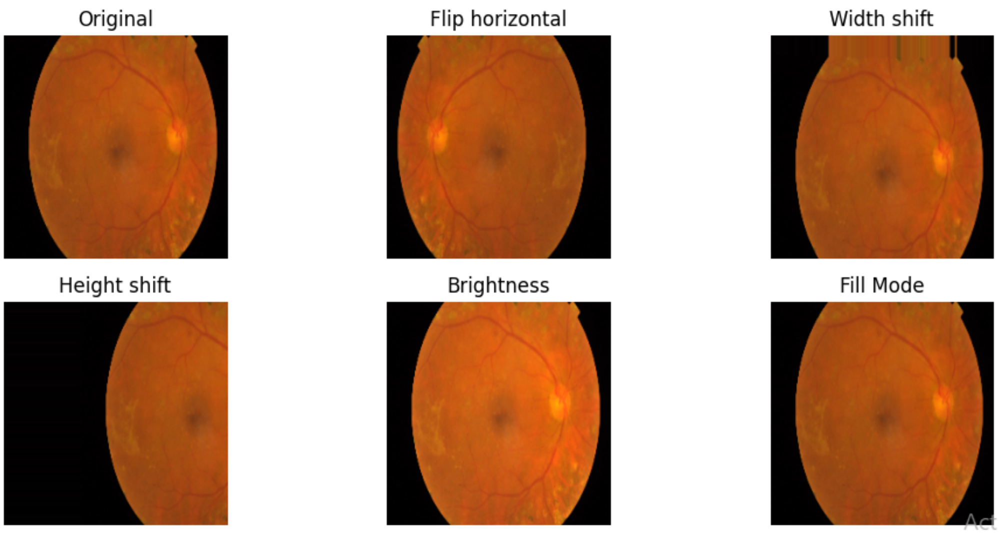


Figure 3.3 – Augmented data for Aptos 2019 Blindness Detection images

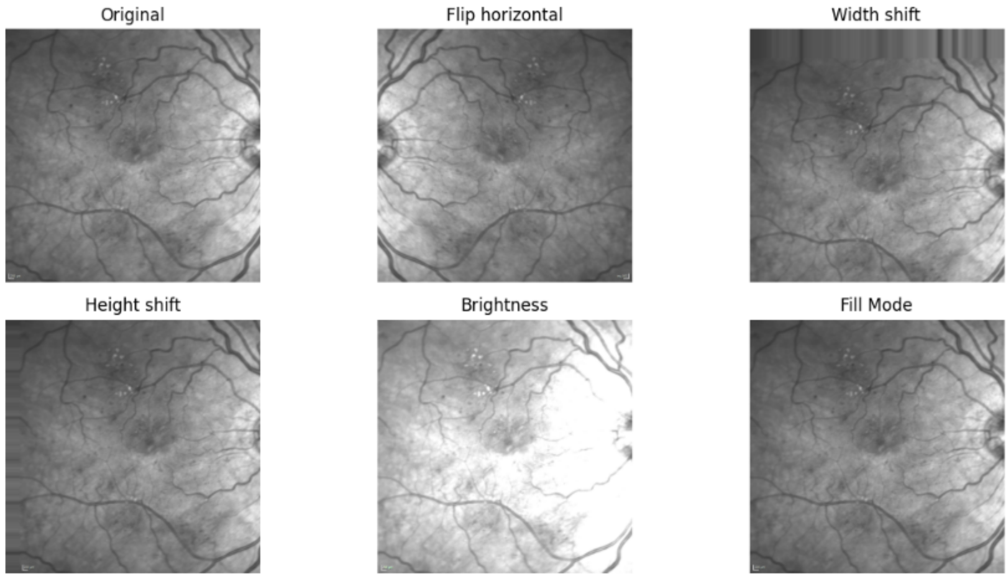


Figure 3.4 – Augmented data for OLIVES fundus images

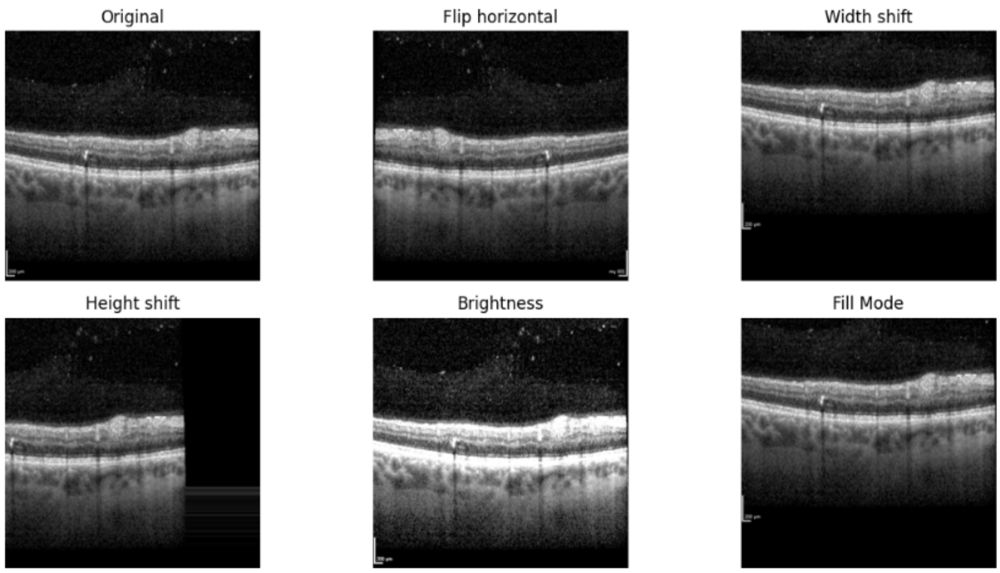


Figure 3.5 – Augmented data for OLIVES OCT images

3.3 Handling Class Imbalance

Both datasets utilized in this work presented a significant class imbalance , In an effort to overcome the undesirable effects of such an imbalance model bias on the majority class and poor generalization on minority classes we used diverse strategies pertinent to each dataset.

3.3.1 Oversampling

For the first dataset Aptos 2019, we handled the issue through oversampling specific to the minority classes: 2 (Mild),4 (Severe), and 5 (Proliferative DR) . We did this by replicating samples of minority classes to artificially increase their frequency in the training set.

In practice, we did this with a factor of 3 such that for every instance of class 2, 4 and 5 we added two more synthetic instances through sampling with replacement. This tripled the number of examples for these classes, bringing them closer in proportion to the majority class. The final training set was then shuffled so that the model would not learn from any sequence patterns created by oversampling.

3.3.2 Class weighting

For the second dataset OLIVES, we employed a different approach with class weighting. Instead of modifying the dataset itself, we computed class weights from the 'balanced' mode of scikit-learn compute_class_weight function . The mode automatically gave higher weight to minority classes and lower weight to majority classes proportionally to the inverse frequency of each class, these weights were then passed on to the training function of the model so that errors on unusual classes would be given more importance in the loss function. This helps to offset bias towards standard classes and forces the model to treat all classes equally despite their respective frequencies being different.

4 Models Architecture

4.1 Model used

As part of this project, we studied and compared some of the convolutional neural network (CNN) architectures in order to identify the best-performing one for our classification.

CNN from scratch: (used for aptos dataset only) We initially built a proprietary CNN architecture from scratch, employing no pre-trained models. This facilitated us to grasp the basics of CNN and have direct control over the number of convolutional layers, filter sizes, activation functions, and other parameters. However, even though this model was entirely tailored to our data, its performance and generalization ability were found to be modest.

EfficientNetB0 : (used for aptos dataset only)

We then experimented with EfficientNetB0, which is one of the pre-trained models that is highly efficient offering a good balance between accuracy and computational cost. We used transfer learning on the features learned on ImageNet and then fine-tuned the model to our task.

EfficientNetB2 : (used for olives dataset only)

In addition to the experiments above, we used EfficientNetB2 for the olives dataset, a more powerful variant of the EfficientNet family compared to B0, it had greater feature extraction capability, which better suited to the complexity of the olives classification task. As with EfficientNetB0, ImageNet-weighted transfer learning was used, followed by task-specific fine-tuning.

InceptionResNetV2 : (used for both datasets)

Lastly, we used InceptionResNetV2, which is a deeper and more rich architecture compared to EfficientNetB0. It is a network that makes use of the strength of both Inception modules and residual connections for effectively learning multi-scale features and maintaining gradient flow in very deep networks. It has proven useful for applications with extensive and hierarchical feature extraction.

4.2 Architectural Modifications and Adaptations

4.2.1 APTOS 2019 Blindness Detection Dataset :

CNN from scratch:

Component	**CNN from Scratch**
Base Model	None – built from scratch
Input Size	224
Conv Blocks	• Conv Block 1: 32 filters, 3×3, ReLU • Conv Block 2: 64 filters, ReLU • Conv Block 3: 128 filters, ReLU • Conv Block 4: 256 filters, ReLU
Pooling	MaxPooling2D after each Conv layer
Dense Layers	• 256 units + ReLU • 128 units + ReLU
Dropout	• Dropout(0.2) after each Dense layer
Output Layer	Dense(5) + softmax

Table 3.2 – Architecture details of CNN from scratch

EfficientNetB0 and InceptionResNetV2:

Component	EfficientNetB0	InceptionResNetV2
Base Model	Pretrained on ImageNet (include_top=False)	
Input Size	224×224 RGB	256x256 RGB
Convolutional Blocks	Handled internally by the base model	
Pooling	GlobalAveragePooling2D	
Dense Layers	• 256 units + ReLU • 128 units + ReLU	
Dropout	Dropout(0.2) after each Dense layer	
Output Layer	Dense(5) + softmax	
Training Strategy	Custom layers trained first, then fine-tune full model	

Table 3.3 – Comparison of architecture components: EfficientNetB0 vs InceptionResNetV2

4.2.2 OLIVES Dataset :

EfficientNetB2:

Component	EfficientNetB2
Base Model	Pretrained on ImageNet (include_top=False)
Input Size	224×224 RGB
Convolutional Blocks	Handled internally by the base model
Pooling	GlobalAveragePooling2D
Dense Layers	• 128 units + ReLU • 64 units + ReLU
Dropout	Dropout(0.2) after first Dense layer Dropout(0.3) after second Dense layer
Output Layer	Dense(2) + softmax
Training Strategy	Train custom layers first (5 epochs) Then fine-tune last 100 layers (20 epochs)

Table 3.4 – Model architecture and training strategy for EfficientNetB2

InceptionResNetV2:

Component	**InceptionResNetV2**
Base Model	Pretrained on ImageNet (include_top=False)
Input Size	224×224 RGB
Convolutional Blocks	Handled internally by the base model
Pooling	GlobalAveragePooling2D
Dense Layers	Dense(256) + ReLU
Dropout	Dropout(0.3) after the dense layer
Output Layer	Dense(2) + softmax
Training Strategy	• Train custom head for 5 epochs • Fine-tune the last 100 layers for 10 epochs

Table 3.5 – Model architecture and training strategy for InceptionResNetV2

5 Training Strategy

5.1 Data Splitting

For both datasets, the data was devided into 90% for training and 10% for validation/test. Splitting was done in a stratified manner to maintain class distribution over subsets .

5.2 Hyperparameters

5.2.1 Loss Function

For the APTOS dataset, we used a proprietary categorical focal loss with gamma = 1.5 for EfficientNetB0 and gamma = 2.0 for InceptionResNetV2 with alpha = 0.25, emphasizing more challenging to classify instances.

On the OLIVES dataset, we used a regular categorical cross-entropy loss .

5.2.2 Optimizer

The two models were optimized using the Adam optimizer due to its adaptive learning rate and quick convergence. It was used in initial training and fine tuning.

5.2.3 Training Configuration

In the APTOS dataset, initial training involved learning rate 1e-4 with batch size of 32 for 5 epochs and subsequent fine tuning through learning rate 1e-5 for the remaining 30 epochs for EfficientNetB0 ,20 epochs for InceptionResNetV2, and 25 epochs for CNN. In OLIVES dataset, the initial training also began with learning rate of 1e-4, but batch size of 16 in EfficientNetB2, and batch size of 32 in InceptionResNetV2 for 5 epochs in the first training phase. The fine tuning was with learning rate of 5e-5 for 20 epochs.

5.3 Training Techniques

To aid generalization and stabilize the learning, both data sets used the identical training procedures.

- **ReduceLROnPlateau** decreased the learning rate when performance plateaued by a factor of 0.5 for APTOS and 0.2 for OLIVES.
- **ModelCheckpoint** was also applied on both environments to save best performing model weights during training .

These methods combined to provide enhanced convergence, less overfitting, and improved performance on the validation sets.

5.4 Evaluation Metrics

For measuring the model's performance, we used various evaluation metrics. At both training and validation, we monitored accuracy and loss to evaluate global learning behavior and detect signs of overfitting.

In order to have a more precise performance per class, especially with class imbalance, we also computed the precision, recall, and F1-score on the validation set. These metrics inform us about how good the model is at discriminating between individual class, especially the minority class. The results were then reported through a confusion matrix to further analyze prediction errors .

6 Implementation Environment

6.1 Hardware used

The development and training of the model were carried out using two different environments:

— **Local machine**:
  — **Processor**: AMD Ryzen 7 5800X3D
  — **Graphics Card**: AMD Radeon RX 6800 XT

  - **Memory**: 32 GB RAM
- **Cloud environment**:
  Platform: Kaggle Notebooks, which provides free access to GPU resources , we used GPU T4 x 2 for EfficientNet and GPU P100 for InceptionResNetV2.

6.2 Software Stack

- **Python distribution**: Anaconda
- **Python version**: Python 3.11.4

6.3 Deep learning libraries

TensorFlow / Keras: Keras is high-level API in the TensorFlow platform. It offers a simple and effective interface for solving machine learning tasks, particularly deep learning. This API supports the entire ML workflow from data preprocessing to model training, tuning, and deployment [93].

Other libraries:

- `NumPy`: NumPy is a fundamental Python library for scientific computing. It provides tools for arrays with dimensions and has a variety of functions that are able to perform math, statistics, data sorting, etc., quickly and effectively [94].
- `Pandas`: Pandas is a free Python library that offers fast, robust, and flexible tools for data analysis and manipulation [95].
- `Matplotlib`: Matplotlib is a Python library that creates various types of plots, from simple charts to more complex, animated, or interactive plots [96].
- `scikit-learn`: Scikit-learn is an open source library providing simple and high performance predictive data analysis tools, built on NumPy, SciPy, and Matplotlib, designed to be convenient to use ,reusable, and commercially usable under the BSD license [97].

7 Results and Discussion

7.1 Aptos 2019 Blindness detection datase

7.1.1 Results without Data Augmentation and Oversampling

To establish the baseline performance of our models and estimate the difficulty of training, we initially trained our three architectures(CNN from scratch, EfficientNetB0, and InceptionResNetV2) without the use of any data augmentation or oversampling methods. This provides the first step to see how each model performs on the original imbalanced dataset . The following plots are the training and validation accuracy and loss curves of each model:

- **CNN**

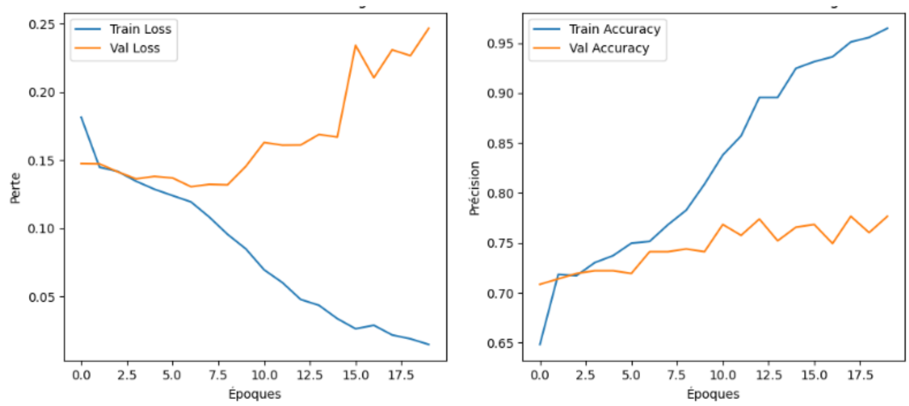


Figure 3.6 – Training and validation accuracy/loss curves before data augmentation for CNN

- **EfficientNetB0**

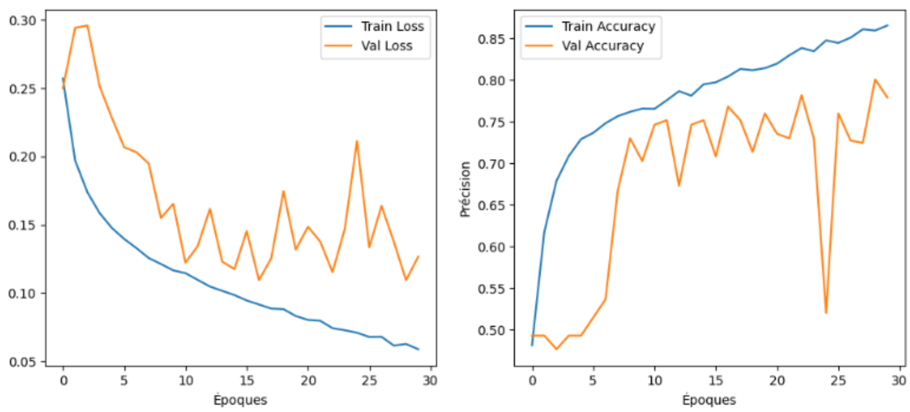


Figure 3.7 – Training and validation accuracy/loss curves before data augmentation for EfficientNetB0

- **InceptionResNetV2**

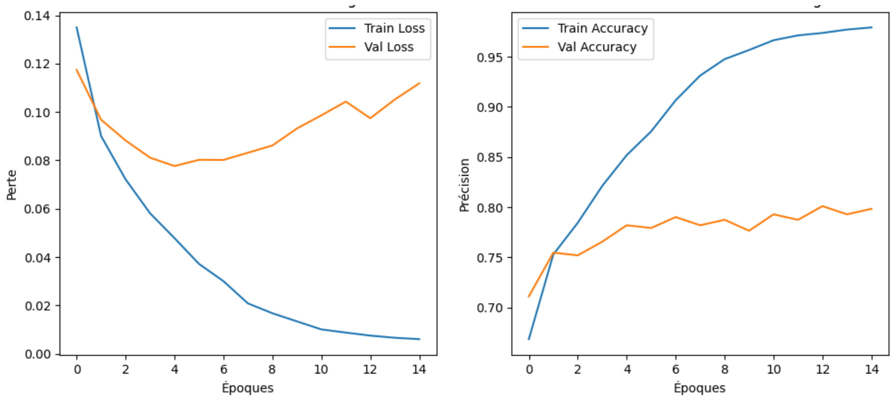


Figure 3.8 – Training and validation accuracy/loss curves before data augmentation for InceptionResNetV2

In all three models, we saw indications of overfitting. Training accuracy kept going up steadily, whereas validation accuracy plateaued , and validation loss started to rise after several epochs. This indicated that the models weren't generalizing too well to new data .

The following are the confusion matrices for the three models:

- **CNN**

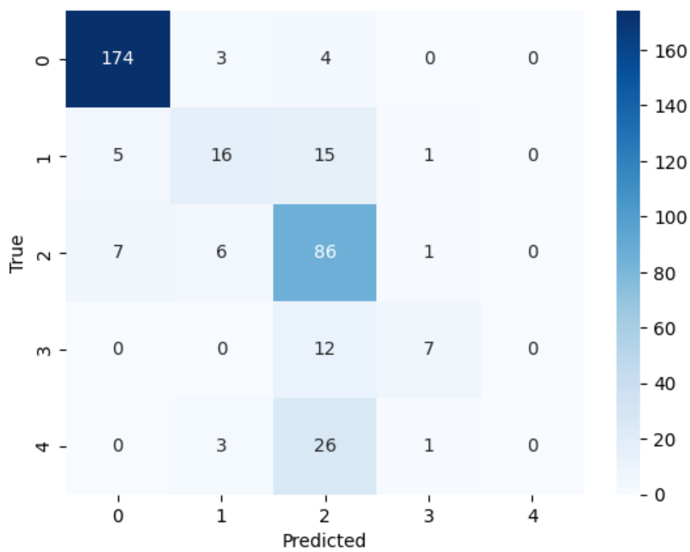


Figure 3.9 – Confusion matrix before oversampling for CNN

- **EfficientNetB0**

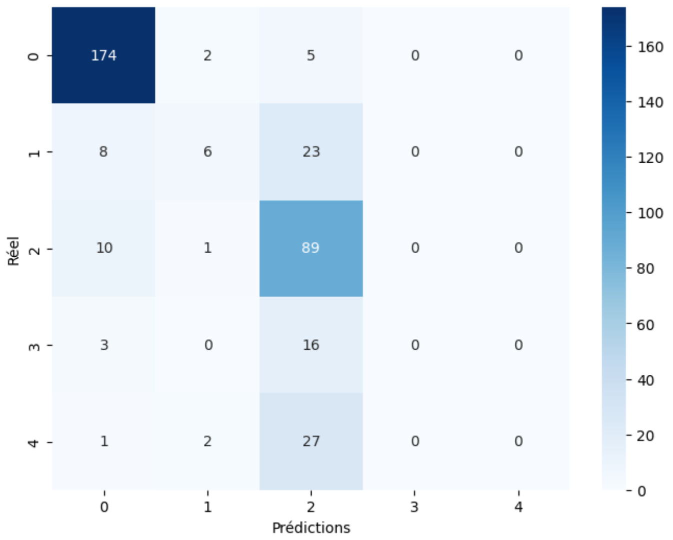


Figure 3.10 – Confusion matrix before oversampling for EfficientNetB0

- **InceptionResNetV2**

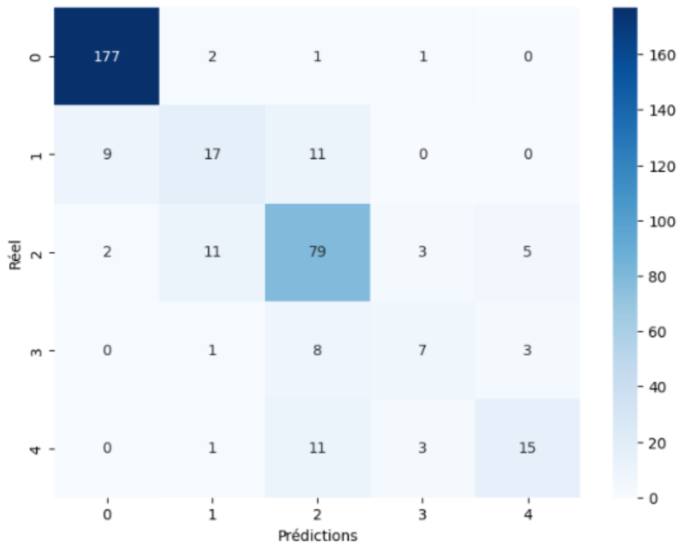


Figure 3.11 – Confusion matrix before oversampling for InceptionResNetV2

These confusion matrices also reveal the class imbalance problem. The models were doing well for the most common class (class 0 and class 2) and consistently misclassifying minority classes like class 1, 3, and 4 or failing to predict them , as observed with classes 3 and 4 in EfficientNetB0 and class 4 in CNN. This is a clear indication that the learning process was trying to learn majority class patterns and was ignoring underrepresented classes.

The f1-score values before using oversampling, for each class of our three models are in the following table:

Class	**CNN f1-score**	**EfficientNetB0 f1-score**	**InceptionResNetV2 f1-score**
0	95%	92%	96%
1	49%	25%	46%
2	71%	68%	75%
3	48%	00%	42%
4	00%	00%	57%

Table 3.6 – F1-scores per class for each model before oversampling

As it can be observed from this comparison, bot CNN and EfficientNetB0 struggled the most with correctly identifying classes 1, 3, and 4. F1-score for them is significantly worse or doesn't exist than for the majority class (class 0 and class 2), which corroborates the fact that class imbalance significantly impacts model performance.

7.1.2 Final Results with Data Augmentation and Oversampling

After data augmentation and oversampling, the performance of ous models was significantly enhanced to mitigate overfitting and class imbalance.

- **CNN**

  Figure 3.12 represents the CNN model's training and validation accuracy and loss curves. The accuracy value on the training set was **76%** while on the validation set was **77%**. Loss dropped to **0.1029** for training and **0.1210** for validation.

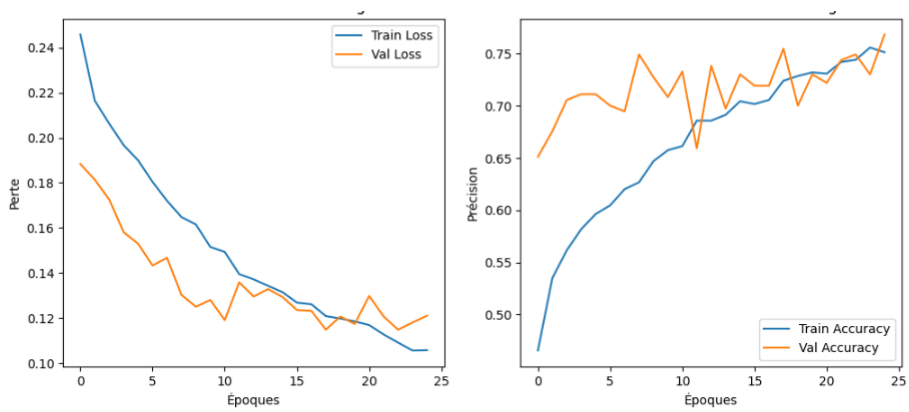


Figure 3.12 – Training and validation accuracy/loss curves for CNN

Figure 3.13 displays the confusion matrix of the CNN model, showing its predictions across the five diabetic retinopathy classes.and the classification report provides the detailed performance metrics for each class.

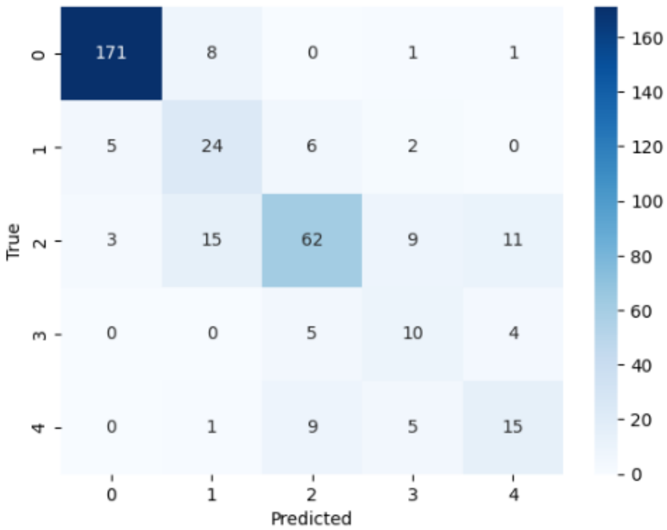


Figure 3.13 – Confusion matrix for CNN

Class	Precision	Recall	F1-score
0	96%	94%	95%
1	50%	65%	56%
2	76%	62%	68%
3	37%	53%	43%
4	48%	50%	49%

Table 3.7 – Classification report for CNN

- **EfficientNetB0**

  Figure 3.14 shows the training and validation loss and accuracy curves over 30 epochs. As can be seen, accuracy kept going up, reaching **82%** for the training set and **81%** for the validation set. The loss also kept dropping consistently, reaching **0.0824** for training and **0.0911** for the validation set.

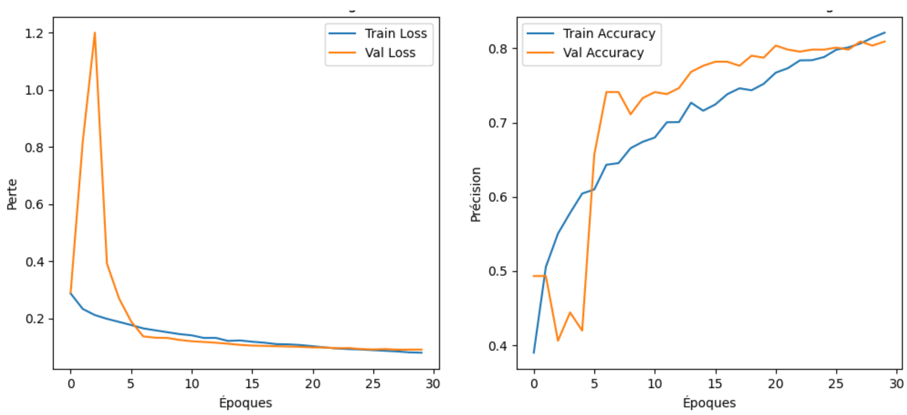


Figure 3.14 – Training and validation accuracy/loss curves for EfficientNetB0

Figure 3.15 also illustrates the confusion matrix, which reveals a relatively well balanced classification for the five classes. The classification report also bears witness to that improvement.

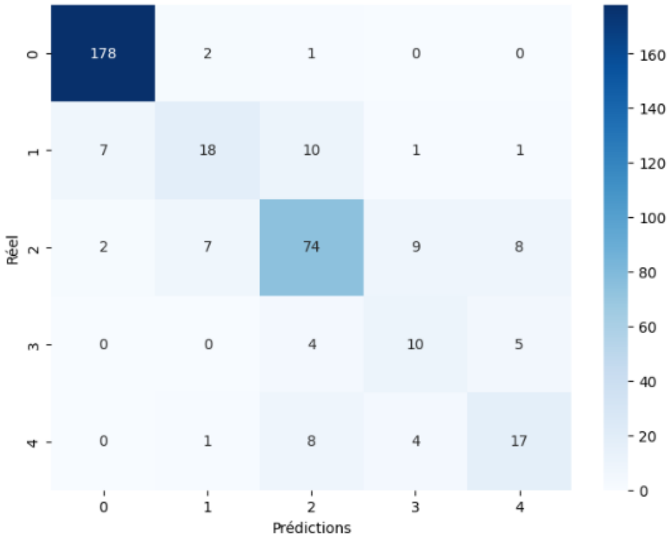


Figure 3.15 – Confusion matrix for EfficientNetB0

Class	**Precision**	**Recall**	**F1-score**
0	95%	98%	97%
1	64%	49%	55%
2	76%	74%	75%
3	42%	53%	47%
4	55%	57%	56%

Table 3.8 – Classification report for EfficientNetB0

- **InceptionResNetV2**

  Figure 3.16 shows the training and validation loss and accuracy curves over 15 epochs for the InceptionResNetV2 model. As shown, the training accuracy steadily increased to reach approximately **94%**, while the validation accuracy improved and stabilized around **81%**. The training loss continuously decreased to below **0.02**, and the validation loss reached a value close to **0.08**, indicating good convergence.

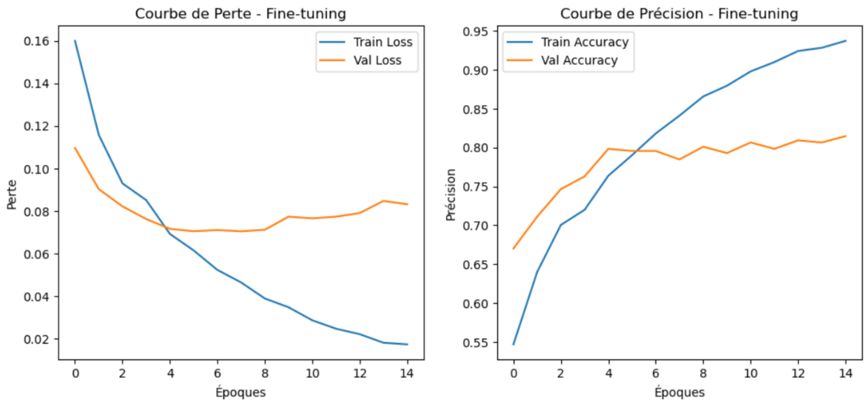


Figure 3.16 – Training and validation accuracy/loss curves for InceptionResNetV2

The confusion matrix illustrated in Figure 3.17 and classification report below also confirm that after oversampling, the model showed a slight increase in F1-scores, especially for the minority classes, confirming the benefit of balancing the dataset.

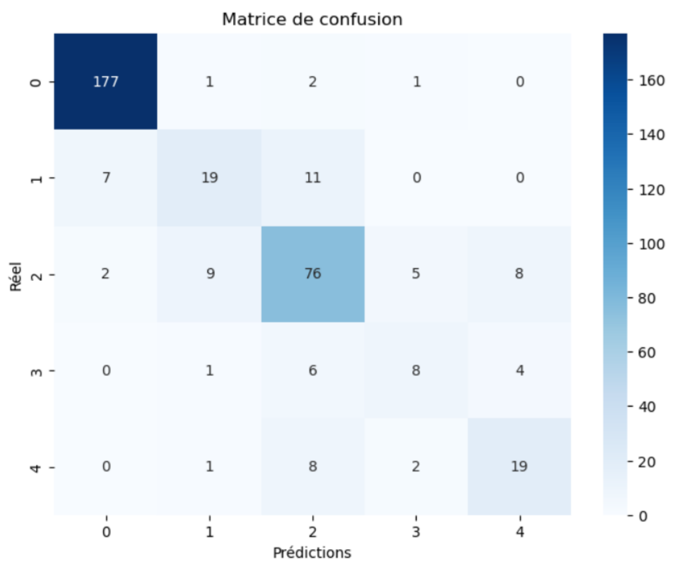


Figure 3.17 – Confusion matrix for InceptionResNetV2

Class	Precision	Recall	F1-score
0	0.95	0.98	0.96
1	0.61	0.51	0.56
2	0.74	0.76	0.75
3	0.50	0.42	0.46
4	0.61	0.63	0.62

Table 3.9 – Classification report for InceptionResNetV2

7.2 OLIVES dataset

7.2.1 Results without Data Augmentation and classweight

- **EfficientNetB2**

  Figure 3.18 presents the training and validation accuracy and loss of the EfficientNetB2 model before applying data augmentation and class weighting. The training accuracy increased steadily, reaching over 97%, while the validation accuracy fluctuated throughout the 20 epochs. The training loss decreased continuously, but the validation loss showed instability and peaked at several points, indicating poor generalization and overfitting.

Figure 3.18 – Training and validation accuracy/loss curves for EfficientNetB2

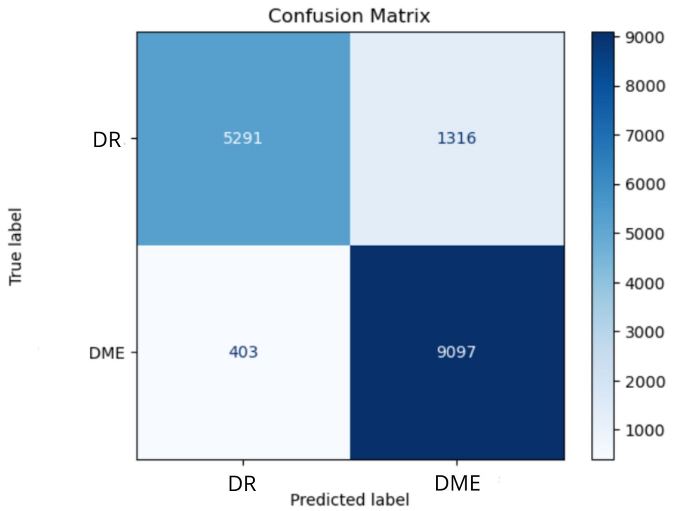


Figure 3.19 – Confusion matrix for EfficientNetB2

Class	Precision	Recall	F1-score
DR	0.93	0.80	0.86
DME	0.87	0.96	0.91

Table 3.10 – Classification report for EfficientNetB2

- **InceptionResNetV2**

  Figure 3.20 presents the training and validation accuracy and loss of the InceptionResNetV2 model. The training accuracy rises sharply, reaching 99% by the fifth epoch, while the validation accuracy stagnates around 84–88%, indicating a significant generalization gap. The loss curves further confirm this issue: while the training loss decreases steadily (approaching 0), the validation loss increases. This rise in validation loss is a classic sign of overfitting.

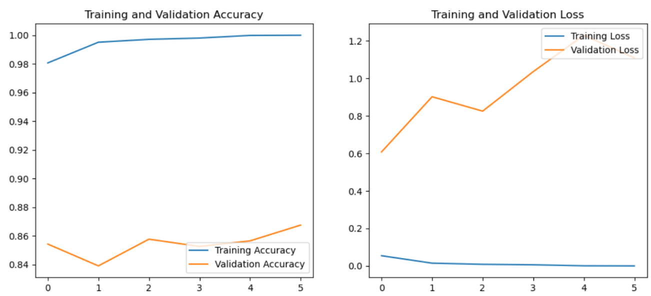


Figure 3.20 – Training and validation accuracy/loss curves for InceptionResNetV2

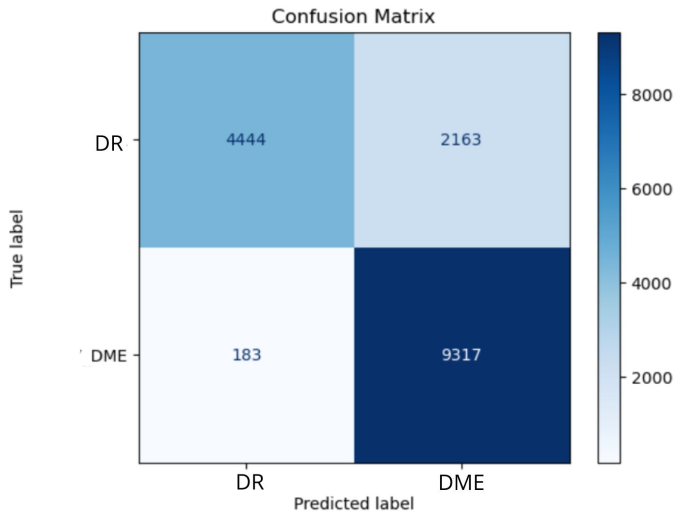


Figure 3.21 – Confusion matrix for InceptionResNetV2

Class	Precision	Recall	F1-score
DR	0.96	0.67	0.79
DME	0.81	0.98	0.89

Table 3.11 – Classification report for EfficientNetB2

7.2.2 Final Results with Data Augmentation and classweight

- **EfficientNetB2**

  Figure 3.22 presents the training and validation accuracy and loss of the EfficientNetB2 model. The accuracy improved consistently for the 20 epochs to **93%** training and **92%** validation. Training loss fell to **0.1624**, and validation loss reached **0.2046**.

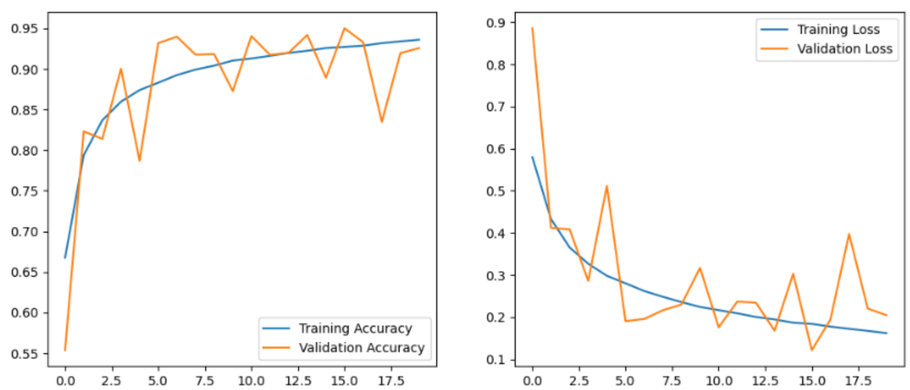


Figure 3.22 – Training and validation accuracy/loss curves for EfficientNetB2

Figure 3.23 presents the corresponding confusion matrix, indicating the good predictions of the model between DR class and DME class.

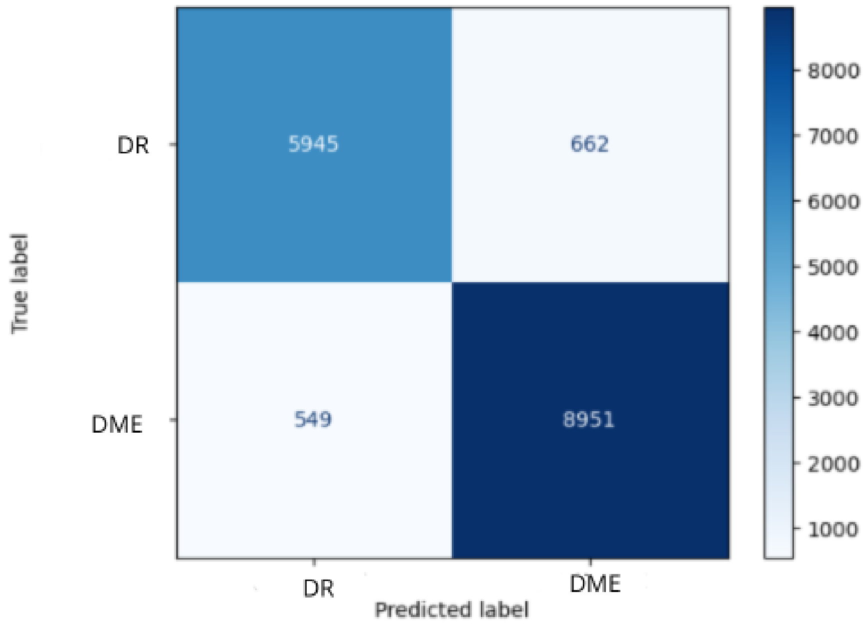


Figure 3.23 – Confusion matrix for EfficientNetB2

Class	Precision	Recall	F1-score
DR	0.92%	0.90%	0.91%
DME	0.93%	0.94%	0.94%

Table 3.12 – Classification report for EfficientNetB2

- **InceptionResNetV2**

  Figure 3.24 displays the training and validation curves of the InceptionResNetV2 model after applying data augmentation and class weighting. While the training accuracy continues to improve and reaches nearly **100%**, the validation accuracy remains low and fluctuates slightly between **84%** and **87%**. Moreover, the validation loss increases significantly across epochs, while the training loss converges close to zero.

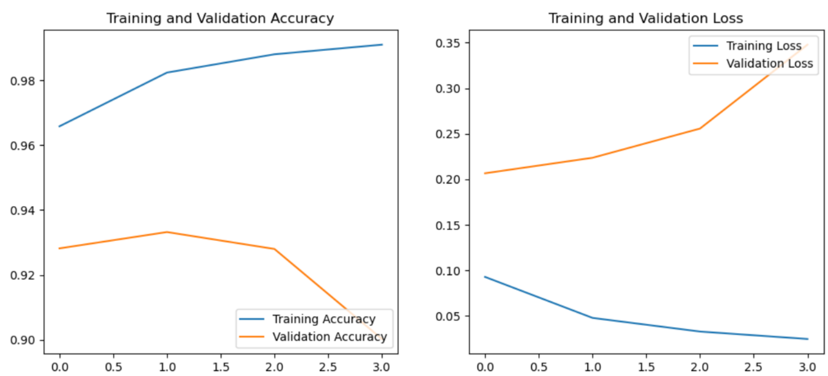


Figure 3.24 – Training and validation accuracy/loss curves for InceptionResNetV2

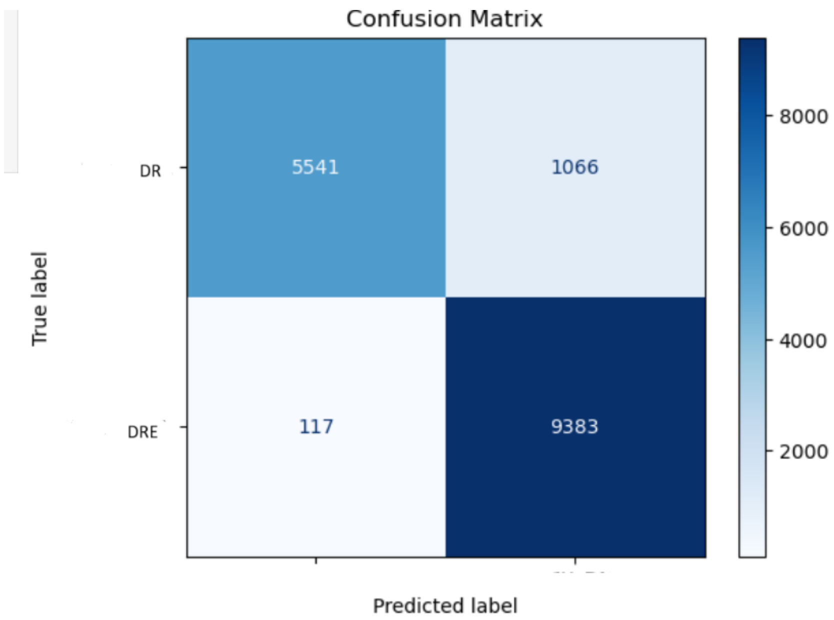


Figure 3.25 – Confusion matrix for InceptionResNetV2

Class	Precision	Recall	F1-score
DR	0.98%	0.84%	0.90%
DME	0.90%	0.99%	0.94%

Table 3.13 – Classification report for InceptionResNetV2

7.3 Discussion

7.3.1 Aptos 2019 Blindness detection database

-Initially, training without any augmentation or rebalancing techniques revealed a clear tendency toward overfitting, particularly for the CNN developed from scratch. The learning curves show a continuous increase in training accuracy, while the validation accuracy quickly plateaued at 77%, accompanied by a rise in validation loss.

-In addition, the confusion matrices showed a clear bias of the models toward the majority classes (notably classes 0 and 2), while minority classes (1, 3, and 4) were frequently misclassified, or even completely ignored. This issue was reflected in the F1-scores, which were very low or even zero for these classes in the case of CNN and EfficientNetB0.

-A notable exception was observed with InceptionResNetV2, which, even without oversampling, was able to detect some of the minority classes, in particular class 4 with a score of 57%, something that none of the other two models was able to achieve. This reflects a higher inherent robustness of InceptionResNetV2 in dealing with unbalanced data, likely due to its depth and its ability to learn multi-scale features.

-Following this observation, data augmentation was effective in generalizing as can be seen from the stabilization of the validation curves and the gap closure between the training and validation performance. However, the class imbalance problem persisted, which justified the application of oversampling — a strategy that allowed better representation of the under-represented classes. This improvement is visible not only in the confusion matrices, but also in the more balanced distribution of F1-scores across the different classes

-Although the InceptionResNetV2 model had a strong ability to identify some minority classes before rebalancing, its performance in the final stage exhibited a significant overfitting behavior, as seen from the learning curves. It became less effective in generalizing well on new data, despite achieving a good end accuracy of 81.47%, which was only a slight improvement from its initial performance (79.84%). On the other hand, EfficientNetB0 showed greater enhancement, from 73% to 81% accuracy, by having more stable and robust training and validation learning curves, signifying better ability for generalization. Such results show that EfficientNetB0 has the best performance-stability-robustness trade-off and is therefore more suitable for automated clinical use in real-world practice — especially when there is an imbalanced data scenario.

7.3.2 OLIVES database

In the initial training on the OLIVES dataset, regularization techniques were not applied ,no class weighting and no data augmentation. Under these conditions, we observed a severe case of overfitting, reflected in the divergence of training and validation loss and accuracy curves .The EfficientNetB2 model achieved high accuracy on the training set but performed poorly on the validation data . This is a sign of biased learning with the model memorizing training examples rather than learning to generalize. The same pattern was observed for the InceptionResNetV2 model .

The imbalance had a direct impact on the performance metrics. For every EfficientNetB2, F1-scores were 81% for DR and 91% for DME, indicating the model bias toward the majority class . InceptionResNetV2 was similar, F1-scores being 79% for DR and 89% for DME. These imbalances indicate that the models will underrepresent the minority class, which in medical applications can be particularly problematic given the relevance of identifying rare cases .

Class weights were introduced to balance the learning process. a weight of 1.22 on the minority class (DR) and a weight of 0.85 on the majority class (DME) were assigned according to class frequency.

The effect of these remedial actions was clearly reflected in the results obtained after fine-tuning. EfficientNetB2 achieved **93%** validation accuracy and **94%** in train, with 91% and 94% F1-scores for DR and DME, respectively. Validation loss was significantly better at 0.2046 from 0.693 , indicating better model convergence and excellent reduction of overfitting . Moreover, loss and accuracy curves were better aligned between the training and validation datasets, indicating better generalization capability. The model was better balanced in prediction and more stable in classifying both classes .

InceptionResNetV2 also finished training with an accuracy of **99%** and **90%** in validation, and validation loss of 0.3478 and train loss of 0.0253. The F1-scores showed a more balanced classification in both classes with 90% in DR and 94% in DME . A slight instability in the validation curves was observed from the epoch 3, however, which is a sign of risk of overfitting in the later training epochs.

Comparing both architectures, EfficientNetB2 was the more stable and balanced model on all the metrics considered in the evaluation.

8 Conclusion

This chapter described training process, evaluation, and performance comparison of various deep learning models for diabetic retinopathy classification. It highlighted challenges such as overfitting and class imbalance, which were addressed using data augmentation, oversampling, and class weighting. EfficientNetB0 and EfficientNetB2 were the most stable and robust models for the multiclass and binary tasks, respectively. InceptionResNetV2 showed potential but was limited by overfitting thereafter. These results highlight the value of adapting training strategies and carefully choosing models to develop reliable and clinically relevant diagnostic systems.

General Conclusion

This thesis explored the application of deep learning techniques for automatic diagnosis of diabetic retinopathy (DR) from both color fundus photographs and multimodal ophthalmic data. The study was conducted in two stages: stage one on the APTOS 2019 dataset, which was a multiclass classification task (five levels of severity of DR), and stage two on the OLIVES dataset, which was a binary classification task between DR and Diabetic Macular Edema (DME).

For APTOS, three convolutional neural networks were compared: a CNN built from scratch, EfficientNetB0 model, and InceptionResNetV2 model. Initially, all models suffered from overfitting and poor generalization, especially in classifying minority classes such as Severe and Proliferative DR. In response to these limitations, data augmentation and oversampling were used. These interventions achieved significant improvements, particularly for EfficientNetB0, where it achieved the best trade-off between performance and generalization with the final accuracy of 81% . Although InceptionResNetV2 did have some capacity to identify minority classes before rebalancing, it displayed signs of overfitting in later stages.

In the second part of the study, using the OLIVES dataset, two transfer learning models were tested: EfficientNetB2 and InceptionResNetV2. The new challenge of multimodal images (fundus + OCT) and class imbalance between DR and DME was encountered in this dataset. EfficientNetB2 was the most stable and generalizable model, achieving validation accuracy of 92%. Although InceptionResNetV2 had slightly higher raw accuracy, it was less stable and overfitted. EfficientNetB2 was thus the optimal model for binary classification under clinical conditions.

Overall, this work confirms the value of deep learning as a tool for DR diagnosis. It validates the significance of selecting appropriate models and training approches based on the type of dataset—particularly in handling class imbalance and preventing overfitting. The results demonstrate that EfficientNet-based architectures, combined with tailored data preparation strategies, offer robust, accurate, and clinically viable solutions for both multiclass and binary DR classification tasks.

To further improve results, it would be interesting to explore other techniques as ensemble models, adding clinical metadata, and expanding the datasets to include a broader and more diverse population. These enhancements would further improve diagnostic consistency and support real-world deployment in teleophthalmology and clinical screening programs.

In conclusion, this research provides a promising step toward intelligent and accessible DR diagnosis systems, with the potential to contribute significantly to early detection and treatment in real-world medical settings.

Bibliography

[1] Pramila P. Shinde and Seema Shah. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–6, 2018.

[2] Mohsen Soori, Behrooz Arezoo, and Roza Dastres. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*, 3:54–70, 2023.

[3] Lokman Balyen and Tunde Peto. Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. *The Asia-Pacific Journal of Ophthalmology*, 8(3):264–272, 2019.

[4] Md Hosne Mobarak, Mariam Akter Mimona, Md Aminul Islam, Nayem Hossain, Fatema Tuz Zohura, Ibnul Imtiaz, and Md Israfil Hossain Rimon. Scope of machine learning in materials research—a review. *Applied Surface Science Advances*, 18:100523, 2023.

[5] Mamatha Bhat, Madhumitha Rabindranath, Beatriz Sordi Chara, and Douglas A Simonetto. Artificial intelligence, machine learning, and deep learning in liver transplantation. *Journal of hepatology*, 78(6):1216–1233, 2023.

[6] V. Kulkarni, M. Kulkarni, and A. Pant. Quantum computing methods for supervised learning. *Quantum Machine Intelligence*, 3(2), 2021.

[7] Samreen Naeem, Aqib Ali, Sania Anam, and Muhammad Munawar Ahmed. An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*, 2023.

[8] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic markets*, 31(3):685–695, 2021.

[9] Jian Liang, Ran He, and Tieniu Tan. A comprehensive survey on test-time

adaptation under distribution shifts. *International Journal of Computer Vision*, 133(1):31–64, 2025.

[10] Yutaka Matsuo, Yann LeCun, Maneesh Sahani, Doina Precup, David Silver, Masashi Sugiyama, Eiji Uchibe, and Jun Morimoto. Deep learning, reinforcement learning, and world models. *Neural Networks*, 152:267–275, 2022.

[11] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[12] Mohsen Soori, Behrooz Arezoo, and Roza Dastres. Artificial neural networks in supply chain management, a review. *Journal of Economy and Technology*, 1:179–196, 2023.

[13] Maher GM Abdolrasol, SM Suhail Hussain, Taha Selim Ustun, Mahidur R Sarker, Mahammad A Hannan, Ramizi Mohamed, Jamal Abd Ali, Saad Mekhilef, and Abdalrhman Milad. Artificial neural networks based optimization techniques: A review. *Electronics*, 10(21):2689, 2021.

[14] Roza Dastres and Mohsen Soori. Artificial neural network systems. *International Journal of Imaging and Robotics (IJIR)*, 21(2):13–25, 2021.

[15] Neeratyoy Mallik, Edward Bergman, Carl Hvarfner, Danny Stoll, Maciej Janowski, Marius Lindauer, Luigi Nardi, and Frank Hutter. Priorband: Practical hyperparameter optimization in the age of deep learning. *Advances in Neural Information Processing Systems*, 36:7377–7391, 2023.

[16] Ayman H Abd-elaziem and Tamer HM Soliman. A multi-layer perceptron (mlp) neural networks for stellar classification: A review of methods and results. *International Journal of Advances in Applied Computational Intelligence*, 3(10.54216), 2023.

[17] Walter H Delashmit, Michael T Manry, et al. Recent developments in multilayer perceptron neural networks. In *Proceedings of the seventh annual memphis area engineering and science conference, MAESC*, volume 7, page 33, 2005.

[18] Rahman Ali, Jamil Hussain, and Seung Lee. Multilayer perceptron-based self-care early prediction of children with disabilities. *Digital health*, 9:20552076231184054, 06 2023.

[19] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.

[20] Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic markets*, 31(3):685–695, 2021.

[21] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9:611–629, 2018.

[22] Hiromu Yakura, Shinnosuke Shinozaki, Reon Nishimura, Yoshihiro Oyama, and Jun Sakuma. Malware analysis of imaged binary samples by convolutional neural network with attention mechanism. pages 127–134, 03 2018.

[23] Keiron O'shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[24] Susmita Das, Amara Tariq, Thiago Santos, Sai Sandeep Kantareddy, and Imon Banerjee. Recurrent neural networks (rnns): architectures, training tricks, and introduction to influential research. *Machine learning for Brain disorders*, pages 117–138, 2023.

[25] Ibomoiye Domor Mienye, Theo G Swart, and George Obaido. Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9):517, 2024.

[26] Saidul Islam, Hanae Elmekki, Ahmed Elsebai, Jamal Bentahar, Nagat Drawel, Gaith Rjoub, and Witold Pedrycz. A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, 241:122666, 2024.

[27] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers. *AI Open*, 3:111–132, 2022.

[28] Pragati Baheti. Activation functions in neural networks [12 types use cases], march 2021. Accessed on March 10, 2025.

[29] Nagesh Singh Chauhan. Optimization algorithms in neural networks, Dec 2020. Accessed on March 14, 2025.

[30] Nuria Gómez, Luis López, Alberto Albert, and Javier Llamas. Image classification with convolutional neural networks using gulf of maine humpback whale catalog. *Electronics*, 9:731, 04 2020.

[31] Regularization techniques in deep learning: Ultimate guidebook, n.d. Accessed on March 14, 2025.

[32] Anastasiia Safonova, Gohar Ghazaryan, Stefan Stiller, Magdalena Main-Knorn, Claas Nendel, and Masahiro Ryo. Ten deep learning techniques to address small data problems with remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 125:103569, 2023.

[33] LI Xuhong, Yves Grandvalet, and Franck Davoine. Explicit inductive bias for transfer learning with convolutional networks. In *International conference on machine learning*, pages 2825–2834. PMLR, 2018.

[34] Yanan Zhao, Chunshen Long, Wenjing Shang, Zhihao Si, Zhigang Liu, Zhenxing Feng, and Yongchun Zuo. A composite scaling network of efficientnet for improving spatial domain identification performance. *Communications Biology*, 7(1):1567, 2024.

[35] JV Thomas Abraham, A Muralidhar, Kamsundher Sathyarajasekaran, and N Ilakiyaselvan. A deep-learning approach for identifying and classifying digestive diseases. *Symmetry*, 15(2):379, 2023.

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[37] Kshitij Ijari and Carlos D Paternina-Arboleda. Sustainable pavement management: Harnessing advanced machine learning for enhanced road maintenance. *Applied Sciences*, 14(15):6640, 2024.

[38] Nasser Mesfer N Alzahrani. Enhancing image classification performance using inceptionnet: A comparative study.

[39] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[40] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[41] DM Anisuzzaman, Jeffrey G Malins, Paul A Friedman, and Zachi I Attia. Fine-tuning large language models for specialized use cases. *Mayo Clinic Proceedings: Digital Health*, 3(1), 2025.

[42] Abil Elly, Abill Robert, and Abiodun Okunola. The role of ensemble learning in improving efficiency of machine learning algorithms. 12 2024.

[43] Shaoze Cui, Yanzhang Wang, Dujuan Wang, Qian Sai, Ziheng Huang, and TCE Cheng. A two-layer nested heterogeneous ensemble learning predictive method for covid-19 mortality. *Applied Soft Computing*, 113:107946, 2021.

[44] Wendy Flores-Fuentes, Oleg Sergiyenko, Julio C Rodríguez-Quiñonez, and Jesús E Miranda-Vega. Application of information theory to computer vision and image processing, 2024.

[45] Heang-Ping Chan, Ravi K Samala, Lubomir M Hadjiiski, and Chuan Zhou. Deep learning in medical image analysis. *Deep learning in medical image analysis: challenges and applications*, pages 3–21, 2020.

[46] Kuei-Hu Chang. Natural language processing: Recent development and applications, 2023.

[47] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56(2):1–40, 2023.

[48] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

[49] S Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L Prince, Daniel Rueckert, and Ronald M Summers. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5):820–838, 2021.

[50] Zakaria Rguibi, Abdelmajid Hajami, Dya Zitouni, Amine Elqaraoui, and Anas Bedraoui. Cxai: Explaining convolutional neural networks for medical imaging diagnostic. *Electronics*, 11(11):1775, 2022.

[51] Yuxin Wang, Qiang Wang, Shaohuai Shi, Xin He, Zhenheng Tang, Kaiyong Zhao, and Xiaowen Chu. Benchmarking the performance and energy efficiency of ai accelerators for ai training. In *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, pages 744–751. IEEE, 2020.

[52] Qianyu Huang and Tongfang Zhao. Data collection and labeling techniques for machine learning. *arXiv preprint arXiv:2407.12793*, 2024.

[53] Jingjing Xu, Wangchunshu Zhou, Zhiyi Fu, Hao Zhou, and Lei Li. A survey on green deep learning. *arXiv preprint arXiv:2111.05193*, 2021.

[54] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18, 2020.

[55] James L Cross, Michael A Choma, and John A Onofrey. Bias in medical ai: Implications for clinical decision-making. *PLOS Digital Health*, 3(11):e0000651, 2024.

[56] Konstantinos Mavrogiorgos, Athanasios Kiourtis, Argyro Mavrogiorgou, Andreas Menychtas, and Dimosthenis Kyriazis. Bias in machine learning: A literature review. *Applied Sciences*, 14(19):8860, 2024.

[57] Francois Chollet. The future of deep learning. *future*, 8(2), 2017.

[58] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[59] Donald S Fong, Lloyd Aiello, Thomas W Gardner, George L King, George Blankenship, Jerry D Cavallerano, Fredrick L Ferris III, Ronald Klein, and American Diabetes Association. Retinopathy in diabetes. *Diabetes care*, 27(suppl_1):s84–s87, 2004.

[60] Cheena Mohanty, Sakuntala Mahapatra, Biswaranjan Acharya, Fotis Kokkoras, Vassilis C Gerogiannis, Ioannis Karamitsos, and Andreas Kanavos. Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors*, 23(12):5726, 2023.

[61] Anila Sebastian, Omar Elharrouss, Somaya Al-Maadeed, and Noor Almaadeed. A survey on deep-learning-based diabetic retinopathy classification. *Diagnostics*, 13(3):345, 2023.

[62] Zhengwei Yang, Tien-En Tan, Yan Shao, Tien Yin Wong, and Xiaorong Li. Classification of diabetic retinopathy: Past, present and future. *Frontiers in endocrinology*, 13:1079217, 2022.

[63] Timothy HM Fung, Bakula Patel, Emma G Wilmot, and Winfried MK Amoaku. Diabetic retinopathy for the non-ophthalmologist. *Clinical Medicine*, 22(2):112–116, 2022.

[64] Tien-En Tan and Tien Yin Wong. Diabetic retinopathy: Looking forward to 2030. *Frontiers in Endocrinology*, 13:1077669, 2023.

[65] Zhengwei Yang, Tien-En Tan, Yan Shao, Tien Yin Wong, and Xiaorong Li. Classification of diabetic retinopathy: Past, present and future. *Frontiers in endocrinology*, 13:1079217, 2022.

[66] Anas Bilal, Liucun Zhu, Anan Deng, Huihui Lu, and Ning Wu. Ai-based automatic detection and classification of diabetic retinopathy using u-net and deep learning. *Symmetry*, 14(7):1427, 2022.

[67] Ryan Lee, Tien Y Wong, and Charumathi Sabanayagam. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye and vision*, 2:1–25, 2015.

[68] Angel Ayala, Tomas Ortiz Figueroa, Bruno Fernandes, and Francisco Cruz. Diabetic retinopathy improved detection using deep learning. *Applied Sciences*, 11(24):11970, 2021.

[69] Prawej Ansari, Noushin Tabasumma, Nayla Nuren Snigdha, Nawfal Hasan Siam, Rachana V. N. R. S. Panduru, Shofiul Azam, J. M. A. Hannan, and Yasser H. A. Abdel-Wahab. Diabetic retinopathy: An overview on mechanisms, pathophysiology and pharmacotherapy. *Diabetology*, 3(1):159–175, 2022.

[70] Kuan-Yu Lin, Wen-Hui Hsih, Yen-Bo Lin, Chen-Yu Wen, and Tien-Jyun Chang. Update in the epidemiology, risk factors, screening, and treatment of diabetic retinopathy. *Journal of diabetes investigation*, 12(8):1322–1325, 2021.

[71] Massimo Porta and Francesco Bandello. Diabetic retinopathy: a clinical update. *Diabetologia*, 45:1617–1634, 2002.

[72] Abdessalam MOUALDI. Diabetic retinopathy detection using deep learning techniques. Master's thesis, University of Kasdi Merbah Ouargla, 2023. Master's Thesis.

[73] Ehimare S Enaholo, Mutali J Musa, and Marco Zeppieri. Optical coherence tomography. In *StatPearls [Internet]*. StatPearls Publishing, 2024.

[74] Avinash V Varadarajan, Pinal Bavishi, Paisan Ruamviboonsuk, Peranut Chotcomwongse, Subhashini Venugopalan, Arunachalam Narayanaswamy, Jorge Cuadros, Kuniyoshi Kanai, George Bresnick, Mongkol Tadarati, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nature communications*, 11(1):130, 2020.

[75] Jacqueline Chua, Ralene Sim, Bingyao Tan, Damon Wong, Xinwen Yao, Xinyu Liu, Daniel SW Ting, Doreen Schmidl, Marcus Ang, Gerhard Garhöfer, et al. Optical coherence tomography angiography in diabetes and diabetic retinopathy. *Journal of Clinical Medicine*, 9(6):1723, 2020.

[76] Anna Bryl, Małgorzata Mrugacz, Mariusz Falkowski, and Katarzyna Zorena. The effect of diet and lifestyle on the course of diabetic retinopathy—a review of the literature. *Nutrients*, 14(6):1252, 2022.

[77] Stefano Fogli, Marzia Del Re, Eleonora Rofi, Chiara Posarelli, Michele Figus, and Romano Danesi. Clinical pharmacology of intravitreal anti-vegf drugs. *Eye*, 32(6):1010–1020, 2018.

[78] Burcu Nurözler Tabakcı and Nurten Ünlü. Corticosteroid treatment in diabetic macular edema. *Turkish journal of ophthalmology*, 47(3):156, 2017.

[79] Thanitsara Rittiphairoj, Tahreem A Mir, Tianjing Li, and Gianni Virgili. Intravitreal steroids for macular edema in diabetes. *Cochrane Database of Systematic Reviews*, (11), 2020.

[80] Gauri S Pande and Pravin Tidake. Laser treatment modalities for diabetic retinopathy. *Cureus*, 14(10), 2022.

[81] Makoto Gozawa, Yoshihiro Takamura, Seiji Miyake, Takehiro Matsumura, Masakazu Morioka, Yutaka Yamada, and Masaru Inatani. Photocoagulation of the retinal nonperfusion area prevents the expression of the vascular endothelial growth factor in an animal model. *Investigative Ophthalmology & Visual Science*, 58(13):5646–5653, 2017.

[82] Deepakkumar Mishra, Shilpkala Gade, Katie Glover, Ravi Sheshala, and Thakur Raghu Raj Singh. Vitreous humor: composition, characteristics and

implication on intravitreal drug delivery. *Current eye research*, 48(2):208–218, 2023.

[83] Yohei Tomita, Deokho Lee, Kazuo Tsubota, Kazuno Negishi, and Toshihide Kurihara. Updates on the current treatments for diabetic retinopathy and possibility of future oral therapy. *Journal of clinical medicine*, 10(20):4666, 2021.

[84] Wejdan L Alyoubi, Wafaa M Shalash, and Maysoon F Abulkhair. Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, 20:100377, 2020.

[85] María Herrero-Tudela, Roberto Romero-Oraá, Roberto Hornero, Gonzalo C Gutiérrez Tobal, María I López, and María García. An explainable deep-learning model reveals clinical clues in diabetic retinopathy through shap. *Biomedical Signal Processing and Control*, 102:107328, 2025.

[86] Sheikh Muhammad Saiful Islam, Md Mahedi Hasan, and Sohaib Abdullah. Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images. *arXiv preprint arXiv:1812.10595*, 2018.

[87] Md Mohaimenul Islam, Hsuan-Chia Yang, Tahmina Nasrin Poly, Wen-Shan Jian, and Yu-Chuan Jack Li. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Computer Methods and Programs in Biomedicine*, 191:105320, 2020.

[88] Nur Nadirah Harun, Shahida Jamaludin, and Nur Hafizati Mohtar. Diabetic retinopathy detection with deep learning: A resnet-cnn model enhanced by attention mechanism and ensemble learning. *Open Journal of Applied Sciences*, 15(3):688–699, 2025.

[89] Karthik, Maggie, and Sohier Dane. Aptos 2019 blindness detection. `https://kaggle.com/competitions/aptos2019-blindness-detection`, 2019. Kaggle.

[90] İsmail Akgül, Ömer Çağrı Yavuz, and Uğur Yavuz. Deep learning based models for detection of diabetic retinopathy. *Tehnički glasnik*, 17(4):581–587, 2023.

[91] Mohit Prabhushankar, Kiran Kokilepersaud, Yash-yee Logan, Stephanie Trejo Corona, Ghassan AlRegib, and Charles Wykoff. Olives dataset: Ophthalmic labels for investigating visual eye semantics. In *Proceedings of the*

Neural Information Processing Systems Track on Datasets and Benchmarks 2 (NeurIPS Datasets and Benchmarks 2022), 2022.

[92] Mohit Prabhushankar, Kiran Kokilepersaud, Yash yee Logan, Stephanie Trejo Corona, Ghassan AlRegib, and Charles Wykoff. Olives dataset: Ophthalmic labels for investigating visual eye semantics, June 2022.

[93] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[94] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[95] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[96] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.