République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de laRecherche Scientifique

A.MIRA-BEJAIA University
Faculty of Exact Sciences
Department of Computer Science

Master's thesis
In
Computer Science

Option
advanced information system

Theme

Diabetic Retinopathy classification using transfer learning and GAN

Made by:

Mr. ABICHE Yacine
Mr. AMOKRANE Akram

The jury is made up of:

President : Mr K. Amroun	Université de Béjaia.
Examiner : Mr F. Bouchebbah	Université de Béjaia.
Supervisor : Ms D. Boukredera	Université de Béjaia.
Co-Supervisor : Ms S. Ait Kaci Azzou	Université de Béjaia.

U. A/M Béjaia, July 2023.

Acknowledgments

First and foremost, it seems appropriate to begin this thesis by expressing our gratitude, firstly to God for granting us the strength and courage to successfully complete this modest work.

Our heartfelt appreciation and gratitude go to our supervisors, Ms Samira AIT KACI AZZOU and Ms Djamila BOUKREDERA, who have supported and guided us throughout this professional experience with patience and enthusiasm.

We would also like to thank the members of the jury for agreeing to examine and evaluate our work.

To all those who have contributed, directly or indirectly, through their advice, encouragement, or friendship, to the completion of this work, we express our deep gratitude.

For their encouragement, moral support, and the patience they have shown us throughout the year, we are sincerely grateful to all members of our families.

Dedications

This humble work is dedicated to:

Our dear parents who supported and encouraged us throughout

our education,

Our brothers and sisters,

Our teachers,

Our friends,

And to all the people who provided us with assistance.

Yacine & Akram

Contents

List of figures

List of Tables

General Introduction

In recent decades, there has been a significant global increase in the prevalence of diabetes-related diseases. The number of adults affected by diabetes has risen to approximately 537 million worldwide, and this number is projected to further increase to 643 million by 2030 and 783 million by 2045. Algeria, like many other countries, is also experiencing a rise in the prevalence of diabetes. Currently, diabetes affects around 14.4% of the Algerian population aged between 18 and 69 years. It is predicted that the number of adults with diabetes in Algeria will reach 3.4 million by 2045, reflecting an increase of 8.1% compared to the present [27].

This disease often leads to blindness in patients between 20 to 74 years of age due to a condition caused by uncontrolled diabetes known as Diabetic Retinopathy (DR) [34]. Diabetic Retinopathy (DR) is a disorder that damages the retinal blood vessels causing them to grow abnormally, sweal and eventually tear. It is characterized by various signs and symptoms that manifest in retinal tissue. These symptoms include microaneurysms, exudates (leakage of fluid), hemorrhages (bleeding), and swelling of the blood vessels. [34]

In general, Diabetic Retinopathy (DR) can be divided into two stages: non-proliferative DR (NPDR) and proliferative DR (PDR) [9]. NPDR can be further classified into three types based on severity: mild NPDR, moderate NPDR, and severe NPDR [58]. Overall, DR can be graded using a 5-point scale, consisting of the following categories: no DR, mild DR, moderate DR, severe DR, and proliferative DR [58]. This grading system provides a comprehensive assessment of the severity and progression of DR, allowing for appropriate diagnosis and treatment.

Diagnosis of DR can be done either manually by an ophthalmologist or through an automated system. There are pros and cons to both of these methods of DR detection. The only benefit of manual detection is that, it does not require any computer processing or sophisticated algorithms for the DR detection process. However, it is important to acknowledge the limitations associated with this method of diagnosing diabetic retinopathy (DR). Here are some of the limitations of manual diagnosis:

- About 80 percent of DR patients are in developing or underdeveloped countries with limited resources.These countries often lack ophthalmologists and basic DR detection mechanisms
- Plenty of people harmed by DR do not visit an eye-care professional unless the DR situation extends to the severe NPDR or PDR stage. Detection of DR at an early stage is very important in order to save the patient's vision.
- Detecting and grading DRs in the early stages requires time and expertise.
- Traditional measures to identify DR involve ophthalmologists for assessment and diagnosing capability, which is time-consuming and very costly work.

- Manual evaluation of DR patients shows discrepancies among practitioners :there is often a significant level of variability in the manual evaluation of DR among different practitioners.
- Manual screening techniques are insufficient to meet the growing demand for DR diagnostics worldwide.
- Sometimes the signs of DR at its initial phase are so small that even an expert ophthalmologist cannot recognize them properly.

Therefore, it has been proven that manual screening techniques are insufficient to keep pace with the growing need for diagnostic methods of DR worldwide [ref]

Rapid advancements in Artificial Intelligence (AI) have led the way for early detection of DR through automated systems, offering numerous benefits compared to manual detection methods. These benefits include:

- ✓ reducing the workload on ophthalmologists.
- ✓ and minimizing the risk of human error.
- ✓ automated systems have the potential to detect lesions and abnormalities more effectively and efficiently than manual methods.

Therefore, automated detection of DR is essential. Such automated systems can be developed using either machine learning or deep learning approaches.

Although machine learning-based frameworks have shown resilience in detecting DR, their effectiveness is highly dependent on hand-crafted features that are still difficult to generalize. In contrast, deep learning (DL) methods have provided techniques for automatic feature extraction from fundus images to overcome these drawbacks[74].

Deep learning approaches require large amounts of data, memory and computing power [16] which can be very challenging especially in the medical field. The number of medical images available is usually small, their acquisition and labeling is a very costly process and datasets obtained are usually unbalanced. This is why, in the recent years, the Generative Adversarial Network (GAN)[18] has become the technique of choice for image generation and conversion in medical imaging[97]. On the other hand, when it comes to the classification of medical images, transfer learning offers a significant advantage by minimizing the time and computating power required, making it a valuable approach [39].

In the field of ophthalmology, particularly in the detection of Diabetic Retinopathy (DR), the application of GANs is still in its early stages, and the existing literature on this topic is relatively limited. Apart from the scarcity of data, fundus images face various challenges, such as limitations of imaging devices, variations in examiner skill, and anatomical differences in the eye. The quality of the acquired images can significantly impact the diagnostic performance of ocular images.

The contributions made in this work include:

1. Defining a transfer learning-based CNN architecture for the detection and classification of retinopathy into 3 and 5 grades(a part of this work has been presented in a conference « Colloque sur les Objets et systèmes Connectés COC 2023 »).
2. Several transfer learning-based models were proposed and trained on two different datasets to highlight the impact of data on model performance.

3. After parameter optimization, the best model is selected, which will serve as the discriminator in the GAN.

4. A 5-layers CNN is implemented to define the Generator, which aims to denoise the images.

5. The GAN is implemented by training the Generator and Discriminator together.

The organization of this thesis is as follows:

In the first chapter, we will introduce Machine Learning techniques that have the potential to revolutionize the healthcare field, focusing specifically on Deep Learning, Transfer Learning, and Generative Adversarial Networks (GANs).

In the second chapter, we will provide an overview of diabetic retinopathy (DR) and cover essential concepts related to its classification. Additionally, we will review existing works in the field of DR classification, highlighting the advancements and challenges in this area.

In the final chapter, we will unveil the architecture employed and present the results obtained for the detection and multi-class classification of the condition.

We will conclude this thesis with a conclusion that summarizes the essence of our work.

Chapter 1

Artificial Intelligence

I Introduction

As computing power continues to advance, the potential of artificial intelligence (AI) and machine learning (ML) to transform healthcare becomes increasingly evident[57]. Among the various ML techniques, deep learning and transfer learning have significantly improved the accuracy of analyzing medical images, including X-rays, MRI scans, and eye fundus images. Additionally, the advent of Generative Adversarial Networks (GANs) in ML has expanded the possibilities by enabling the generation of realistic synthetic data for tasks such as image synthesis and data augmentation[48]. GANs have proven to be particularly valuable in healthcare due to their ability to address the challenges posed by imbalanced datasets[48].

The combination of deep learning, transfer learning, and GANs has propelled AI research and applications to new heights. These techniques offer powerful tools for solving complex problems, enabling advancements in various fields such as computer vision, natural language processing, and data generation. As researchers continue to explore and refine these methods, AI's potential to transform industries and address societal challenges will only continue to expand.A hierarchy of Artificial Intelligence Concepts: Machine Learning, Deep Learning, Transfer Learning, and GANs is illustrated in Figure 1.1.

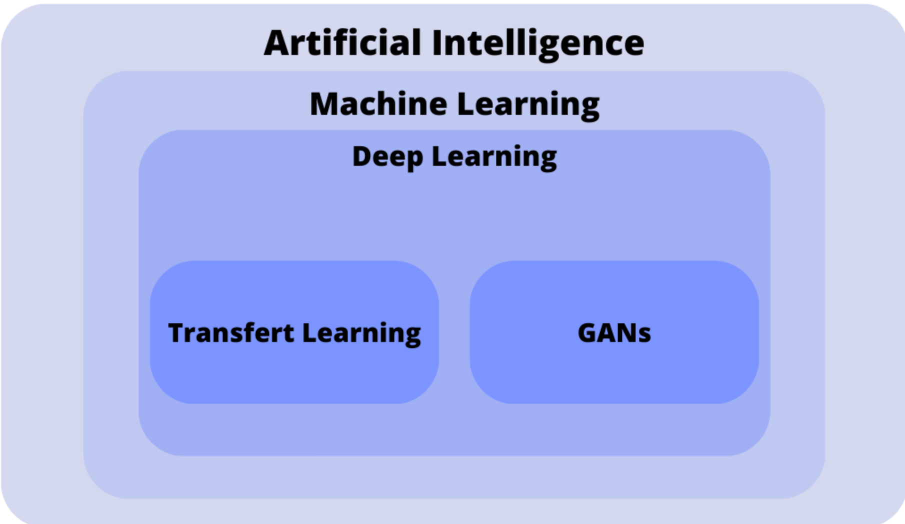


Figure 1.1: Artificial intelligence subfields

II Machine learning

Machine Learning (ML) is a subfield of artificial intelligence that focuses on the development of algorithms and models capable of automatically learning patterns and making predictions or decisions based on data[82]. It was defined in the 1950s by AI pioneer Arthur Samuel as the field of study that gives computers the ability to learn without explicitly being programmed. It is an interdisciplinary field that combines principles from statistics, mathematics, and computer science to enable computers to learn from and improve with experience. The goal of machine learning is to develop algorithms that can generalize and make accurate predictions or decisions on unseen data[98].

Machine learning its self is divided to three branches as illustrated in Figure 1.2: supervised learning, unsupervised learning and reenforced learning.

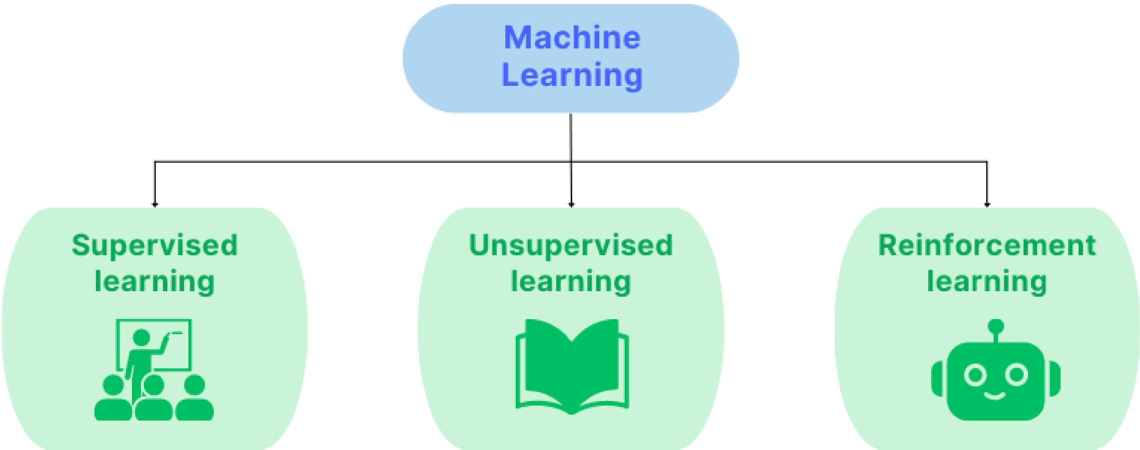


Figure 1.2: Branches of ML

II.1 Supervised learning

Supervised learning is a popular branch of machine learning that deals with learning from labeled data. In supervised learning, the training dataset consists of input data along with their corresponding output labels[11]. The aim is to learn a mapping function that can accurately predict the output labels for new, unseen input data[53] as shown in Figure 1.3. There are various algorithms used in supervised learning, including linear regression, logistic regression, support vector machines (SVM)[14], decision trees[50], and random forests[12], among others. Each algorithm has its own strengths and is suited for different types of problems. The training process in supervised learning involves optimizing the model parameters to minimize the difference between the predicted labels and the true labels in the training data. This process often involves the use of loss functions and optimization techniques such as gradient descent. Supervised learning is widely used in various domains, such as image classification, sentiment analysis, fraud detection, and medical diagnosis[53]. Its applications range from simple tasks to complex problems that require advanced modeling techniques.

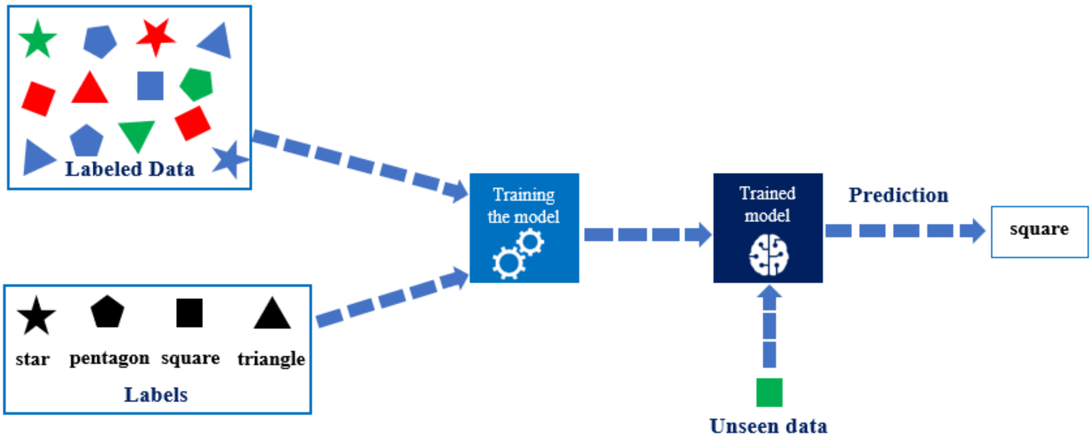


Figure 1.3: supervised learning

II.2 Unsupervised learning

Unsupervised learning is another important branch of machine learning that deals with learning from unlabeled data as illustrated in Figure 1.4. Unlike supervised learning, there are no predefined output labels available in unsupervised learning[11]. Instead, the goal is to discover hidden patterns, structures, or relationships within the data. Clustering is a common technique used in unsupervised learning, where the aim is to group similar data points together based on their intrinsic properties. Other techniques include dimensionality reduction, anomaly detection, and generative modeling[29]. In unsupervised learning, the models learn directly from the input data without any explicit guidance or supervision. The algorithms aim to capture the underlying structure of the data and identify meaningful patterns or clusters. This can provide insights into the data and help in understanding the inherent characteristics of the dataset. Unsupervised learning has various applications, such as customer segmentation, recommender systems, outlier detection, and data preprocessing. It is particularly useful when labeled data is scarce or expensive to obtain, as it can uncover valuable information and patterns in the absence of explicit labels[11].

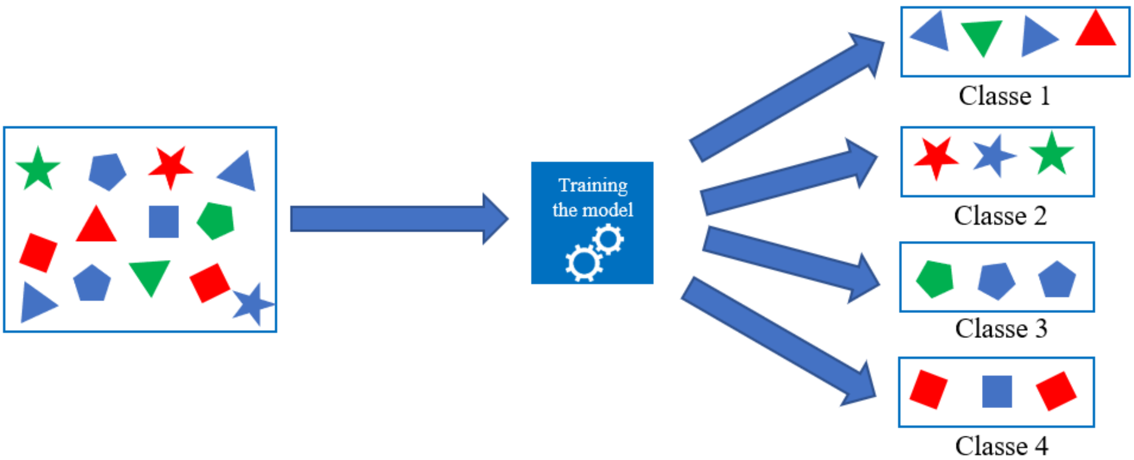


Figure 1.4: Unsupervised learning

II.3 Reinforcement learning

Reinforcement learning focuses on how an agent can learn to make decisions and take actions in an environment to maximize a cumulative reward. It draws inspiration from the concept of how humans and animals learn through trial and error[96]. In reinforcement learning as illustrated in Figure 1.5, an agent interacts with an environment, observes its state, and takes actions to transition to new states. The agent receives feedback in the form of rewards or penalties based on its actions, guiding it towards desirable outcomes. Through a process of exploration and exploitation, the agent learns a policy—a mapping from states to actions—that maximizes the expected long-term reward[96]. Reinforcement learning algorithms, such as Q-learning and policy gradients, utilize various techniques like value function estimation, exploration strategies, and temporal difference learning to iteratively improve the agent's decision-making capabilities. Reinforcement learning has been successfully applied to various domains, including robotics, game playing, recommendation systems, and autonomous vehicles, offering the potential for intelligent systems to learn and adapt in dynamic environments.

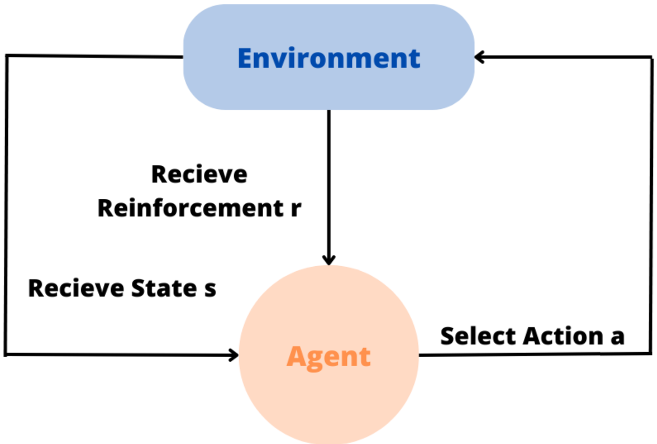


Figure 1.5: Reinforcement learning.

In conclusion, machine learning plays a crucial role in solving a wide range of real-world problems and advancing the field of artificial intelligence. However, it is still facing challenges like the feature extraction and preparation of the data to be usable by traditional ML techniques like SVM and Random Forest which requires human involvement.

III Deep learning

Deep learning is a subfield of machine learning that focuses on training artificial neural networks to learn hierarchical representations of data. It has gained significant attention and revolutionized various domains, including computer vision, natural language processing, and speech recognition[42]. Unlike traditional machine learning algorithms, deep learning models can automatically learn intricate patterns and complex representations directly from raw data which makes them more autonomous than traditional ML models. Figure 1.6 represents the general structure of a deep neural network.

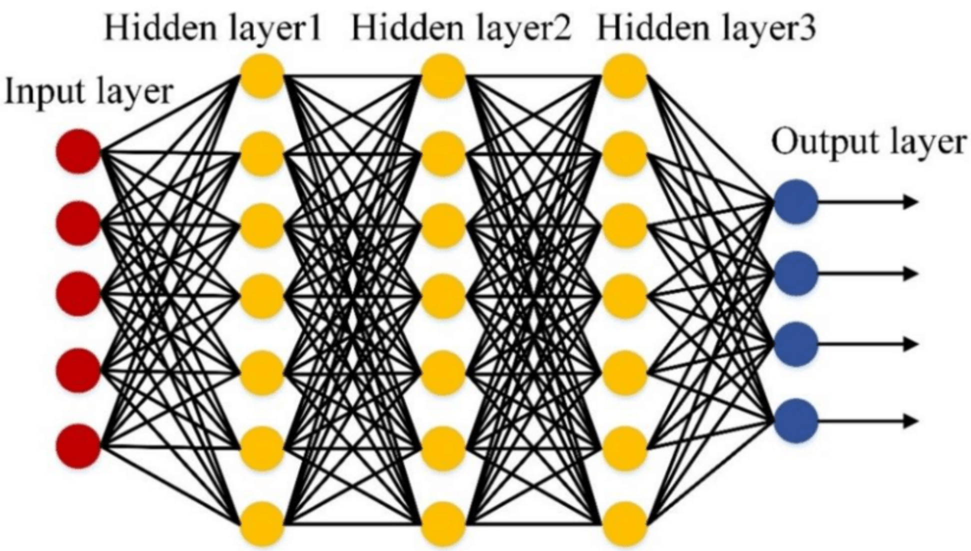


Figure 1.6: Structure of the deep neural network [62]

III.1 Neural Networks and Deep Neural Networks

Neural networks are the fundamental building blocks of deep learning. They consist of interconnected nodes, or artificial neurons like the one illustrated in Figure 1.7, organized into layers. Each neuron receives input signals, applies a transformation using an activation function, and produces an output signal. Deep neural networks (DNNs) refer to neural networks with multiple hidden layers, enabling them to learn increasingly abstract and high-level features[38].

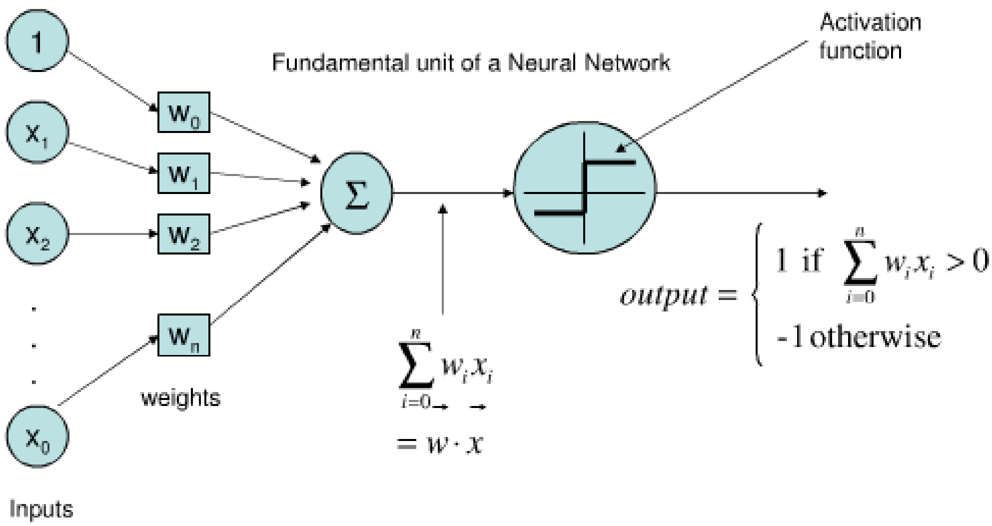


Figure 1.7: Single Layer Perceptron Network[81].

III.2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a specialized type of deep neural networks designed for processing grid-like data, such as images. They employ convolutional layers, which apply filters to input images, capturing local patterns and spatial dependencies as illustrated in Figure 1.8. CNNs also incorporate pooling layers to downsample feature maps, reducing computational complexity and enhancing translation invariance. With their ability to automatically learn hierarchical representations, CNNs have achieved remarkable performance in image classification, object detection, and other computer vision tasks[30].

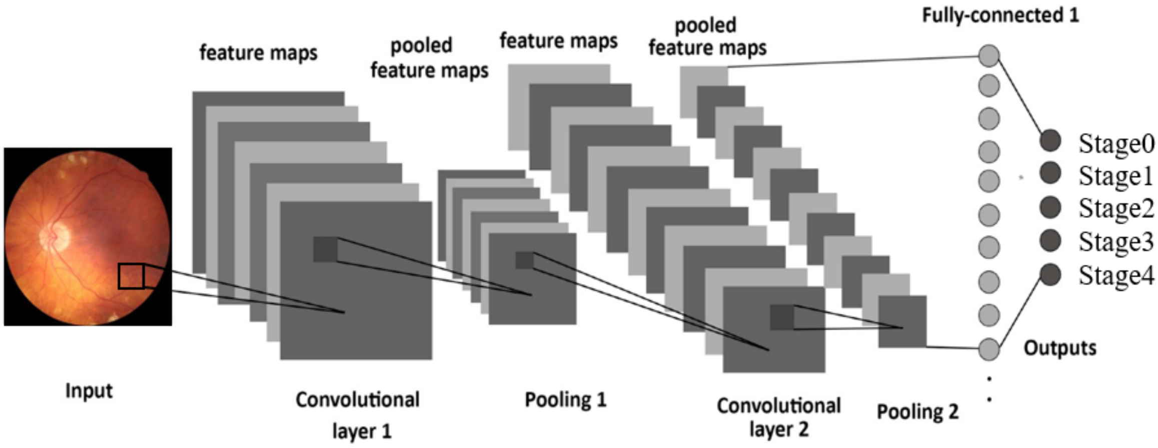


Figure 1.8: CNN structure explaining convolutions and poolings

III.3 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are another category of deep neural networks that excel in handling sequential and temporal data, such as speech and text. RNNs have recurrent connections that allow information to persist and flow through the network, enabling them to capture dependencies over time. This sequential memory makes RNNs suitable for tasks like natural language processing, speech recognition, machine translation, and sentiment analysis. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular variants of RNNs that address the vanishing gradient problem and enable better modeling of long-term dependencies[72]. Figure 1.9 is an unrolled RNN representation.

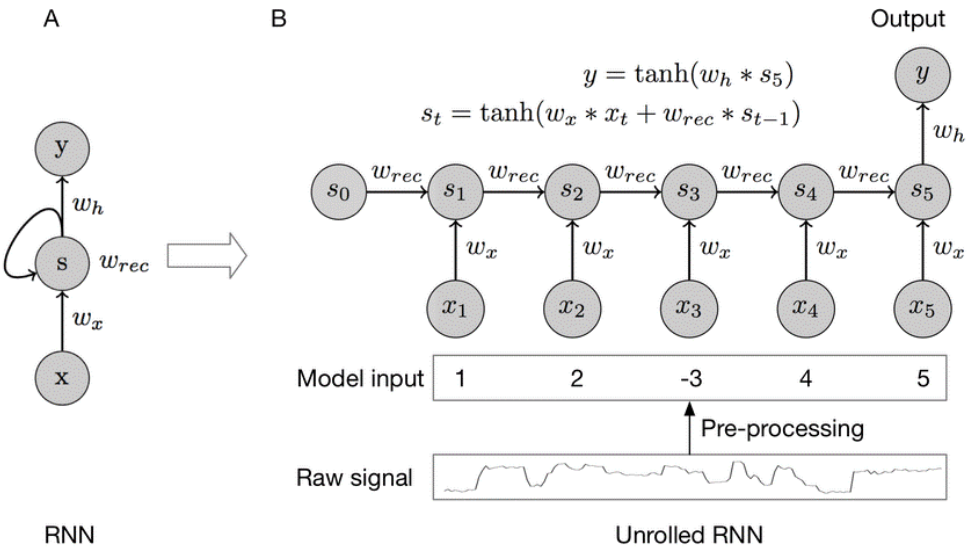


Figure 1.9: Illustration of a recurrent neural network. (A) A typical rolled RNN representation. (B) An easy-to-understand unrolled RNN representation[56].

III.4 Applications and Advancements in Deep Learning

Deep learning has made significant contributions to a wide range of fields and has been instrumental in driving advancements in various applications[85]. Its ability to learn intricate patterns and extract high-level representations has opened up new possibilities and achieved breakthroughs in several domains. Some notable applications and advancements in deep learning include:

- Computer Vision: Deep learning has revolutionized computer vision tasks, such as image classification, object detection, semantic segmentation, and image generation.
- Natural Language Processing (NLP): Deep learning techniques have greatly improved the field of NLP, enabling tasks such as sentiment analysis, machine translation, question-answering systems, text summarization, and language generation.
- Speech Recognition: Deep learning models, particularly recurrent neural networks, have significantly advanced automatic speech recognition (ASR) systems. These models have demonstrated impressive performance in speech-to-text conversion, enabling voice assistants, transcription services, and voice-controlled systems.
- Healthcare and Biomedicine: Deep learning has shown great potential in medical image analysis, disease diagnosis, and drug discovery. Convolutional Neural Networks (CNNs) have been used for early detection of diseases, such as diabetic retinopathy, cancer, and cardiovascular diseases, by analyzing medical images.
- Autonomous Systems: Deep learning has played a vital role in the development of autonomous systems, including self-driving cars, drones, and robots. Deep neural networks enable these systems to perceive and interpret their environment, make real-time decisions, and adapt to changing situations. This has led to significant advancements in the field of robotics and autonomous navigation.
- Generative Models: Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have gained attention in the deep learning community. GANs can generate realistic synthetic data, such as images and videos, while VAEs can learn meaningful latent representations and reconstruct missing or corrupted data. These models have applications in data augmentation, image synthesis, and creative content generation.

As deep learning has advanced, it has encountered a set of challenges including limited availability of labeled data, the substantial computational resources and time needed for training large models, and the issue of overfitting in complex networks. These challenges have prompted the extensive adoption of transfer learning, a technique that effectively tackles these obstacles. Transfer learning harnesses the power of pre-trained models trained on extensive datasets, enabling the transfer of knowledge and enhancing performance even when confronted with limited labeled data. By embracing transfer learning, researchers can overcome the limitations inherent in deep learning, leading to enhanced efficiency and effectiveness in diverse domains.

IV Transfer learning

Transfer learning is a powerful technique in deep learning that leverages knowledge learned from one task or domain and applies it to a different but related task or domain[86]. It enables the reuse of pre-trained models or features, saving computational resources and improving the performance of models on new tasks with limited labeled data. Figure 1.10 explains the learning process using pre-trained models.

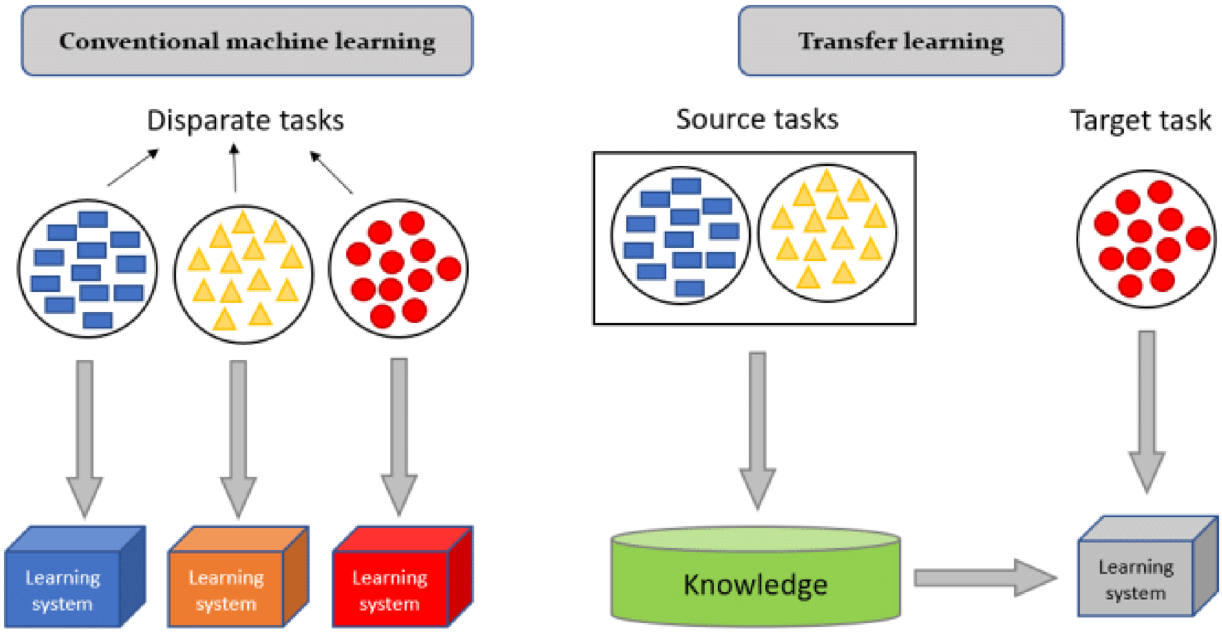


Figure 1.10: Comparative diagram of Learning Processes between Conventional Machine Learning and Transfer Learning[78]

IV.1 Pretrained Models and Feature Extraction

Pre-trained models are deep learning models that have been trained on large-scale datasets, typically for generic tasks such as image classification on well-known datasets like ImageNet[78]. These models have learned rich representations of data that can be transferred to new tasks. In transfer learning, the pretrained models act as a valuable source of knowledge that can be utilized[61]. Feature extraction is a common approach in transfer learning, where the pre-trained models are used as powerful feature extractors. The earlier layers of the pretrained models capture low-level features such as edges and textures, while the deeper layers capture more complex and abstract features. By using these pretrained features as inputs, new models can focus on learning task-specific patterns and relationships[61]. Here are the most commonly used pretrained models for feature extraction as illustrated in Fiure 1.11:

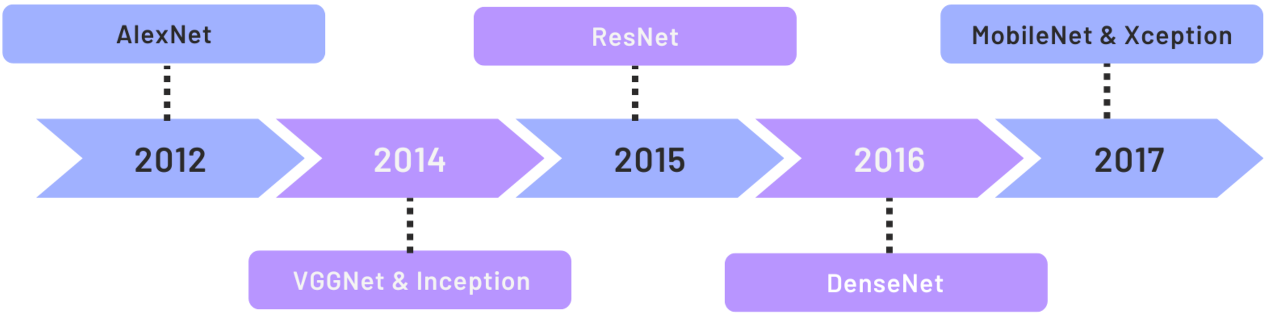


Figure 1.11: A timeline that represents key milestones in the development of CNNs and TL models

IV.1.1 AlexNet

AlexNet is a groundbreaking convolutional neural network (CNN) architecture that played a crucial role in advancing deep learning. It was proposed by Krizhevsky, Sutskever, and Hinton in 2012 [51]. The network architecture of AlexNet consists of multiple convolutional layers with varying filter sizes, followed by max-pooling layers and fully connected layers. Notably, AlexNet introduced the concept of using rectified linear units (ReLU) as activation functions, which helped alleviate the vanishing gradient problem. Additionally, it utilized techniques such as local response normalization and dropout regularization to improve generalization performance. AlexNet achieved a significant breakthrough in image classification by winning the ImageNet Large Scale Visual Recognition Challenge in 2012[70], substantially outperforming traditional methods. Its success paved the way for the subsequent development of deeper and more complex CNN architectures. AlexNet's impact on the field of deep learning continues to be profound, inspiring further advancements in computer vision and other domains. Figure 1.12 illustrates the impact of the ReLU activation function.

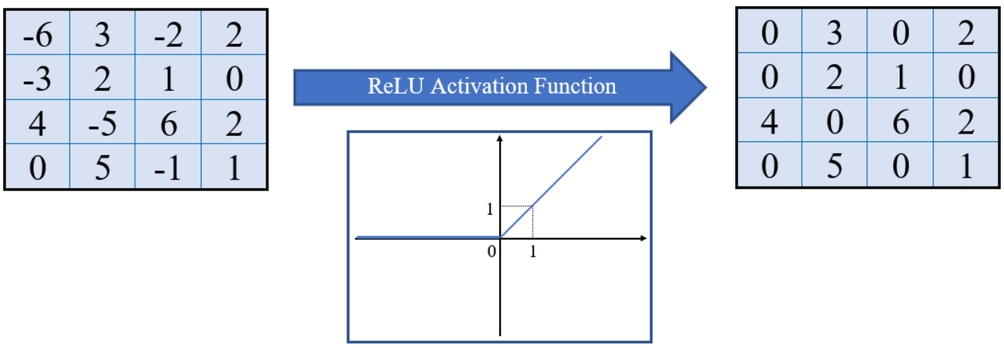


Figure 1.12: ReLU Activation function

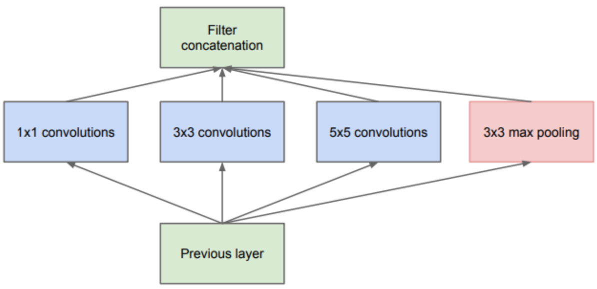
IV.1.2 VGGnet

VGGNet is a widely used convolutional neural network (CNN) architecture known for its simplicity and effectiveness in image classification. It was proposed by Simonyan and Zisserman in 2014 [76]. The network architecture consists of multiple convolutional layers with small 3x3 filters and a stride of 1, followed by max-pooling layers. This design choice allows for deeper networks while maintaining computational efficiency. VGGNet has achieved impressive results on various image classification tasks, including the ImageNet Large Scale

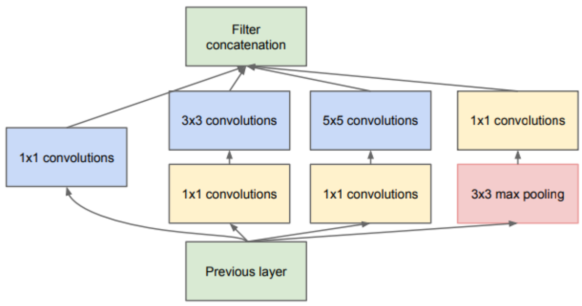
Visual Recognition Challenge. Variants like VGG16 and VGG19, which have 16 and 19 weight layers, respectively, have been particularly influential in the field. While newer architectures have surpassed VGGNet in terms of performance, it remains an important benchmark for evaluating the effectiveness of CNNs[71].

IV.1.3 Inception & Xception

Inception, proposed by Szegedy et al. in 2014[77], is a deep convolutional neural network architecture that introduced the concept of inception modules that are shown in Figure 1.13. These modules perform multiple convolutions with different filter sizes in parallel, allowing the network to capture information at different scales. The Inception architecture achieves state-of-the-art performance while maintaining computational efficiency. Xception, proposed by Chollet in 2017[17], is an extension of the Inception architecture that replaces the traditional convolutional layers with depthwise separable convolutions. This modification significantly reduces the number of parameters and computations, making the network more efficient. Xception achieves competitive performance with fewer parameters than previous architectures. Both Inception and Xception have made significant contributions to deep learning by introducing innovative module designs and pushing the boundaries of network efficiency and performance.



(a) Inception module, naïve version



(b) Inception module with dimension reductions

Figure 1.13: Inception module

IV.1.4 ResNet

ResNet, introduced by He et al. in 2015[32], is a revolutionary deep convolutional neural network architecture that addresses the challenge of training deep networks. It introduces residual connections, enabling the network to learn residual mappings. ResNet's key innovation is the use of residual blocks with skip connections, allowing effective gradient propagation and enabling the training of extremely deep networks. ResNet has achieved exceptional performance in image classification and other computer vision tasks. Its success has inspired subsequent architectures and made residual connections a fundamental technique in deep learning. Figure 1.14 illustrates the concept of residual learning.

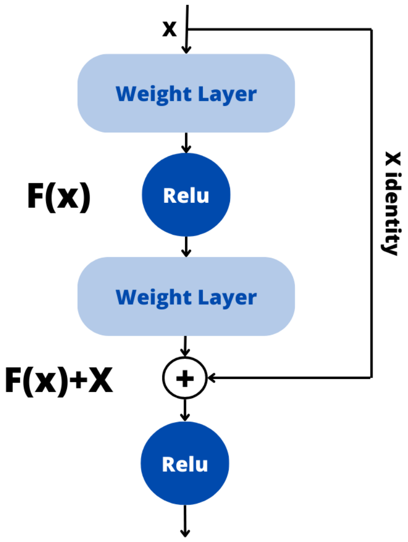


Figure 1.14: Residual learning

IV.1.5 DenseNet

DenseNet, introduced by Huang et al. in 2016[37], is a groundbreaking deep convolutional neural network architecture that offers a novel approach to addressing the challenges of training deep networks. It introduces densely connected blocks, where each layer is connected to every other layer in a feed-forward manner as shown in Figure 1.15. This dense connectivity promotes feature reuse and gradient flow, allowing for better parameter efficiency and improved learning[94]. DenseNet has achieved remarkable performance in image classification and other computer vision tasks. Its unique design has inspired subsequent architectures and solidified the importance of dense connections as a fundamental technique in the field of deep learning.

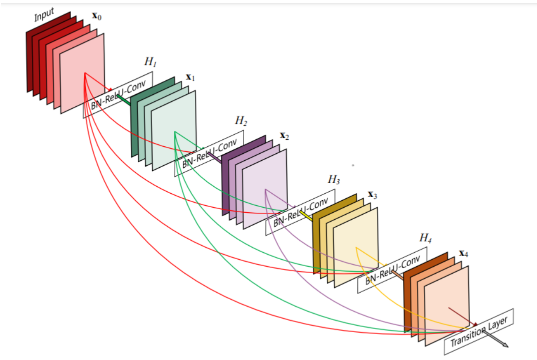


Figure 1.15: A 5-layer dense block with a growth rate of k = 4. Each layer takes all preceding feature-maps as input.[37]

IV.1.6 MobileNet

MobileNet, proposed by Howard et al. in 2017[35], is a lightweight convolutional neural network architecture specifically designed for mobile and embedded devices. It introduces depthwise separable convolutions, which factorize the standard convolutional operation into a depthwise convolution followed by a pointwise convolution as illustrated in Figure 1.16. This reduces the computational complexity and model size while maintaining good accuracy.

MobileNet achieves a good trade-off between model size and accuracy, making it suitable for resource-constrained environments. It has been widely adopted in various applications that require real-time or on-device deep learning inference.

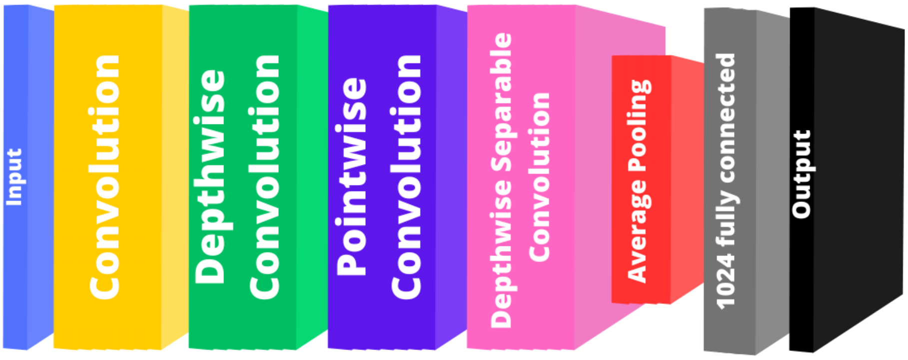


Figure 1.16: The structure of the MobileNet V1 network.

IV.2 Comparison of the pretrained models

The pretrained models mentioned (VGGNet, AlexNet, ResNet, Inception, Xception, MobileNet) differ manly in their architecture and design principles. Here are some key differences between them:

IV.2.1 Architecture:

- AlexNet: It was one of the first deep convolutional neural networks (CNNs) with a deeper architecture compared to previous models. It introduced the concept of using ReLU activation functions and dropout for regularization.

- VGGNet: It has a simple and uniform architecture with stacked convolutional layers and small filter sizes (3x3). It has a large number of trainable parameters.

- Inception: It employs the concept of inception modules, which are composed of multiple parallel convolutional layers of different filter sizes. This allows the network to capture features at different scales.

- ResNet: It introduced residual connections, allowing for the training of very deep networks (e.g., hundreds of layers) by addressing the vanishing gradient problem. It uses skip connections to propagate gradients effectively.

- DenseNet: It introduces dense connections between layers, where each layer is connected to every other layer in a feed-forward fashion. This dense connectivity promotes feature reuse and gradient flow. DenseNet has a compact architecture and uses small filter sizes (3x3).

- Xception: It extends the idea of Inception by replacing the standard convolutional layers with depthwise separable convolutions, which separate the spatial and channel-wise filtering. This reduces computational complexity.

- MobileNet: It focuses on efficiency and is designed for mobile and embedded devices. It utilizes depthwise separable convolutions and pointwise convolutions to reduce both computational cost and model size.

IV.2.2 Computational Efficiency:

- VGGNet, DenseNet, AlexNet, and ResNet have relatively higher computational requirements due to their deeper architectures and larger number of parameters.
- Inception and Xception strike a balance between accuracy and computational efficiency by using parallel convolutions.
- MobileNet is specifically optimized for low-resource environments and mobile devices, aiming for efficient inference with reduced computational cost.

IV.2.3 Performance

- The performance of these models can vary depending on the specific task, dataset, and implementation. Generally, deeper models like ResNet, DenseNet, Inception and Xception tend to achieve higher accuracy on challenging tasks but require more computational resources.
- VGGNet and AlexNet, while slightly less complex, can still deliver good performance in image classification tasks.

IV.3 Fine-tuning and Adapting Pretrained Models

Fine-tuning is another technique in transfer learning that involves updating and adjusting the pretrained models' weights on a new task-specific dataset[92]. This allows the models to adapt to the new task while still benefiting from the previously learned knowledge like it is shown in Figure 1.10. During fine-tuning, the weights of the pretrained models are adjusted based on the gradients computed from the new dataset, with careful consideration to prevent overfitting[92]. When fine-tuning pretrained models, various hyperparameters can be adjusted to achieve optimal performance. These hyperparameters include activation functions, dropout rates, learning rates, and batch sizes, among others. Here are some of the hyperparameters that have the most influence on the model's performance:

IV.3.1 Architecture modifications

Adjustments made to the model's architecture, such as adding or removing layers, changing layer sizes, etc. These modifications can affect the model's capacity to learn complex patterns or reduce computational requirements.

IV.3.2 Learning rate

The learning rate controls the step size at which the model's weights are updated during training. A higher learning rate may result in faster convergence, but it can also lead to instability or overshooting. Conversely, a lower learning rate may slow down training but can help the model converge more accurately.

IV.3.3 Loss function

The loss function is the objective function used to measure the model's performance during training. It quantifies the discrepancy between predicted and true values. Common loss functions include cross-entropy for classification tasks and mean squared error for regression tasks.

IV.3.4 Batch size

The batch size specifies the number of samples processed in each training iteration before updating the model's weights. A larger batch size can lead to faster training as more samples are processed simultaneously, but it requires more memory. Smaller batch sizes offer better generalization but may require more iterations to converge.

IV.3.5 Number of epochs

The number of epochs determines the number of times the entire dataset is passed through the model during training. Increasing the number of epochs allows the model to see the data more times, potentially improving performance. However, too many epochs can lead to overfitting.

IV.3.6 Activation functions

Activation functions introduce non-linearity to the model's layers, enabling the network to learn complex relationships. Common activation functions include ReLU (Rectified Linear Unit), sigmoid, and tanh. The choice of activation function can impact the model's ability to capture and represent different patterns in the data.

IV.3.7 Optimizer

The optimizer determines the algorithm used to update the model's weights based on the gradients computed during backpropagation. Popular optimizers include Adam, SGD (Stochastic Gradient Descent), and RMSprop. Each optimizer has its own update rules, convergence characteristics, and hyperparameters.

IV.3.8 Dropout rate

Dropout is a regularization technique where a fraction of neurons are randomly dropped out during training. It helps prevent overfitting by reducing the co-dependency between neurons, forcing the network to learn more robust representations. The dropout rate determines the fraction of neurons to be dropped out.

IV.3.9 Early stopping

Early stopping is a technique that stops training when the performance on a validation set starts deteriorating. It helps prevent overfitting by finding the optimal point where the model's generalization is the best. Early stopping requires monitoring the validation loss or other performance metrics during training

IV.3.10 Freezing layers

Freezing layers refers to the decision of keeping certain layers of the pretrained model fixed during fine-tuning. By freezing specific layers, their weights are not updated during training. This can be beneficial when the pretrained layers already capture relevant features, and freezing them prevents them from being modified.

These hyperparameters play a crucial role in fine-tuning and customizing pretrained models for specific tasks and datasets. Proper tuning of these parameters can significantly impact the model's performance and generalization capabilities.

IV.4 Benefits and Challenges of Transfer Learning

Transfer learning offers several benefits in machine learning and Deep learning. It helps in overcoming the limitations of restricted labeled data, as it can effectively leverage knowledge from large-scale datasets. By utilizing pretrained models or features, transfer learning reduces the need for extensive training on new datasets, saving computational time and resources. It also improves generalization performance, especially when the new task has a similar underlying structure or patterns to the original task. However, transfer learning also presents certain challenges. One challenge is determining the appropriate level of similarity between the original and new tasks or datasets to ensure effective knowledge transfer. Overfitting can be a concern when fine-tuning, as the pretrained models may already have strong features that are task-agnostic. Balancing the amount of relearning and preserving the learned knowledge is crucial.

In conclusion, transfer learning is a valuable technique in both machine learning and deep learning that enables the transfer of knowledge from pre-trained models or features to new tasks. By leveraging the learned representations, it enhances the performance and generalization capabilities of models, particularly in scenarios with limited labeled data. The use of pretrained models, feature extraction, and fine-tuning are common strategies in transfer learning, offering numerous benefits while addressing the associated challenges.

V Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a type of Machine Learning models that has received a lot of interest for its capacity to produce realistic and high-quality synthetic data[18]. GANs, invented in 2014 by Ian Goodfellow and his team [29], train a pair of competing networks[93]. The generator (G) network generates forgeries with the goal of creating realistic images, whilst the discriminator (D) network receives both forgeries and real images and attempts to identify them[18]. GANs have been used to produce impressive results in image synthesis, text generation, music generation, and even video synthesis. To generate realistic data with GANs, it involves sampling from the latent space by providing random noise as input to the generator. By manipulating the input latent vectors, it is possible to control specific attributes or generate diverse variations of the data.

V.1 GAN Architecture and Components

The generator and discriminator are the key components of Generative Adversarial Networks (GANs). Both of those components are trained in an iterative process, competing against each other[93]. The generator transforms random noise or a latent vector into synthetic data samples without explicit knowledge of the real data. The discriminator network, which acts as an adversary to the generator, is responsible for distinguishing between real and generated samples. The discriminator is provided with both real and generated samples and learns to differentiate between them[18]. Metaphorically, the generator and the discriminator can be seen as the forger or counterfeit artist and the police or authority respectively as shown in Figure 1.17. The generator aims to produce synthetic samples that are indistinguishable from genuine ones, despite having no explicit knowledge of the real data. On the other hand, the discriminator strives to improve its ability to differentiate between real and generated samples.

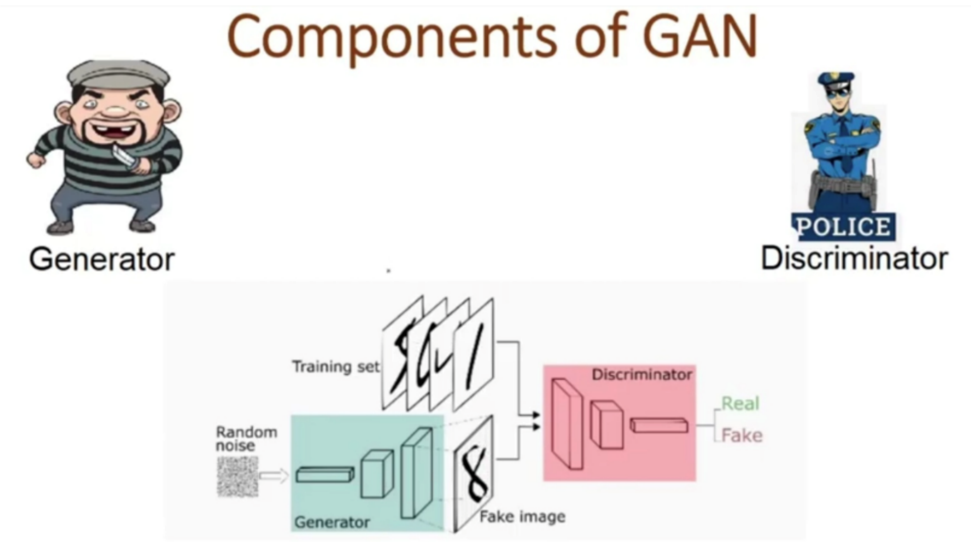


Figure 1.17: Components of GANs[5].

1. **Discriminator** – It is a supervised approach, which means it is a simple classifier that predicts whether data is real or fake. It is trained on real data and provides feedback to a generator[5].

2. **Generator** – It is an unsupervised learning approach. It generates fake data based on the original (real) data. It is implemented as a neural network with hidden layers, activation functions, and a loss function. Its objective is to generate fake images that can fool the discriminator into classifying them as real. Once the discriminator is successfully fooled by the generator, the training process stops, and we can consider a generalized GAN model to be created[5].

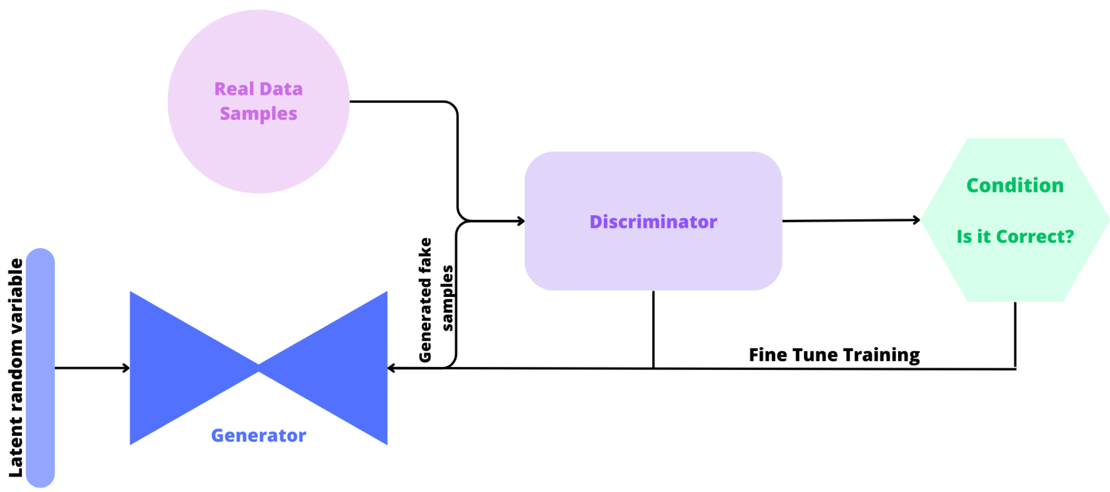


Figure 1.18: GAN training process

The generator takes random noise or a latent vector as input and generates fake samples as shown in Figure 1.18. The goal of the generator is to produce samples that are indistinguishable from real data. The discriminator network, on the other hand, tries to distinguish between real and generated samples. It is trained with real samples from the training data and generated samples from the generator. The discriminator's objective is to correctly classify real data as real and generated data as fake.

V.2 Training GANs and Generating Realistic Data

The training process of Generative Adversarial Networks (GANs) involves a dynamic interplay between the generator and discriminator networks[60]. If the discriminator rapidly recognizes the fake data that the generator produces, the generator suffers a penalty. As the feedback loop between the adversarial networks continues, the generator will begin to produce higher-quality and more believable output and the discriminator will become better at flagging data that has been artificially created.

The key aspect of GAN training, is the minimax game. The generator and discriminator networks are updated iteratively, with the generator aiming to minimize the discriminator's ability to distinguish real and fake samples, while the discriminator aims to maximize its discrimination accuracy[54]. This adversarial training leads to the generator progressively improving its ability to generate samples that resemble the real data distribution. Hence the GAN network is formulated as a minimax game where the Discriminator is trying to minimize its reward V(D, G) and the generator is trying to maximize the Discriminator loss. It can be mathematically described by the formula below:

$$V(G, D) = \mathbb{E}_{p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{p_g(x)}[\log(1 - D(x))] \tag{1.1}$$

Where $V(G, D)$ represents the value function of the Generative Adversarial Network (GAN). The terms $p_{\text{data}}(x)$ and $p_g(x)$ represent the data distribution and the distribution generated by the generator network, respectively. $D(x)$ is the discriminator network, and $\log D(x)$ and $\log(1 - D(x))$ are the logarithmic probabilities of the discriminator's output. The expectation operator $\mathbb{E}$ denotes the average over the respective distributions.

Various training strategies have been devised to enhance the stability and convergence of GAN training. One commonly employed approach is alternating training, where the generator and discriminator are trained in an alternating fashion[8]. During each training iteration, the generator generates samples, which are then used to update the discriminator. Subsequently, the updated discriminator is used to update the generator. This alternating process continues iteratively, allowing both networks to refine their respective capabilities[84].

V.3 Types of GANs

- Deep convolutional GAN (DCGAN)[59]: It is one of the most used, powerful, and successful types of GAN architecture. It is implemented with help of ConvNets in place of a Multi-layered perceptron. The ConvNets use a convolutional stride and are built without max pooling and layers in this network are not completely connected.

- Conditional GAN (CGAN)[60]: Conditional GAN is deep learning neural network in which some additional parameters are used. Labels are also put in inputs of Discriminator in order to help the discriminator to classify the input correctly and not easily full by the generator.

- CycleGAN[99]: It performs the task of Image Translation. Suppose we have trained it on a horse image dataset and we can translate it into zebra images.

- Super resolution GAN (SRGAN): Its main function is to transform low resolution to high resolution known as Domain Transformation[93].

V.4 Applications and Innovations in GANs

GANs have found numerous applications and have sparked innovations across various domains. In the field of ophtalmology, GANs have been used for tasks such as denoising, resolution augmentation, domain transfer and data augmentation as shown in Figures 1.19, 1.20, 1.21 and 1.22.

- Denoising:

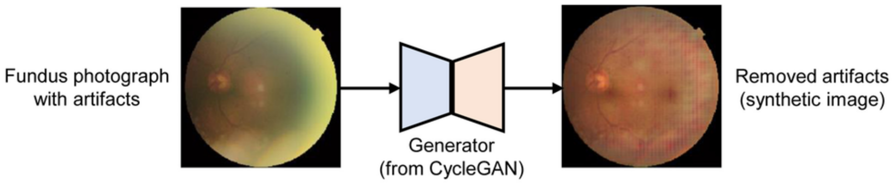


Figure 1.19: Denoising in fundus photography[89]

- Resolution augmentation

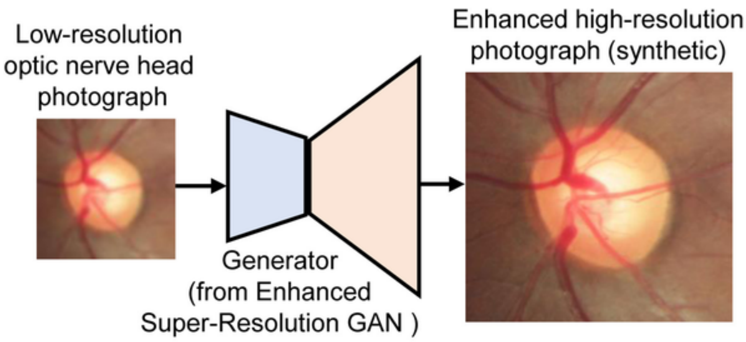


Figure 1.20: Super-resolution for optic nerve head photography[31]

- Domain transfer

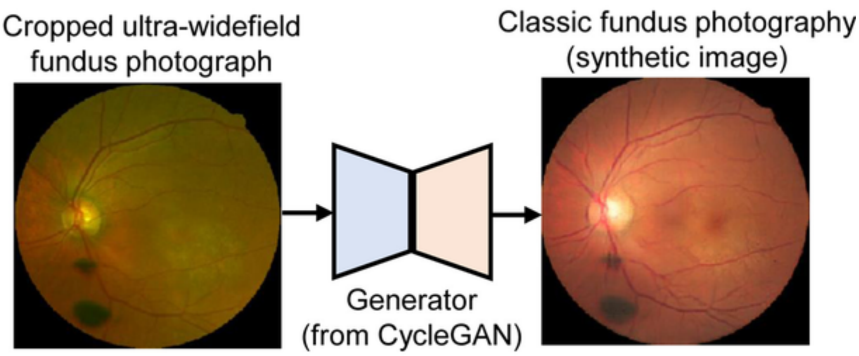


Figure 1.21: Ultra-widefeld to classic fundus photography domain transfer [91]

- Data augmentation

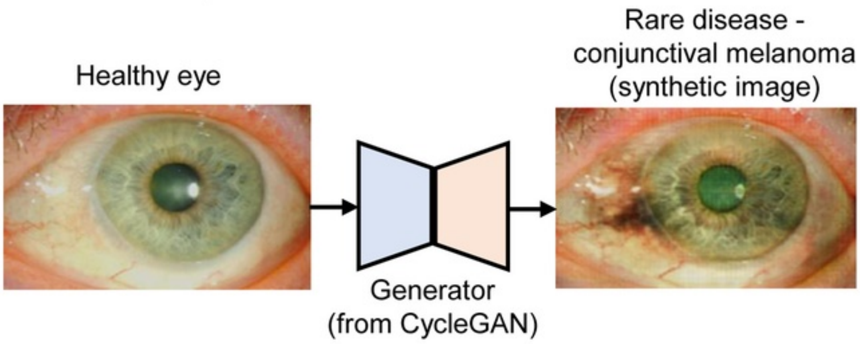


Figure 1.22: Data augmentation for ocular surface images[90]

As research in GANs continues, ongoing advancements aim to address challenges such as mode collapse (where the generator produces limited diversity in generated samples), stability during training, and improving the interpretability and control of generated data. GANs have the potential to revolutionize creative content generation, data augmentation, and enable breakthroughs in various fields by providing a powerful tool for generating realistic and diverse synthetic data.

V.5 conclusion

In this chapter we cover the fundamental concepts of machine learning, including supervised learning, unsupervised learning and reinforcement learning. In addition, we explored the importance of transfer learning, exploiting pre-existing knowledge to improve model performance. We also examined significant advances in the field of deep learning, in particular generative adversarial networks (GANs), which offer promising capabilities for generating realistic data samples. In the following chapter, we'll look at how ML techniques can be used to process ophthalmological images, and more specifically to detect diabetic retinopathy.

Chapter 2

State of art: Diabetic retinopathy classification

I Diabetic Retinopathy

Diabetic retinopathy (DR) is a leading cause of blindness in the mid-aged population[34], primarily affecting individuals who have had diabetes for a prolonged period of time. The condition is characterized by damaged retinal Blood Vessels (BV) due to persistently elevated blood glucose levels[9]. This damage can lead to swelling and abnormal growth of the said BVs causing fluid leakage in the retinal tissue resulting in blurry vision and potential vision loss. Specialists look for specific signs, known as lesions, to diagnose DR. As illustrated in Figure 2.1 the main lesions associated with DR diagnosis include Micro-aneurysms, hemorrhages, and soft and hard exudates[75].

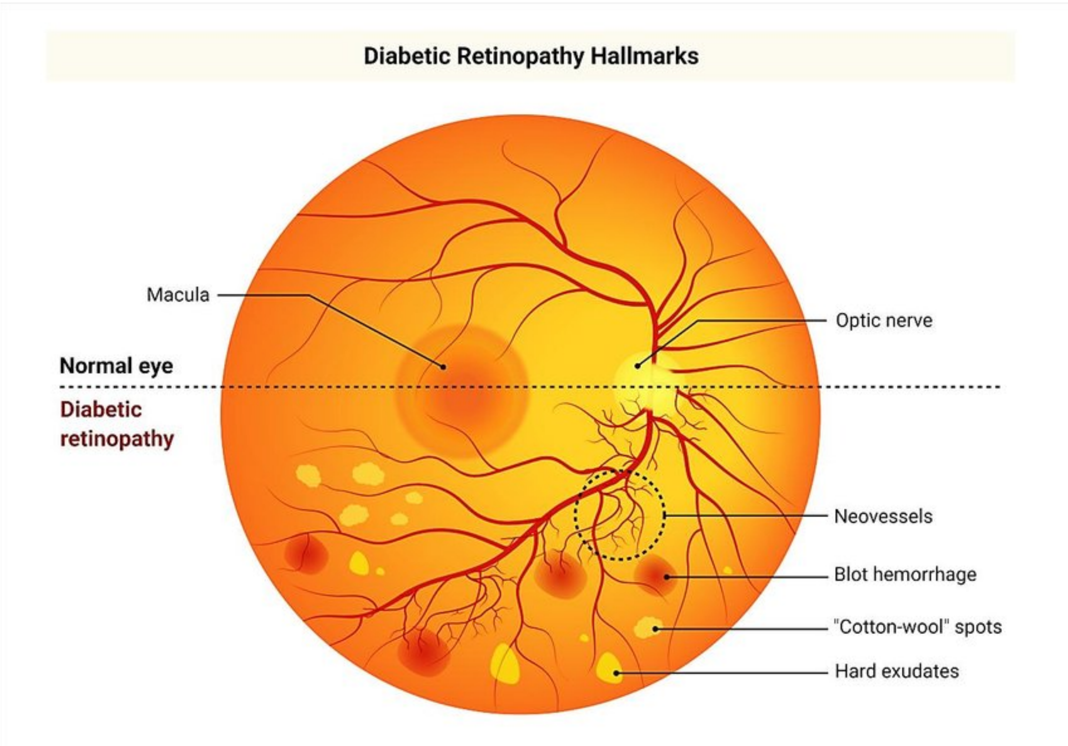


Figure 2.1: Diabetic Retinopathy hellmarks[25].

The presence and severity of these abnormalities are used to determine the stage of the disease. DR can be classified into four levels: mild, moderate and severe non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) as shown in the table below:

Table 2.1: Stages of DR[1]

Stage	Dilated Ophthalmoscopy Observable Signs	Severity
Stage 0	No signs	No DR
Stage 1	Micro-aneurysms	Mild NPDR
Stage 2	Micro-aneurysms, retinal dot and blot hemorrhages, hard exudates or cotton wool spots without signs of severe non-proliferative diabetic retinopathy	Moderate NPDR
Stage 3	More than 20 intra-retinal hemorrhages in each of 4 quadrants, definite venous beading in 2 or more quadrants, prominent intra-retinal microvascular abnormality (IRMA) in 1 or more quadrants without signs of proliferative retinopathy	Sever NPDR
Stage 4	Neovascularization or Vitreous/pre-retinal hemorrhage	Proliferative DR

Various techniques are employed by specialists to diagnose DR, including Dilated Eye Examination, Optical Coherence Tomography (OCT), and Fundus Photography. However, these techniques require significant human involvement and can be time-consuming, making it challenging to keep up with the demand for screening and diagnosis[83]. Consequently, many diabetic patients fail to adhere to recommendations for annual eye exams.

For diabetic patients, timely and appropriate care is essential to preventing long-term vision loss, enhancing patients' quality of life, and reducing the financial burden brought on by visual impairment[25]. Therefor a faster, easier, and more accurate method of diagnosing DR is required to encourage regular eye exams on an annual or semi-annual basis[7].

In conclusion, DR is a common eye disease among individuals with diabetes that can potentially lead to complete vision loss. Early detection is key as treatment is more effective and less costly during the initial stages. Given the overwhelming workload faced by medical professionals in diagnosing DR, especially considering the regular eye exams required for diabetics, computers have proven to be efficient and effective in assisting with diagnosis in general and the classification of DR in particular.

II Classification systems

Computer-aided diagnosis (CAD) systems have gained significant popularity as a research topic, with applications for many conditions including detection and classification of diabetic retinopathy (DR)[34]. Currently, ophthalmologists must carefully examine the patient's retina for common symptoms of the condition. However, it is time consuming and costly because it must be performed by a specialist. Computers, on the other hand, provide the benefit of automated classification, resulting in a higher accuracy and a lighter workload for specialists[83]. The use of Convolutional Neural Networks (CNNs) has particularly excelled in automatically classifying retinal images by extracting symptoms and labeling the said images without human intervention[83]. Despite the differences between the methods (such as Random forests, SVM, Decision trees, DNN, CNN and many more) the operation of developing such a model has to go through the same major phases: dataset collection, preprocessing, feature extraction, classification and evaluation as shown in Figure 2.2

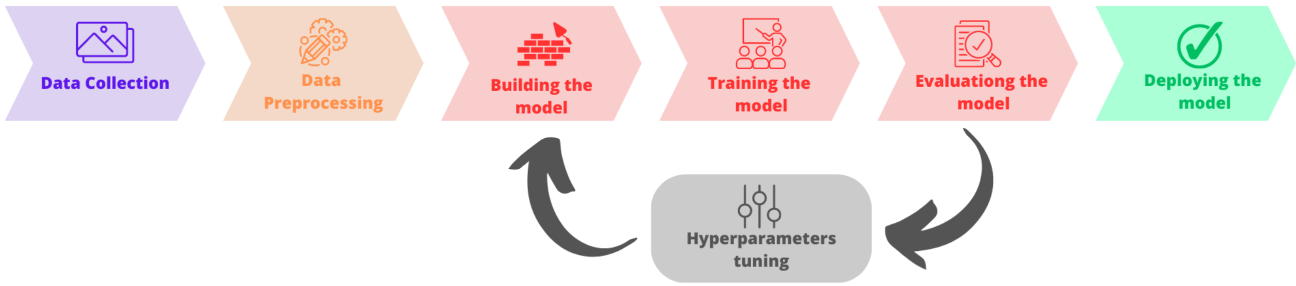


Figure 2.2: Machine Learning Process

II.1 Data collection

Datasets play a crucial role in the development and evaluation of computer-aided diagnosis (CAD) systems for diabetic retinopathy (DR). Since DR is a universal problem, there has been significant advances to find solutions which lead to the collection of so much data. Fundus images datasets the most present in research papers and they can be found on various platforms such as APTOS[45], EyePACS[19], Messidor[20] and more. These datasets vary in size, diversity, quality and severity stages or classes in this case. Here is a table showcasing the most used datasets and their corresponding characteristics:

Table 2.2: DR datasets

Dataset	Availability	Severity classes	Size (Number of Images)
Messidor-2[20]	public	5 (No DR to PDR)	1 748
IDRiD[64]	public	5 (No DR to PDR)	516
APTOS 2019[45]	public	5 (No DR to PDR)	3 662
Kaggle Diabetic Retinopathy Dataset (EyePACS)[19]	public	5 (No DR to PDR)	88 702
e-Ophtha	Restricted	3 (Mild to Severe PDR)	103
RITE[36]	Restricted	4 (No DR to svever PDR)	400
CHASE_DB1[15]	public	2 (No DR to Moderate NPDR)	28
DIARETDB1[24]	Public	4 (No DR to svever PDR)	89

As shown above the most important differences were the size of the dataset, severity classes and its availability. It is important to note that some datasets are imbalanced in the sense that not all classes of severity contain the same number of images. Another problem is the variation of image quality.

II.2 Data preprocessing

The preprocessing phase is critical when it comes to the detection and classification of DR images. It involves a wide range of techniques and procedures that contribute to improving the overall quality and interpretability of the images. By reducing noise, eliminating artifacts, and addressing other unwanted elements, preprocessing aims to enhance the accuracy of the classification system. There are several preprocessing techniques that can be used in the field of DR detection and classification, however there are few that constantly appear in the literature. These are just some of the commonly used ones in the field:

II.2.1 Dataset augmentation

dataset augmentation is the practice of applying a wide array of domain-specific transformations to expand a training set[23]. While dataset augmentation is a widely employed technique for various types of data, it has evolved into an indispensable approach in DR classification. It offers a means to overcome the limitations associated with imbalanced datasets and enhances the overall accuracy and reliability of DR classification models. To create new modified copies of existing images, we apply random geometric transformations such as flipping, rotating, zooming and cropping[23]. Figure 2.3 illustrates the resulting set of augmeted images.

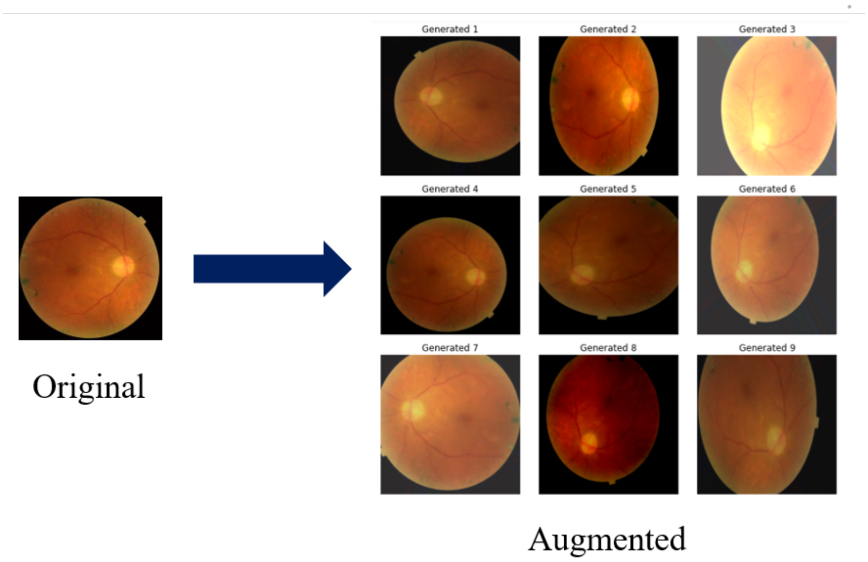


Figure 2.3: Images generated using data augmentation techniques

II.2.2 Image Normalization

In the context of DR classification, normalizing images is a critical step due to the differences in imaging conditions, equipment, or image acquisition techniques. Image normalization involves applying various techniques to enhance comparability and address variations in brightness and contrast. One widely utilized technique in the field of DR classification is scaling the pixel values to a common range, typically between 0 and 1. By doing so, it ensures optimal comparisons across data acquisition methods and texture instances[22] enabling the identification of subtle features and patterns in the images. Figure 2.4 provides an example of the normalization technique on a fundus image.

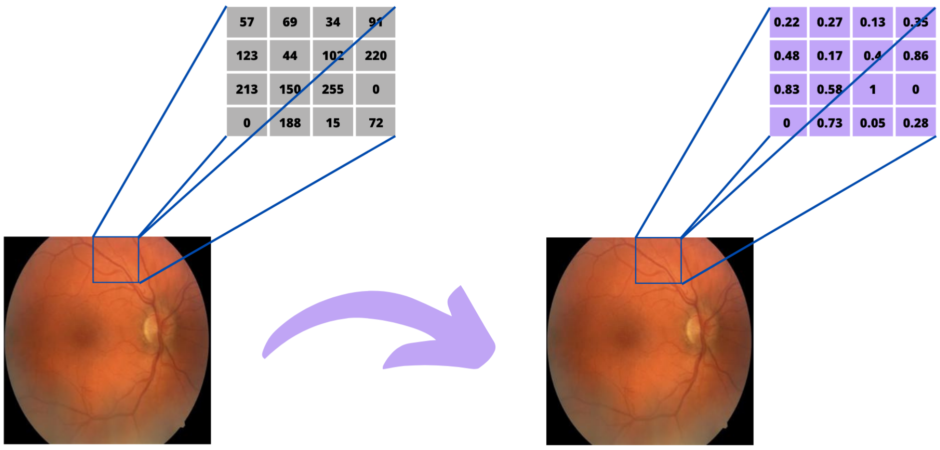


Figure 2.4: fundus images normalization

II.2.3 Image Resizing

Exactly like image normalization, resizing the images to a uniform size ensures optimal comparisons across data acquisition methods and equipment. By enforcing a uniform size, the images can be fed into classification models without the need for further adjustments or resizing during the training and evaluation stages. It also ensures that all images are treated equally in terms of spatial information, allowing the classification algorithms to focus on the intrinsic characteristics of the retinal structures and patterns associated with DR.

II.2.4 Image Cropping

Cropping images means removing irrelevant or unwanted portions of the image to focus on the region of interest. For example, in DR images circle cropping is the most popular since it helps the model to direct its focus towards the central part of the retina, where important features for DR detection and classification are often concentrated. the Figure2.5 shows the copping tecnique used on a fundus image.

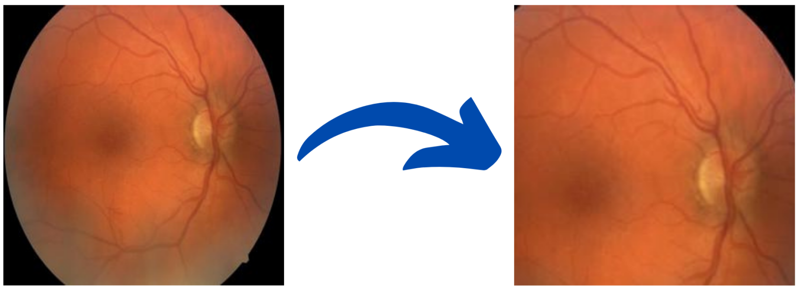


Figure 2.5: fundus images cropping

II.2.5 Image Filtering

Image filtering is a widely utilized technique for medical imaging analysis to enhance image quality, reduce noise, and extract relevant features for accurate classification. It involves applying various filters to modify the pixel values of an image, thereby improving its visual appearance and facilitating the feature extraction tasks. By employing appropriate filtering techniques, the classification of DR images can be enhanced, leading to improved accuracy and insights into the disease. The commonly used filters in DR classification are the Gaussian filter[21], Median filter[88] and the Contrast Limited Adaptive Histogram Equalization (CLAHE) filter[68]. By applying appropriate filters like the ones used in Figure 2.6, DR images can undergo significant improvement in terms of noise reduction, edge enhancement, and contrast adjustment, which in turn facilitates more accurate feature extraction and classification of DR-related abnormalities.

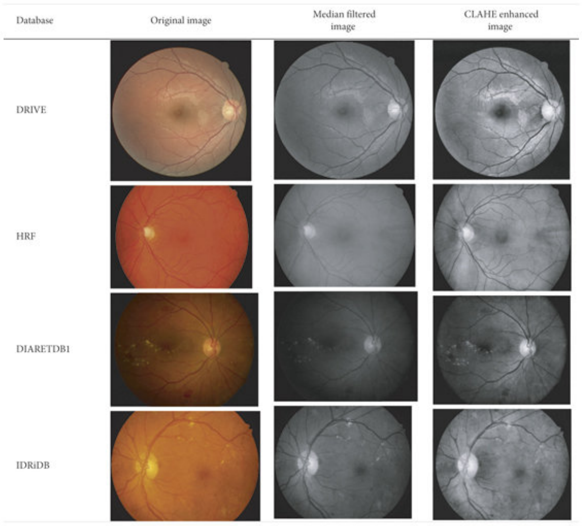


Figure 2.6: fundus images filtering

II.3 Building a model

Building a CNN model for DR classification involves several key steps. The Architecture is the first to define, and to do that we have to make some architecture choices that include: the depth and width of the network, the number of filters and their sizes, activation functions, pooling types and many more. After the architecture we need to define the hyperparameters related to the training process of the network. These are especially important since they are the ones regulating the learning aspect of the model. Hyperparameters that are commonly used in CNN architectures are: learning rate, regularization techniques, batch size and dropout. In the end we also can specify the optimization function and loss function. We can also save time and effort by using a pretrained model such as VGG16 or RESNET50 as a base model then add some layers at the end and use an automatic fine tuning algorithm to find the best hyperparameters.

II.4 Evaluating the model

The model evaluation serves as a critical step in assessing its performance and effectiveness before deployment. Various metrics are commonly employed to Evaluate the quality and predictive capabilities of models in the context of diabetic retinopathy (DR) classification. Among the widely used evaluation metrics are accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC).

Accuracy serves as an overall measure of the model's correctness in accurately predicting the correct class labels. It quantifies the percentage of correctly classified images out of the total images in the validation dataset.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \tag{2.1}$$

Precision, on the other hand, focuses on the model's ability to correctly identify positive instances, such as detecting the presence of DR in retinal images. It measures the proportion of true positive predictions out of all positive predictions made by the model.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2.2}$$

Furthermore, recall assesses the model's capability to capture all positive instances, indicating its ability to correctly detect the presence of DR. It calculates the proportion of true positive predictions out of all actual positive instances present in the evaluation dataset.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2.3}$$

The F1 score combines precision and recall, providing a harmonic mean that considers both false positives and false negatives. This metric offers a balanced assessment of the model's performance, accounting for both the capability to correctly identify positive instances and minimize misclassifications.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.4}$$

The AUC-ROC metric provides a comprehensive evaluation of the model's ability to distinguish between different classes (e.g., DR severity levels) across various classification thresholds. The AUC-ROC curve plots the true positive rate against the false positive rate, illustrating the model's discriminative power. A higher AUC-ROC value signifies a better-performing model with improved class separation.

Additionally, other evaluation metrics such as specificity, sensitivity, and the confusion matrix can offer valuable insights into the model's performance. Specificity measures the model's ability to correctly identify negative instances, while sensitivity evaluates its capacity to correctly detect positive instances. The confusion matrix provides a detailed summary of the model's predictions, indicating true positive, true negative, false positive, and false negative counts.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \tag{2.5}$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2.6}$$

By comprehensively evaluating models using these metrics, researchers and practitioners can gain valuable insights into the model's performance, strengths, and areas of improvement. These evaluations play a vital role in selecting and fine-tuning the most effective DR classification models for real-world applications.

III Related works

In this part we will present an overview on the different works related to DR detection and classification from FIs.

The earliest approaches used machine learning[95, 80, 43, 52, 69, 13, 4] based techniques such as SVM and KNN and Random Forest. Though these approaches obtained promessing results they have certain limitations when it comes to diabetic retinopathy (DR) classification. The main limitations of the traditional machine learning algorithms are difficulty handling high-dimensional data such as DR images and they rely on handcrafted or pre-defined features which is impractical.

After ML approaches, Deep learning, specifically CNNs has emerged as a powerful tool for diabetic retinopathy (DR) classification[28, 49, 6, 2]. CNNs have the ability to automatically extract features from raw image data, eliminating the need for human involvement. This is particularly advantageous in the context of DR, where intricate patterns and subtle abnormalities in retinal images are crucial for accurate classification. Another advantage of CNNs and deep learning in general is the ability to handle high-dimensional data, such as DR images, without significant computational constraints unlike traditional ML approaches. Additionally, to achieve better results some research papers suggest a combination of ML approaches and CNNs[65, 3, 55]. Despite the advantages of CNNs in DR classification, researchers have encountered certain challenges in their application like the availability of large scale, well-annotated dataset, because CNNs are data-hungry and thrive on large amounts of diverse data. They also need a long time for training which lead to the use of pretrained models and what is known as Transfer learning.

Transfer learning has been present in the majority of the latest research papers in the field of diabetic retinopathy (DR) classification due to the imbalanced datasets available and their small size.By leveraging pretrained models and their learned features, researchers can effectively address the data scarcity issue and achieve improved classification performance.

1. For instance, Kassani et al[46] used a Xception based architecture for 5 stages classification and achieved 83% in accuracy. the dataset used is The APTOS dataset[45] and data augmentation was not applied, instead L1 and L2 regularization were employed to tackle the highly imbalanced classes in the dataset.

2. In another study by Mohamed Shaban et al[73], a modified version of VGG19[76] consisting of 18 layers was employed for a 3-class classification task, achieving a validation accuracy of 88-89%. The APTOS dataset[45] was utilized, and the classes were formed by grouping the first two stages into one class and the last two stages into another.

3. An average 5 classes validation accuracy of 91.32% was reported by islam et al[40] using the VGG16[76] pretrained model and the APTOS dataset[45]. the images were PreProcessed but not augmented

4. Rao et al[67] compared multiple pretrained models and reported promising results. For binary classification, an accuracy of 95.59% using the ResNet50[32] based model, for 3 stages classification an accuracy of 88.14% using a InceptionResNetV2[33] based model and a 5 stages classification accuracy of 85 using also the InceptionResNetV2[33] based model.

5. Kaushik et al[47] stacked three CNN models and was able to achieve a binary classification accuracy of 97.92% using the EyePACS dataset[19].

6. Karki and Kulkarni[44] used an ensemble model composed of EfficientNet B1, B2, B3 and B5[79] and a dataset composed of EyePACS[19] and APTOS[45]. They achieved a 92.43% quadratic kappa score for 5-stage classification.

7. Oulhadj et al[63] used deformable registration to eliminate the background. 4 pretrained model were tested and reported a test accuracy of 85.28% using ensemble voting.

8. Athira T R et al[10] used a hyperparameter block to automatically tune the pretrained models which allowed her to achieve a binary classification accuracy of 99.8% and a 3 classes classification of 94.7% with ResNet50[32] model.

9. M A K raiaan et al[66] achieved 98.65% test accuracy for muticlasse classification (5 classes) using a RetNet10[32] model and merging multiple datasets. the dataset was augmented than split which could result in an overlap between the training and testing data.

IV Conclusion

In this chapter, We have defined diabetic retinopathy (DR) and its various stages of complications, and then presented the most recent studies that enable the classification of DR which are sumerized in Table 2.3.

These studies have demonstrated that the use of large datasets, hyperparameter optimization, and transfer learning methods can lead to very interesting results. However, it is important to consider the specific challenges of imbalanced medical datasets when applying these techniques.

Our goal is to develop a model based on a Generative Adversarial Network (GAN) that can perform DR classification. We will leverage the insights gained from these studies to propose, in the next chapter, a discriminator architecture based on transfer learning. This approach will allow us to effectively detect and classify DR.

Table 2.3: Long table caption.

Begining of Table					
Authors/Year	**Methode**	**Dataset**	**Preprocessing**	**Performace**	**Observations**
Kassani et al[46] 2019	Modified Xception	APTOS 2019 (3662 images 5 classes)[45]	Resizing and cropping (size 600*600). Min-pooling. Pixel normalization between -1 and 1.	Accuracy:83.06% Sensitivity 88.24% Specificity 87%	- The study introduced a modified version of Xception along with the aggregation of deep CNN layers. - The model achieved a satisfactory level of accuracy without requiring data augmentation.
Mohamed Shaban et al[73] 2020	Transfer learning (modified VGG19)	APTOS 2019 (3648 images 3 classes)[45]	Image resizing. Data augmentation.	89% AUC :95% Kappa Score :92%	- The achieved accuracy in this article is good. However, there is still room for improvement.
Islam et al[40] 2020	Transfer learning (VGG16)	APTOS 2019 (5590 images 5 classes)	Cropping. Gaussian blurring. Circle cropping. Resized to 512*512.	average accuracy of 91.32%	- The paper does not contain a confusion matrix or a performance metric for unseen data due to the absence of a test set.
Rao et al[67] 2020	Transfer learning	APTOS 2019[45]	Cropping Gaussian blurring Circle cropping Data augmentation Resized to 512*512	**binary:** Accuracy:96.59% Precision: 97% Recall: 97% F1 Score : 96,5% **3-stage:** Accuracy:88.14% Precision:88% Recall: 88% F1 Score : 88,2% **5-stage:** Accuracy:85,02% Precision: 85% Recall: 85% F1 Score : 84,95%	- The results reported for all types of classification were good but can be improved.

Continuation of Table 2.3

Kaushik et al[47] 2021	Stacked CNNs	EyePACS(2471 images 2 classes)[19]	Luminosity normalization Data aumentation	Accuracy:97.77%	- The paper's main focus was on luminosity normalization to address issues associated with multi-sourced images. - However, the authors acknowledged a limitation in the classification of mild diabetic retinopathy (DR) (class 1).
Karki and Kulkarni[44] 2021	ensemble learning and transfer learning	APTOS[45] and EyePACS[19] datasets (5classes)	Down sampling. Data augmentation. Resizing (256*256)	Kappa score:92.43%	- The authors of this paper employed ensemble learning to enhance their model's performance. - However, the improvement achieved through ensemble learning was marginal, with only a 1% increase compared to the best model (EfficientNet B3)[79]. - It's worth noting that implementing ensemble learning can demand significant computational resources.

Continuation of Table 2.3

Oulhadj et al[63] 2022	Transfer learning with ensemble voting for testing	APTOS 2019 (3662 images 5 classes)[45]	Deformable registration	Accuracy:85% Specificity:94% Precision:86% Recall 85% F1-score:84%	- The authors of this paper introduced deformable registration as a technique to mitigate the impact of background space on the dataset images. - By utilizing ensemble voting, they were able to achieve improved results compared to individual models or algorithms. - The combination of deformable registration and ensemble voting proved effective in enhancing the overall performance of the proposed approach.
Athira T R et al[10] 2023	Transfer learning with automatic hyperparameter tuning	APTOS 2019 (3662 images 3 classes)[45]	Image resizing grey scaling gaussian blurring circular cropping	**binary:** Accuracy:99.8% Precision : 99% Recall : 99% F1-score : 99% **3-stage:** Accuracy:94.7% Kappa Score :88.4% Precision : 94% Recall : 94% F1-score : 93%	- The ResNet[32] based networks achived promising results thanks to hyperparameter tuning and skip connected networks.

Continuation of Table 2.3

M A K raiaan et al[66] 2023	Transfer learning (ResNet10)[32]	APTOS[45], Messidor2[20], and IDRiD[64] (5819 images 5 classes)	Artifact removing. Extraction of Regions of Interest. Denoising (NLMD). Image inhancement using CLAHE. Resized to 512*512. Data Augmentation.	Accuracy:98.65% Specificity:99.66% Precision:98.65% Recall 98.65% F1-score:98.65%	- Preprocessing techniques were applied to enhance image quality. - Geometric Augmentation, Elastic Deformation, and Photometric Augmentation were used to augment the dataset. - It is worth noting that the augmentation was performed prior to splitting the dataset, which could result in an overlap between the training and testing data.

End of Table

Chapter 3

GAN: Conception & Realization

I Introduction

Addressing the challenge of insufficient data in medical image databases, particularly in the field of ophthalmology, we propose a GAN-based architecture. This innovative approach not only aims to increase the available data but also offers preprocessing capabilities and facilitates classification based on retinopathy grade.

In this chapter, we will provide a detailed overview of the GAN architecture, including the individual architectures of the discriminator and generator components.

II Generative Adversarial Networks Architecture (GAN)

The GAN architecture, as depicted in the accompanying figure 3.1, comprises a generator and a discriminator that train together to produce high-quality images and classify them according to the degree of diabetic retinopathy (DR). Consequently, the generator of our proposed approach adopts an encoder-decoder architecture to improve the quality of real fundus images. On the other hand, the discriminator is a multi-class convolutional neural network employed for the classification of diabetic retinopathy, taking advantage of transfer learning techniques.

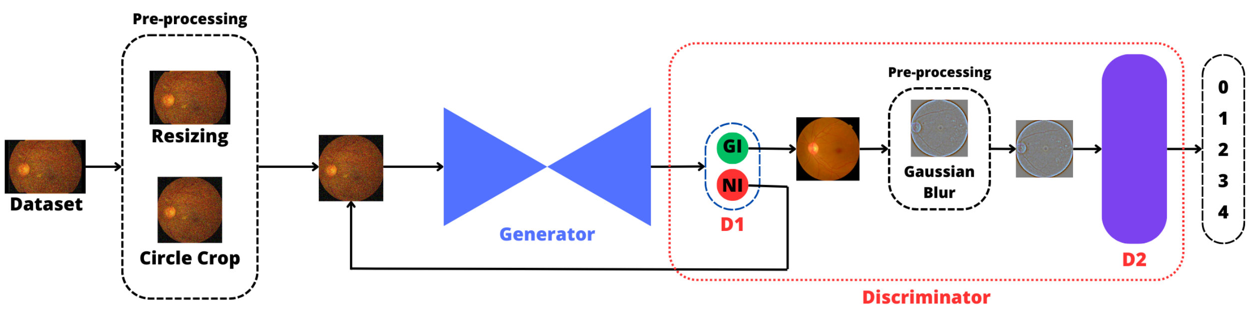
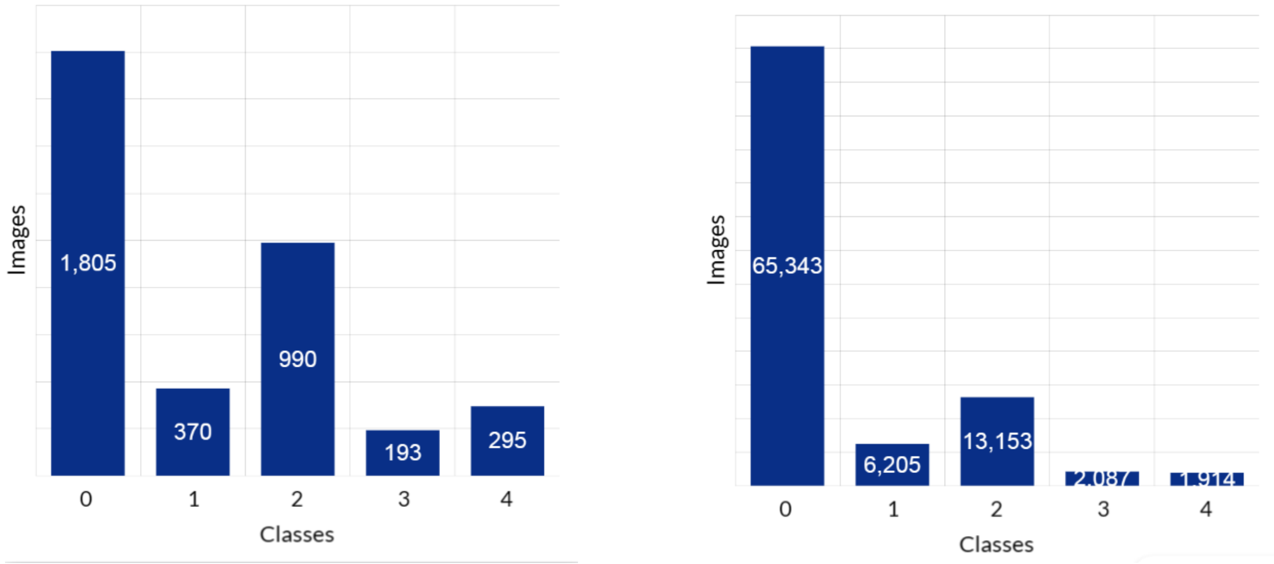


Figure 3.1: Global Architecture of the GAN

II.1 Data Preparation

II.1.1 Datasets

All the models we developed were trained using the publicly available APTOS [18] and EYEPACS [12] databases, which respectively contain 3662 and 88507 images. Both datasets are divided into five classes: No DR (0), Mild (1), Moderate (2), Severe (3), and Proliferative DR (4). Figure 3.2 show the distributions of the databases.



(a) Aptos

(b) EyePacs

Figure 3.2: Datasets classes Divisions

II.1.2 Images Preprocessing

Due to the large amount of data, specific preprocessing techniques are selected to enhance features such as blood vessels, red lesions, and exudates. This is illustrated in Figure x.

2.a Resizing

Since the images in the datasets were collected from different sources and time periods, they had varying sizes and storage capacities. Therefore, in order to achieve optimal performance during processing, it was necessary to resize all the images to a fixed size of 224x224.

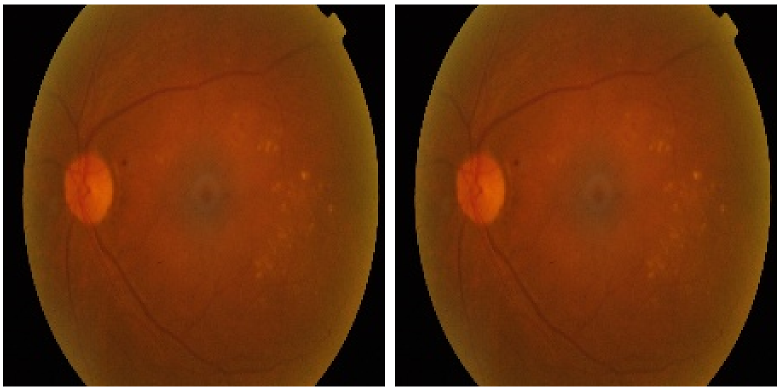
2.b Grey Scale & Gaussian Blur

To achieve better results in the training model, it is important that the images and their features are easy to extract. Therefore, the first technique used for feature enhancement was to convert the color images to grayscale images and then apply a Gaussian blur filter to reduce image noise. To sharpen the edges of the image obtained after the Gaussian blur, a weighted masking method is used. The first processed image is the resized one, and the second processed image is the one obtained after applying the Gaussian blur [10].

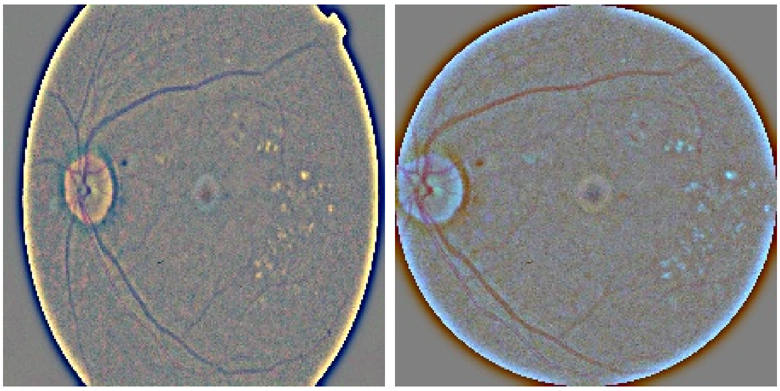
Images with both exudate and red lesion features, and particularly blood vessels, were enhanced compared to the original images.

2.c Circle Crop

To improve performance during model training, a circular cropping technique was used to remove unwanted black pixels (the background) and focus more on the part of the image that contains the fundus of the eye. To apply circular cropping, a mask is created and applied to the image to retain only the pixels within the circle while discarding the rest [10].



(a) Original (b) Resizing



(c) Gaussian Blur (d) Circle Crop

Figure 3.3: Preprocessing Phases

2.d Normalization

To facilitate the learning procedure of our model, we employed a normalization technique on the input images. Specifically, we chose to convert the pixel values of the images from the original range of 0 to 255 to a normalized range of 0 to 1 as shown in figure 3.4. This normalization process ensured that all pixel values were scaled appropriately, allowing for more efficient and stable training of the model. By normalizing the pixel values in this manner, we effectively reduced the potential impact of varying pixel intensity ranges, thereby improving the convergence and performance of our model.

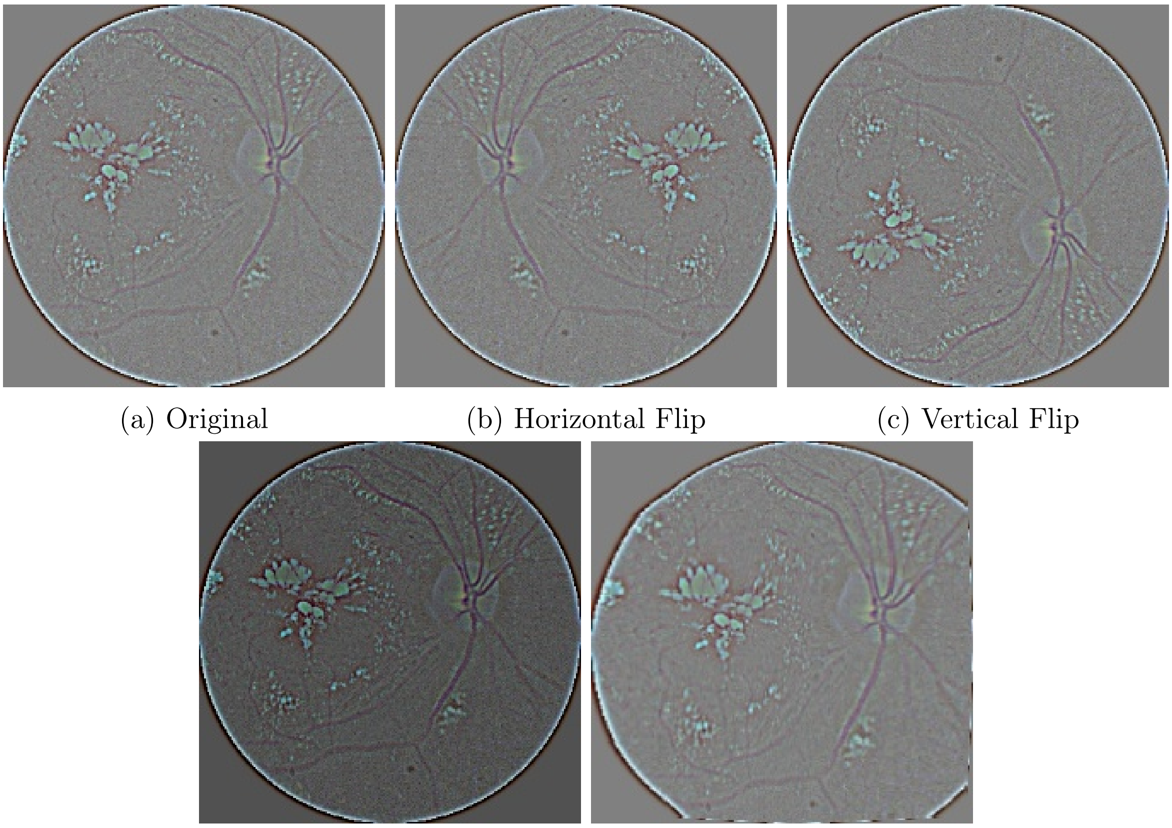
Figure 3.4: Normalization method

II.1.3 Augmentation

Due to the presence of imbalanced classes in the APTOS dataset, we employed various augmentation techniques to expand the dataset and achieve a more balanced representation of the different stages of retinopathy.

We applied multiple augmentation functions to each original image, generating a total of five augmented images for every input. These augmentation functions included grid distortions, horizontal and vertical flips, and brightness adjustments. By leveraging these transformations, we aimed to create additional variations of the original images while preserving the inherent features of retinopathy.

To illustrate the effectiveness of our augmentation procedure, Figure 3.5 showcases a subset of the images generated through the augmentation process.



(a) Original (b) Horizontal Flip (c) Vertical Flip

(d) Brightness (e) Grid Distortion

Figure 3.5: Augmentation Phases

II.2 Discriminator Architecture

In the field of Generative Adversarial Networks (GANs), the role of the Discriminator as a binary classifier, distinguishing real data from generated data, is well-established. However, extending the capabilities of the Discriminator to act as a multi-class classifier represents a significant advancement. This extension not only allows for determining the authenticity of generated data but also classifying them into different categories, thus offering new possibilities for the use of GANs in multi-class classification problems.

In this section, we delve into the detailed design and training of Discriminators specifically designed for the detection (diseased or non-diseased) of retinopathy and the classification into three and five retinopathy stages. We discuss the architecture and models used, the training process, and the obtained results as illustrated in figure 3.6.

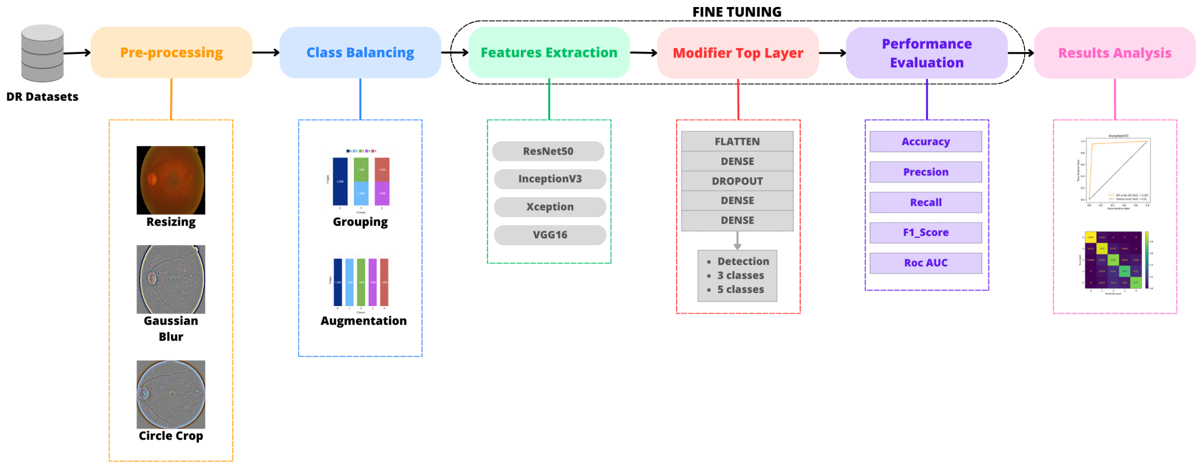


Figure 3.6: Proposed Architecture PA

II.2.1 Methodology

The methodology employed in this study involves the utilization of two datasets, namely the Aptos and EyePACS datasets, for the detection and classification of retinopathy. The overall approach consists of a preprocessing phase, followed by a fine-tuning phase, model comparison, and final testing. As shown in the figure 3.6 and detailed in figure 3.7.

In the preprocessing phase, the collected data from both datasets is preprocessed to ensure compatibility and consistency. This includes image resizing, normalization, and data augmentation techniques to enhance the quality and variety of the dataset. Subsequently, the preprocessed data is split into an 80% training set and a 20% testing set.

Next, the fine-tuning phase begins, where pre-trained models including ResNet50, Xception, Inception, and VGG16 are selected as base models. The aim is to search for the optimal parameters for each model using the training set. Fine-tuning involves adjusting the weights and biases of the pre-trained models to suit the retinopathy detection task. This process is performed using techniques such as gradient descent and backpropagation to optimize the models for improved performance.

Following the fine-tuning phase, the models are trained using the training set with the refined parameters. The training results of each model are then compared and evaluated to identify the best performing model in terms of accuracy and other relevant metrics. The performance comparison helps determine the model that exhibits the highest potential for accurate retinopathy detection and classification.

Finally, the best-performing model identified during the comparison stage is subjected to testing using the separate testing set. This evaluation phase assesses the model's ability to generalize and accurately classify retinopathy cases in unseen data. The performance of the model is evaluated based on various metrics such as accuracy, precision, recall, and F1 score.

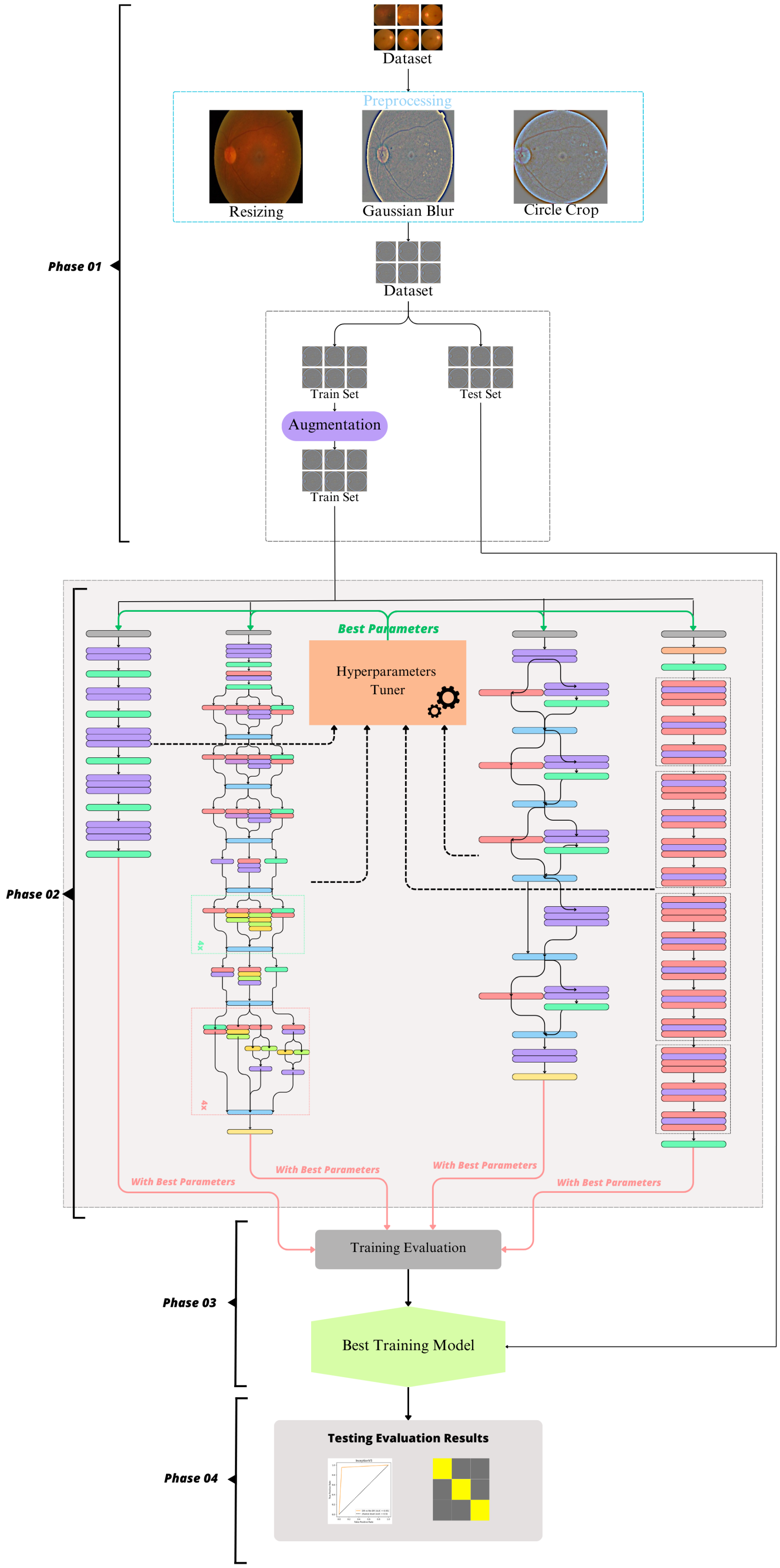


Figure 3.7: Global schema of PA.

II.2.2 Fine Tuning

In our approach, fine-tuning played a pivotal role in optimizing the performance of the pretrained models, namely ResNet50, VGG16, Xception, and InceptionV3, for retinopathy detection and classification. We meticulously searched for the best parameters specific to each model and dataset, including image size, batch size, learning rate, and more.

To begin, we conducted an extensive parameter search to find the optimal configuration for each pretrained model. This involved iteratively adjusting and fine-tuning the parameters while training on the respective datasets, Aptos and EyePACS. The goal was to strike a balance between model complexity and computational efficiency, ensuring optimal performance.

Throughout the fine-tuning process, we carefully monitored the training performance of each model. By evaluating various metrics such as accuracy, loss, precision, and recall, we gained insights into the effectiveness of the models at different parameter settings. This rigorous evaluation enabled us to select the best performing model for each pretrained model.

Ultimately, after thorough training and evaluation, we made a definitive choice by selecting the best model among the four pretrained models. This decision was based on the model's performance in terms of accuracy and its ability to generalize well on both the training and testing datasets. The selected model showcases the finest combination of pretrained models and optimized parameters shown in table table 3.1, providing the highest accuracy for retinopathy detection and classification.

Table 3.1: Fine tuned hyeperparameters

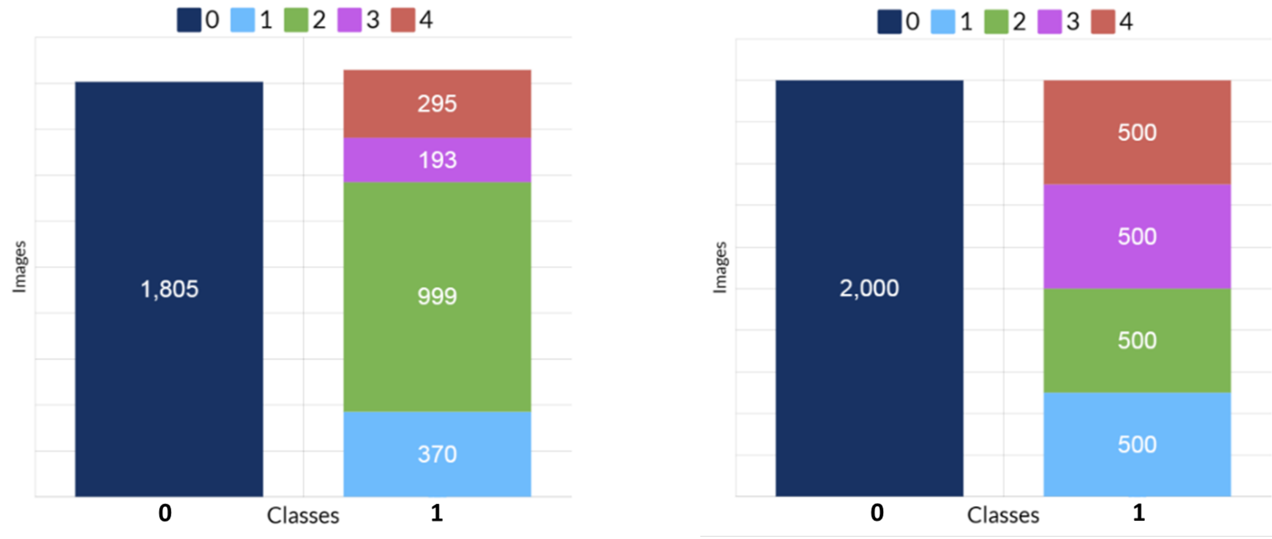
Parameter	Value
Validation Set	20% of dataset
Test Set	30% of validation set
Image Size	224x224
Batch Size	32
Warmup epochs	5
Warmup learning rate	0.00001
Epochs	50
Learning rate	0.0001
Wieght decay	0.02
Early stopping patience	15
Reduce LR patience	5
Regularizer	0.02

II.2.3 DR Detection

In this section, we present the architecture and parameters used in our published article for the detection of retinopathy, along with the results obtained in our study. The objective of our research was to develop an accurate and efficient system for identifying signs of retinopathy in fundus images.

3.a Datasets

Since the objective of our study is to detect retinopathy, we only require class 0 (No DR) and class 1, which combines all other classes (1, 2, 3, 4). Figure 3.8 illustrate how we combine the classes.



(a) Aptos (b) EyePacs

Figure 3.8: Datasets classes Divisions

3.b Architecture

The proposed approach (PAD) aims to detect retinopathy in its initial stage. As shown in Figure 3.9, it involves merging classes 1, 2, 3, and 4 into a single class (1) and a preprocessing phase. After applying the fine-tuning of the best pretrained networks, ResNet50 and Xception are the top performers for the APTOS and EYEPACS datasets, respectively. As mentioned before, a set of fully connected neural networks is added at the end of each pretrained network for the detection of DR disease.

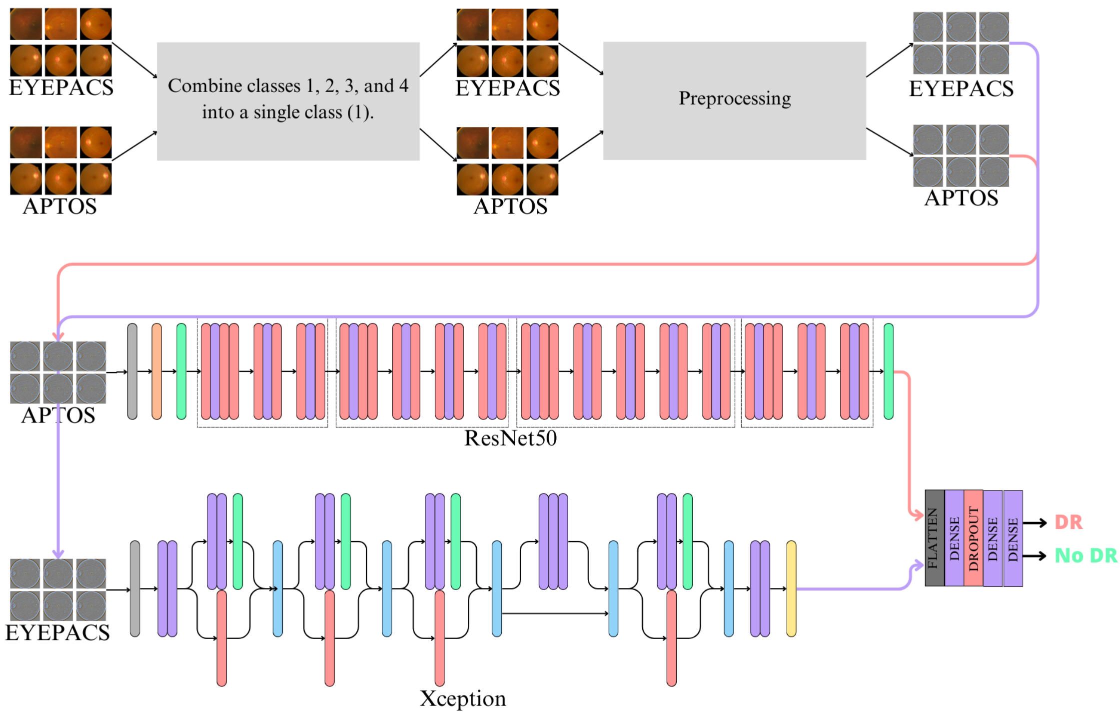


Figure 3.9: Proposed DR detection architecture: PAD

3.b Results

The experiment was conducted with two different datasets to evaluate the models' disease detection capabilities. A test set comprising 220 images from the APTOS dataset and 240 images from the EYEPACS dataset was chosen to evaluate the models.

As shown in Tables 3.2 and 3.3 , the training results using our method on the APTOS dataset achieved an impressive accuracy of 99.22% using the ResNet50 model. However, on the EYEPACS dataset, the accuracy was lower at 71.79% using the Xception model.

Table 3.2: Aptos Testing results

	ResNet50	
APTOS	Accuracy	**0.9922**
	Precision	**0.9961**
	Recall	**0.9923**
	F1_score	**0.9941**

Table 3.3: EyePacs Testing results

	Xception	
EyePacs	Accuracy	**0.7179**
	Precision	**0.7117**
	Recall	**0.6496**
	F1_score	**0.6792**

To have an overall understanding of the model's prediction performance, a confusion matrix and ROC curve were generated for each model. These visualizations are shown in Figures 3.10, 3.11, 3.12, and 3.13. The confusion matrix provides insights into the model's classification accuracy for each class, while the ROC curve illustrates the trade-off between the true positive rate and false positive rate at different classification thresholds. These visualizations help evaluate the models' performance and their ability to accurately detect diabetic retinopathy.

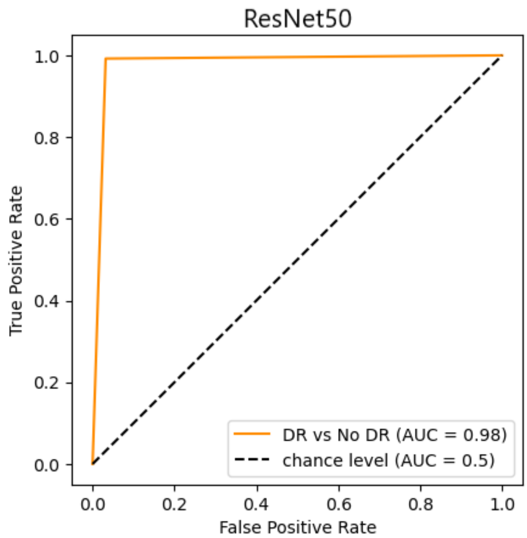


Figure 3.10: Roc curve for Aptos model

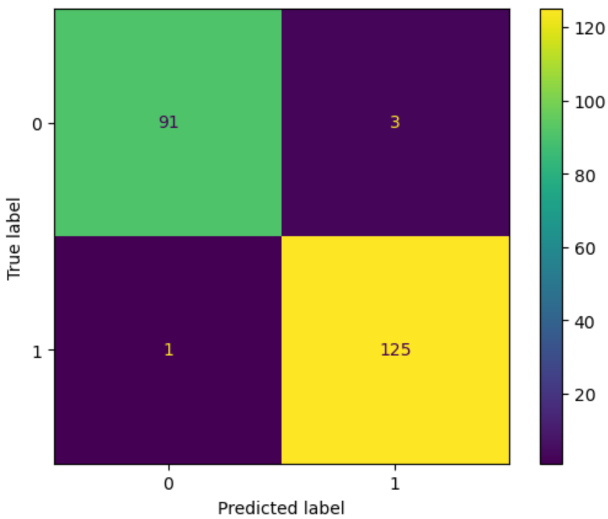


Figure 3.11: Confusion matrix for Aptos model

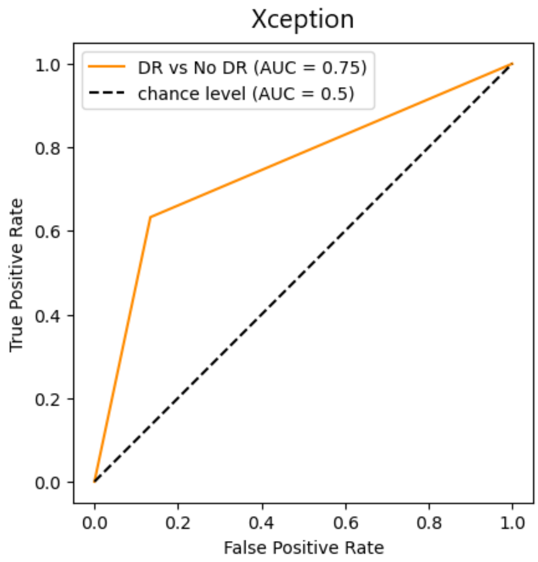


Figure 3.12: Roc curve for Aptos model

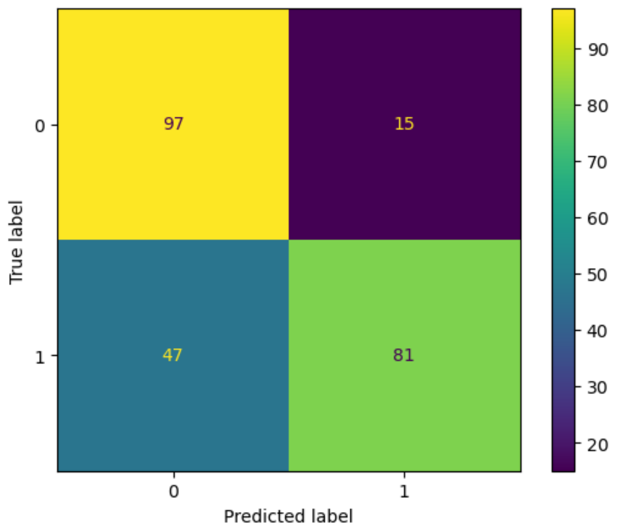


Figure 3.13: Confusion matrix for Aptos model

We compare our results to other existing methods mentioned in the related works section, demonstrating that our approach outperforms them. This comparison is shown in Table 3.4 and Figure 3.14. The results clearly indicate that our Proposed Approach (PAD) achieves higher accuracy and better performance in detecting diabetic retinopathy from fundus images compared to the existing methods. These findings further support the effectiveness of our approach in accurately identifying and classifying retinopathy cases.

Table 3.4: PAD numerical comparaison results

metrics	**[26]**	**[87]**	**[47]**	**[41]**	**[67]**	**[10]**	**PAD**
Accuracy	86%	94%	97.99%	98.36%	96.56%	99.8%	99.22%
Precision	85%	/	100.00%	98.36%	97%	99%	99.61%
Recall	86%	/	96.00%	98.36%	97%	99%	99.23%
F1_Score	85%	/	97.90%	98.37%	96.5%	99%	99.41%

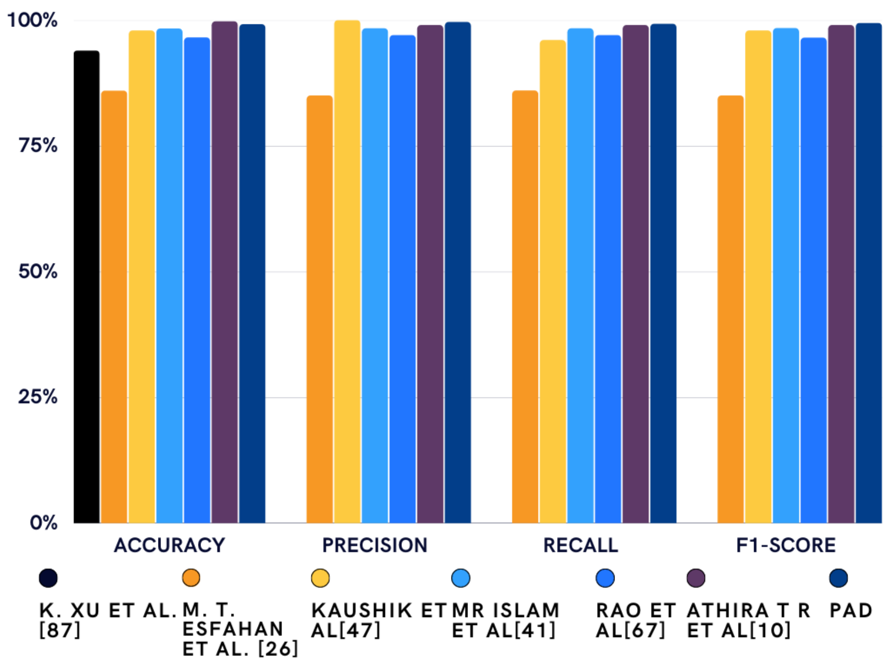


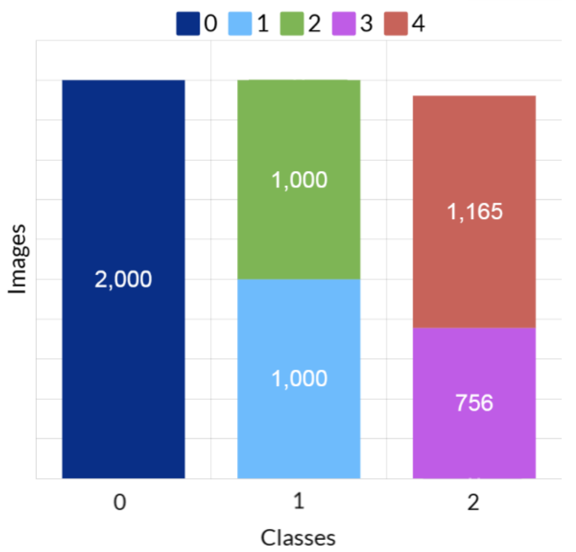
Figure 3.14: PAD results comparison to previous works

II.2.4 DR classification to 3 stages

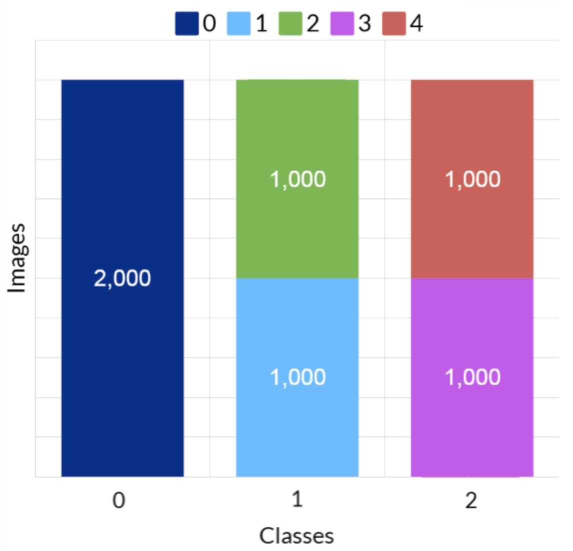
In this section, we present the architecture and parameters used for the classification of retinopathy into 3 classes (normal, early, late), as well as the results obtained in our study. The objective of our research was to develop an accurate and efficient system for identifying signs of retinopathy in fundus images.

a. Datasets

Since the objective of our study in this section is to classify retinopathy into 3 classes, we only need classes 0 (No DR),1 (early) and 2 (late), while the class 1 combines classes 1 and 2 and the class 2 combines classes 3 and 4. Figures 3.15 illustrate how we combined the classes.



(a) Aptos



(b) EyePacs

Figure 3.15: Datasets classes Divisions

b. Architecture

The proposed approach (PA3C) aims to classify retinopathy into 3 stages: Healthy, Mild, and Severe. As illustrated in Figure 3.16, it involves grouping classes 1 and 2 into one class (1), and classes 3 and 4 into another class (2), as well as a preprocessing phase for the images. After applying fine-tuning to the best pretrained networks, ResNet50 and Xception perform the best on the APTOS and EYEPACS datasets, respectively, as mentioned before. A fully connected neural network block is added at the end of each pretrained network for the classification of DR.

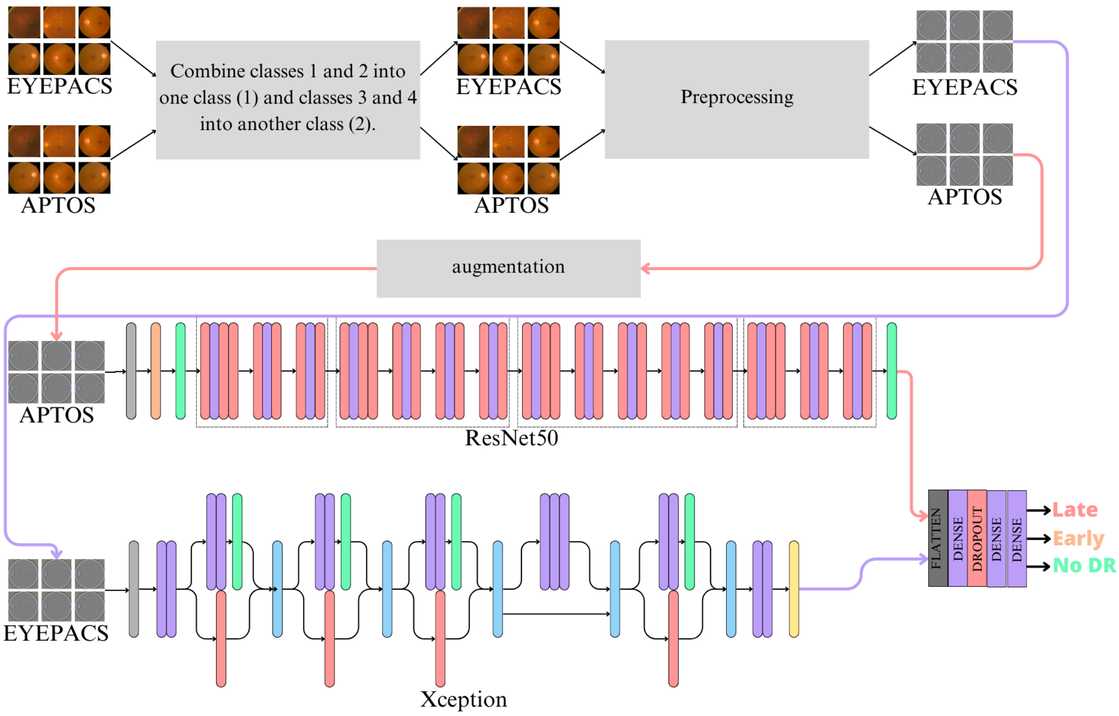


Figure 3.16: DR 3 classes : PA3C architecture

c. Results

Like previous experiment this experiment was conducted using two different databases to evaluate the two models for their disease detection capabilities. A test set consisting of 1184 images from the APTOS dataset and 1200 images from the EYEPACS dataset was chosen to test the models. As shown in Tables 3.5 and 3.6, the training phase results using our method achieved an impressive accuracy of 94.26% on the APTOS dataset using ResNet50, while on the EYEPACS dataset, the accuracy was lower at 67% using Xception.

Table 3.5: Aptos Testing results

		ResNet50
APTOS	Accuracy	**0.9426**
	Precision	**0.9441**
	Recall	**0.9409**
	F1_score	**0.9424**

Table 3.6: EyePacs Testing results

		Xception
EyePacs	Accuracy	**0.6700**
	Precision	**0.6713**
	Recall	**0.6858**
	F1_score	**0.6735**

To provide an overview of the model's performance in terms of prediction, a confusion matrix and a metrics table were created for each model, as shown in tables 3.7 and 3.8 and figures 3.17 and 3.18.

Table 3.7: Performance measures with Aptos

	Classes	Precision	Recall	F1_Score
APTOS	0	93%	97%	95%
	1	82%	80%	81%
	2	86%	85%	85%

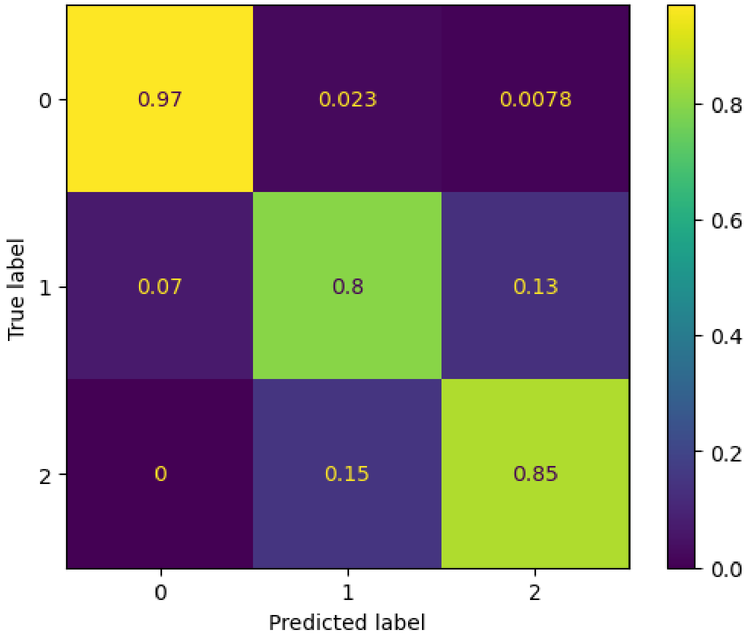


Figure 3.17: Confusion matrix for 3 classes Aptos model

Table 3.8: Performance measures with EyePacs

	Classes	Precision	Recall	F1_Score
EyePacs	0	72%	67%	69%
	1	52%	67%	59%
	2	84%	68%	75%

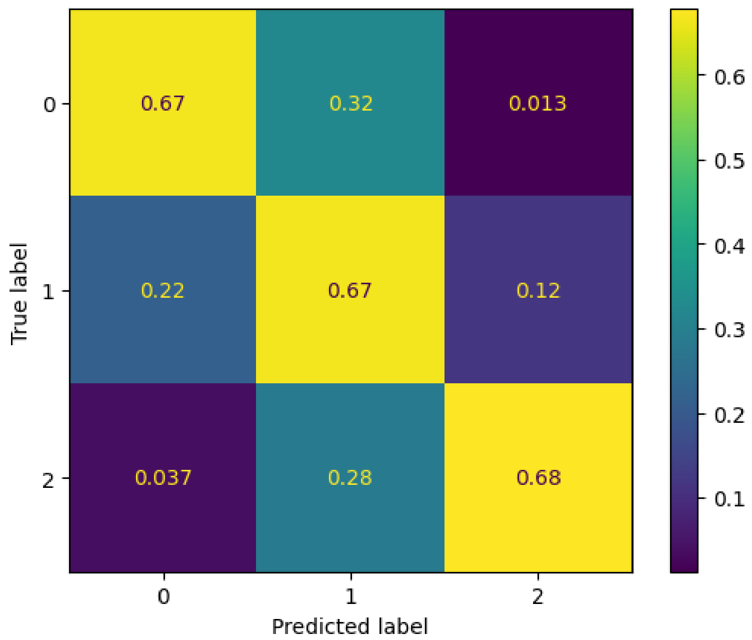


Figure 3.18: Confusion matrix for 3 classes EyePacs model

As shown in table 3.9 Our approach demonstrates favorable results when compared to other existing methods mentioned in the related works section. The performance metrics, including accuracy, precision, and recall, indicate the effectiveness and robustness of our method. These findings validate the efficacy of our approach and its competitive advantage over the state-of-the-art methods

Table 3.9: Performance comparaison results

metrics	**[67]**	**[10]**	**[73]**	**PA3C**
Accuracy	88.14%	94.9%	89%	94.26%
Precision	88%	94%	/	94.41%
Recall	88%	94%	/	94.09%
F1_Score	88.2%	93%	/	94.24%

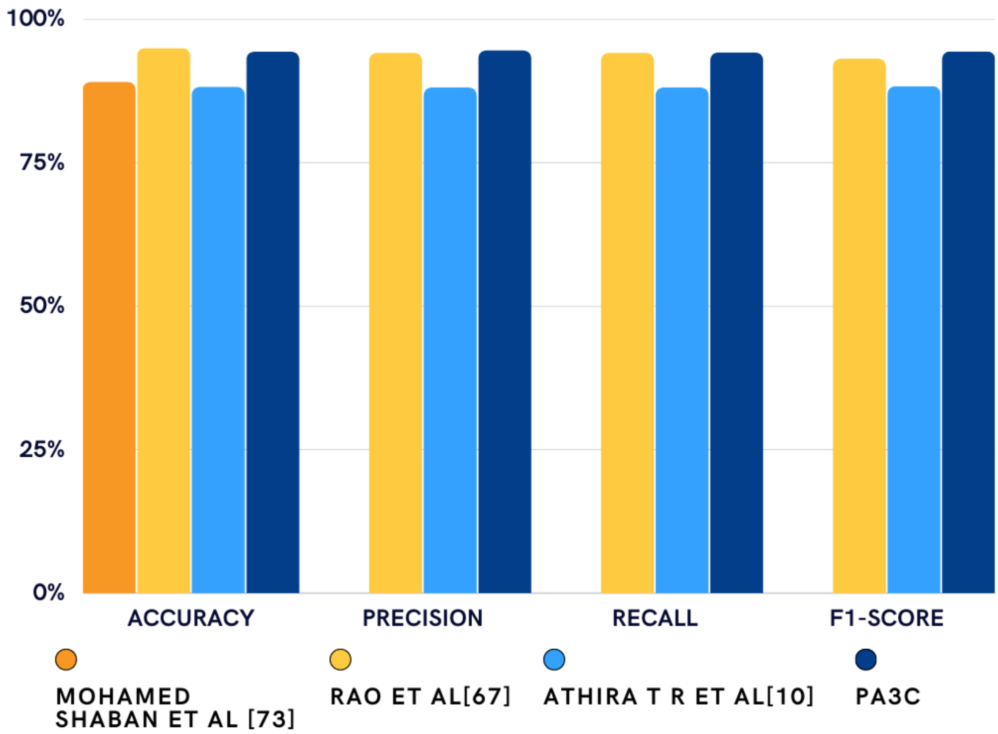


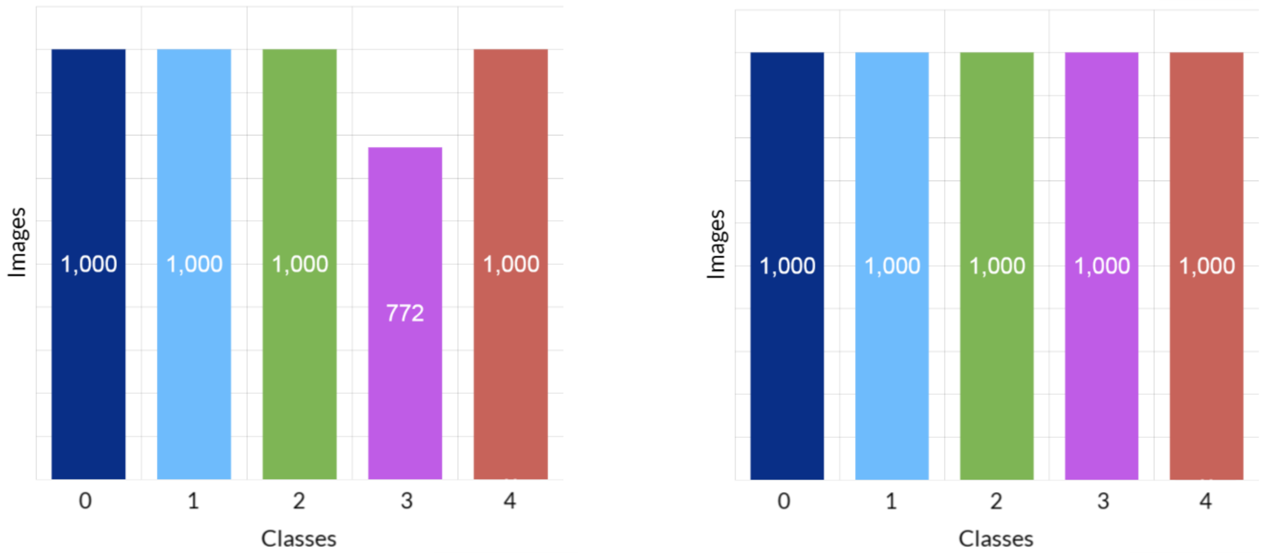
Figure 3.19: PA3C Results comparison to previous works

II.2.5 DR classification to 5 stages

In this section, we present the architecture and parameters used for the classification of retinopathy into 5 classes, as well as the results obtained in our study. The objective of our research was to develop an accurate and efficient system for identifying signs of retinopathy in fundus images.

a. Datasets

Given that the objective of our study in this section is to classify retinopathy into 5 classes, we need all the classes from 0 to 4, Due to the lack of images in some classes in APTOS dataset we applied the augmentation technique to achieve a balanced dataset. Figure 3.20 illustrate the number of images used in this study.



(a) Aptos (b) EyePacs

Figure 3.20: PA5C datasets classes Divisions

b. Architecture

The proposed approach (PA5C) aims to classify retinopathy into 5 stages: Healthy (No DR), Mild, Moderate, Severe and PDR (Proliferative DR). As illustrated in Figure 3.21, it involves a preprocessing phase for the images. After applying fine-tuning to the best pretrained networks, ResNet50 and VGG16 perform the best on the APTOS and EYEPACS datasets, respectively, as mentioned before. A fully connected neural network block is added at the end of each pretrained network for the classification of DR.

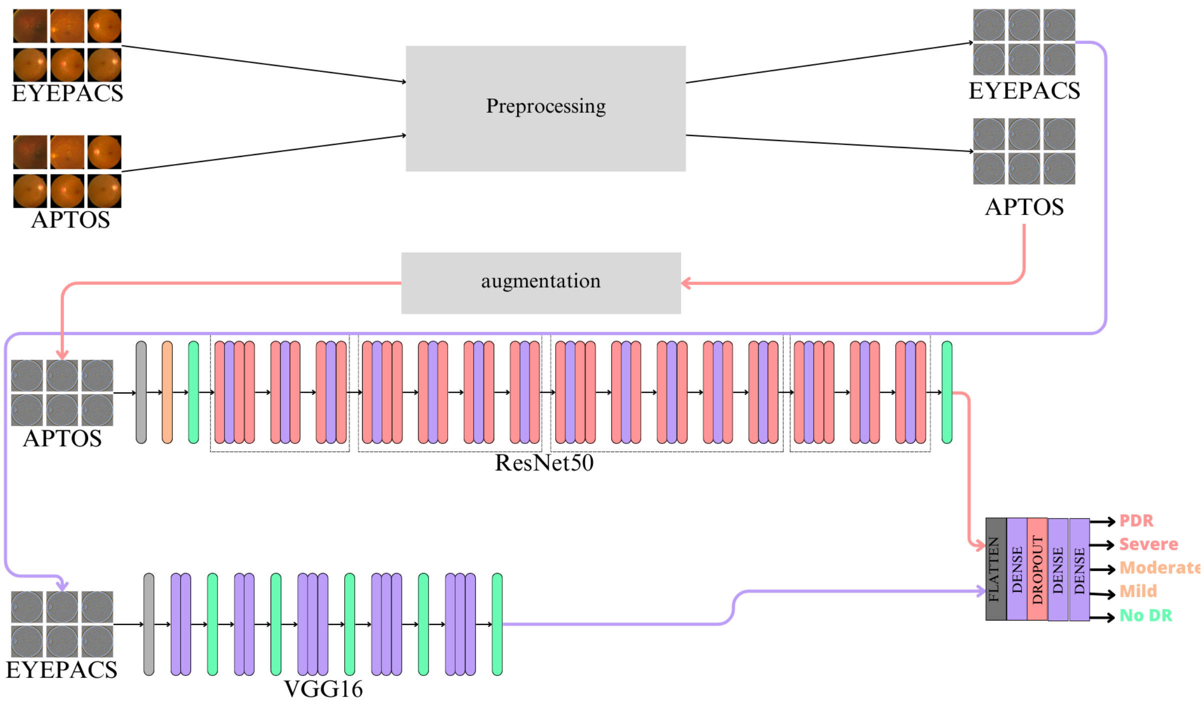


Figure 3.21: DR 5 classes architecture: PA5C

c. Results

Like previous experiment this experiment was conducted using two different databases to evaluate the two models for their disease detection capabilities. A test set consisting of 1184 images from the APTOS dataset and 1200 images from the EYEPACS dataset was chosen to test the models.

As shown in Tables 3.10 and 3.11, the training phase results using our method achieved an acceptable accuracy of 85.42% on the APTOS dataset using ResNet50, while on the EYEPACS dataset, the accuracy was lower at 47.4% using VGG16.

Table 3.10: Aptos Testing results

	ResNet50	
APTOS	Accuracy	**0.8542**
	Precision	**0.8567**
	Recall	**0.8508**
	F1_score	**0.8537**

Table 3.11: EyePacs Testing results

	Xception	
EyePacs	Accuracy	**0.4740**
	Precision	**0.5136**
	Recall	**0.3780**
	F1_score	**0.4354**

To provide an overview of the model's performance in terms of prediction, a confusion matrix and a metrics table were created for each model, as shown in tables 3.12 and 3.13 and Figures 3.22 and 3.23.

Table 3.12: Performance measures with Aptos for RD 5 classification

	Classes	**Precision**	**Recall**	**F1_Score**
APTOS	0	98%	96%	97%
	1	77%	91%	83%
	2	81%	79%	80%
	3	75%	69%	72%
	4	82%	77%	79%

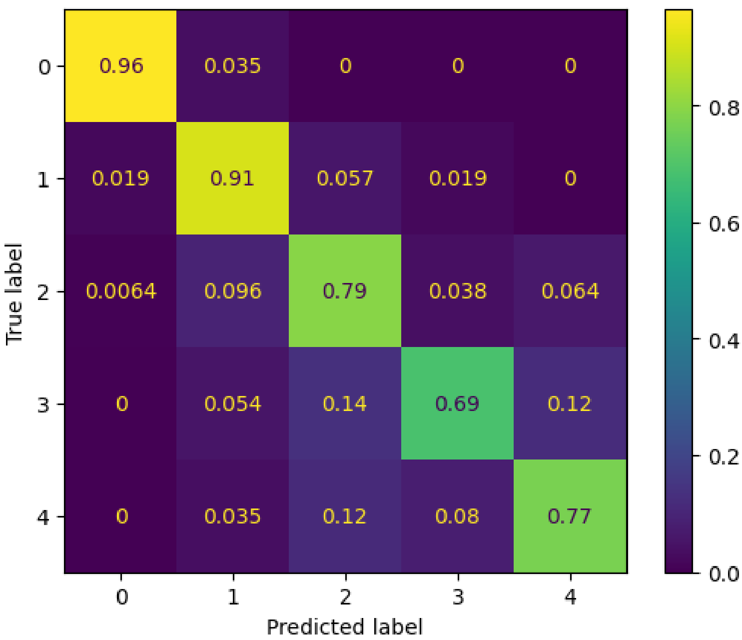


Figure 3.22: Confusion matrix for 5 classes Aptos model

Table 3.13: Performance measures with EyePacs for RD 5 classification

	Classes	Precision	Recall	F1_Score
EyePacs	0	45%	30%	36%
	1	37%	58%	45%
	2	33%	37%	35%
	3	57%	45%	50%
	4	69%	61%	65%

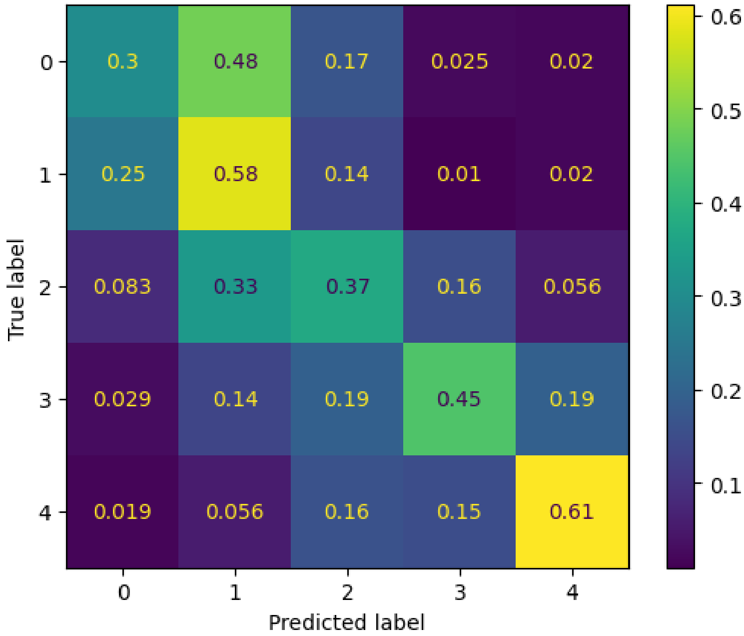


Figure 3.23: Confusion matrix for 5 classes EyePacs model

We conducted a comprehensive comparison of our results with other existing methods discussed in the related works section. The detailed comparison is presented in Tables 3.14 and 3.15 and Figure 3.24. Our Proposed Approach (PA) demonstrates promising results for the classification of diabetic retinopathy. These findings validate the effectiveness and robustness of our approach, positioning it as a valuable solution for diabetic retinopathy diagnosis and classification.

Table 3.14: Performance comparison of PA5C and related works

metrics	[46]	[67]	[63]	PA5C
Accuracy	83.06%	85.02%	85%	85.42%
Precision	/	85%	86%	85.67%
Recall	/	85%	85%	85.08%
F1_Score	/	84%	84%	85.37%

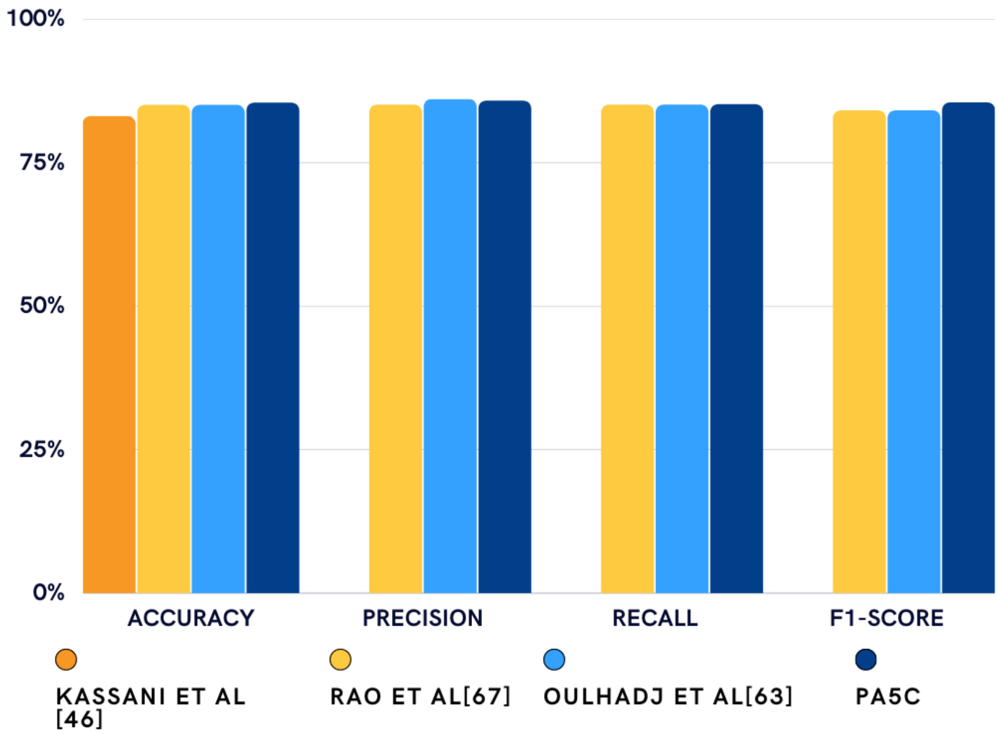


Figure 3.24: PA5C Results comparison to previous works

Table 3.15: results comparison based on class recall

classes	**0**	**1**	**2**	**3**	**4**
Rao et al[67]	96%	91%	72%	81.5%	81.6%
Oulhadj et al[63]	100%	55%	91%	53%	50%
PA5C	96%	91%	79%	69%	77%

II.2.6 Conclusion

Our experiments have shown that models trained with the APTOS dataset consistently achieve impressive accuracy in the detection and classification of retinopathy, while models trained with the EyePACS dataset exhibit significantly lower accuracy. This discrepancy can be attributed to the high-quality images provided by the APTOS dataset, which allow for better identification of patterns and features relevant to retinopathy.

To address the challenge of varying image quality in different datasets, our next step is to explore the implementation of a GAN generator. By using a GAN, we aim to enhance the image quality in datasets like EyePACS, bridging the gap with higher-quality datasets such as APTOS. This approach holds the potential to improve the accuracy and performance of retinopathy detection and classification models across different datasets.

In the upcoming section, we will focus on the development and implementation of the GAN generator as a solution to the issue of differing image quality. By standardizing and improving the quality of images, we anticipate more consistent and accurate results in retinopathy analysis.

II.3 Generator

In this section, we introduce the architecture of the generator. The main function of the generator network is to generate high-quality images using random noise or latent vectors as input. Through training on extensive datasets, the generator learns to capture intricate patterns and structures, resulting in generated images that closely resemble real samples. The objective of the generator is to produce visually appealing images that exhibit enhanced fidelity and reduced noise artifacts. By incorporating innovative architectural designs and optimization strategies, this section investigates the potential of the generator network in enhancing image quality and eliminating noise from DR fundus images.

II.3.1 Data Preparation

a. Dataset

The dataset used in this section was a carefully curated subset of the Aptos dataset, specifically tailored for training our GAN. This subset consisted of 750 images, with equal representation from each class of Diabetic Retinopathy (DR). To ensure a balanced distribution, we selected 150 images from each DR classes as illustrated in figure 3.25.

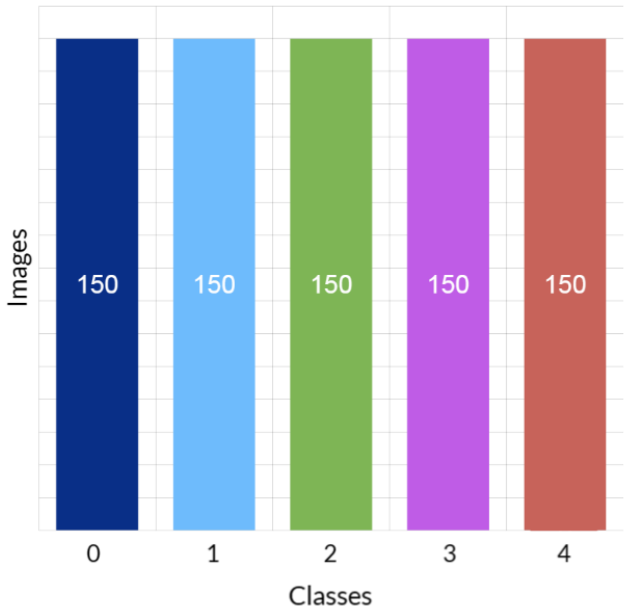


Figure 3.25: GAN training dataset subset

b. Pre-processing

In this section, we employed several preprocessing techniques to enhance the quality and usability of the dataset. Specifically, we applied resizing,circle cropping and normalization.These preprocessing techniques, discussed in detail in the previous section.

As we need to fed our generator with the original image and a noised one we used another prepossessing technique to noise the images named gaussian noise filter as illustrated in figure 3.26.

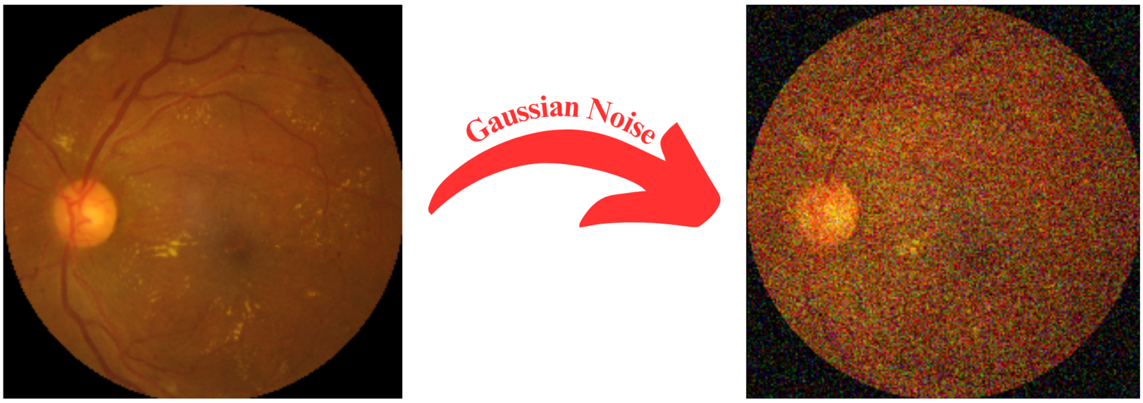


Figure 3.26: Gaussian Noise Filter

II.3.2 Architecture

In this section, we adopted a specific architectural design for both the generator and discriminator (D1) networks to accomplish our goal of denoising images. For the generator, we utilized a sequence of Conv2D layers with decreasing filter sizes (128, 64, 32, 16, 3). Each layer was accompanied by a LeakyReLU activation function, except for the final layer which employed a sigmoid activation function. The absence of an encoder and decoder architecture was intentional as it preserved image details and prevented loss of crucial information during the denoising process.

In contrast, for D1, we leveraged a pretrained ResNet50 model as a base and augmented it with an additional dense block. The dense block consisted of a flatten layer, followed by two dense layers. The first dense layer contained 16 units with a ReLU activation function, facilitating non-linearity and feature extraction. The second dense layer had a single unit with a sigmoid activation function, enabling binary classification between noised (NI) and good images (GI).

For a visual representation of the architecture, refer to Figure 3.27 and 3.28, which provide a clear illustration of the generator and D1 architectures.

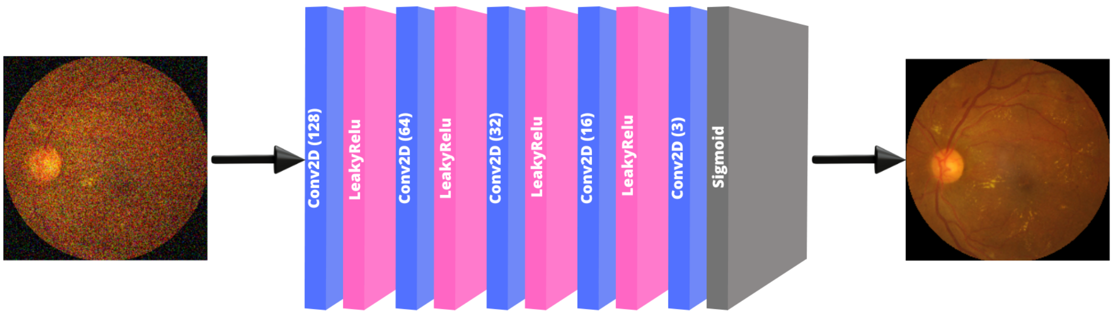


Figure 3.27: Generator Architecture

Figure 3.28: Discriminator(D1) Architecture

II.3.3 Methodology

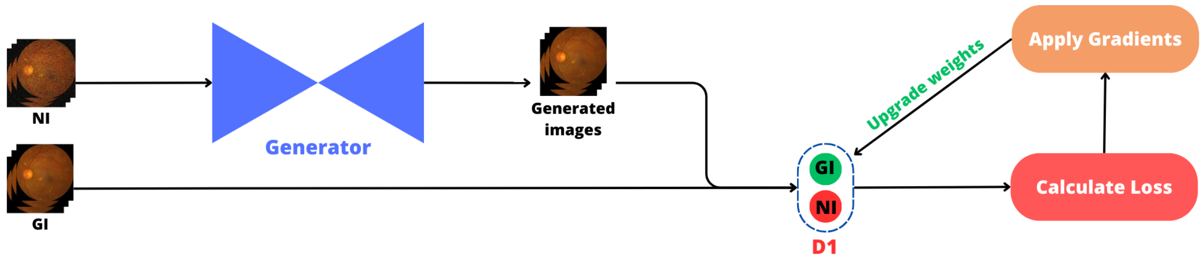
In our methodology, we employed a Generative Adversarial Network (GAN) framework for the denoising of images. The GAN consisted of a generator network, responsible for denoising the images, and two discriminator networks, namely D1 and D2.

Initially, the generator network was trained in collaboration with D1, which was designed to differentiate between a noised image (NI) and a good image (GI) - an image without noise. The generator was fed with pairs of noised and good images, and its objective was to denoise the noised image to make it resemble the good image. D1 provided feedback to the generator, enabling it to learn and improve its denoising capabilities.

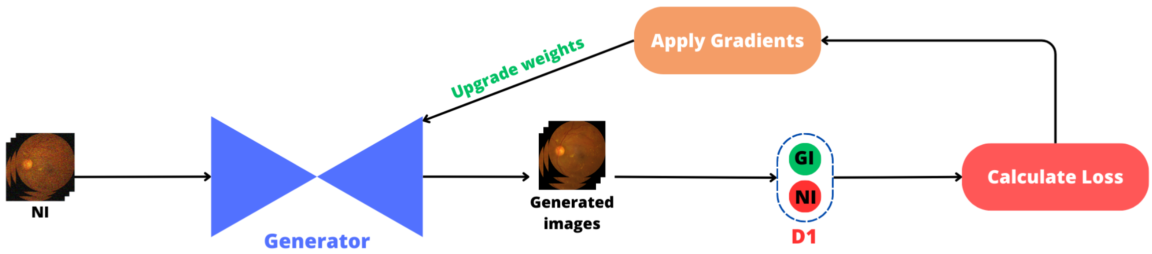
The denoised images classified as good images (GI) by D1 were then passed on to D2, a discriminator trained in the previous section to classify images into five classes of Diabetic Retinopathy (DR). D2 played a crucial role in evaluating the quality of the denoised images and assigning them to the respective DR classes.

On the other hand, if an image was classified as a noised image (NI) by D1, it was returned to the generator for further denoising attempts. This iterative process allowed the generator to refine its denoising abilities until the image was classified as a good image (GI) by D1.

In this section, our focus was on training both the generator network and the D1 discriminator, working collaboratively to enhance the denoising performance of the GAN. Through this iterative training process, we aimed to achieve superior denoising capabilities while ensuring accurate classification of the denoised images into the different DR classes by D2.



(a) Discriminator (D1) training process



(b) Generator training process

Figure 3.29: GAN training phase

II.3.4 Results

In this section, we present the results of our Generative Adversarial Network (GAN) model for denoising images in the context of Diabetic Retinopathy (DR). We discuss the denoising performance of the GAN and its potential for further improvement.

The denoising performance of the GAN was evaluated using Structural Similarity Index (SSIM) metric. However, it is important to note that the initial results obtained from the GAN did not yield satisfactory denoising quality compared to existing methods. The achieved SSIM scores fell below our expectations look at figure 3.30, indicating room for improvement.

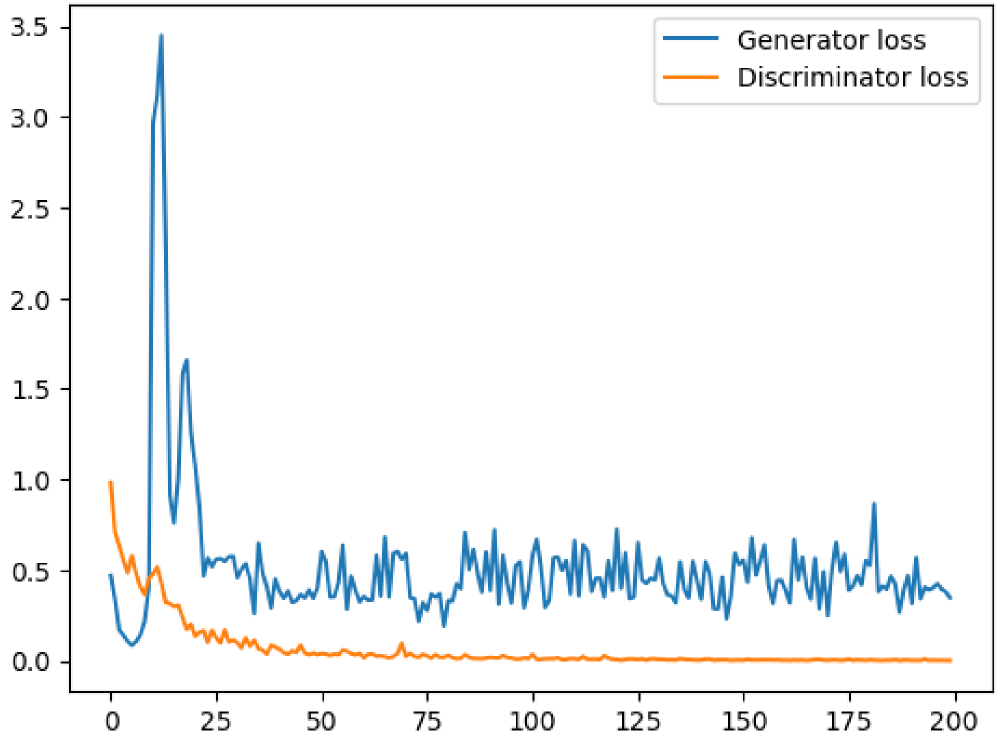


Figure 3.30: GAN loss during training process

In order to evaluate the performance of the generator in denoising images, we conducted extensive testing on a separate test dataset. Figure 3.31 presents a visual comparison between the generated denoised images and their corresponding original images. By closely examining the images, it is evident that the generator has made notable efforts in reducing noise artifacts and enhancing image quality.

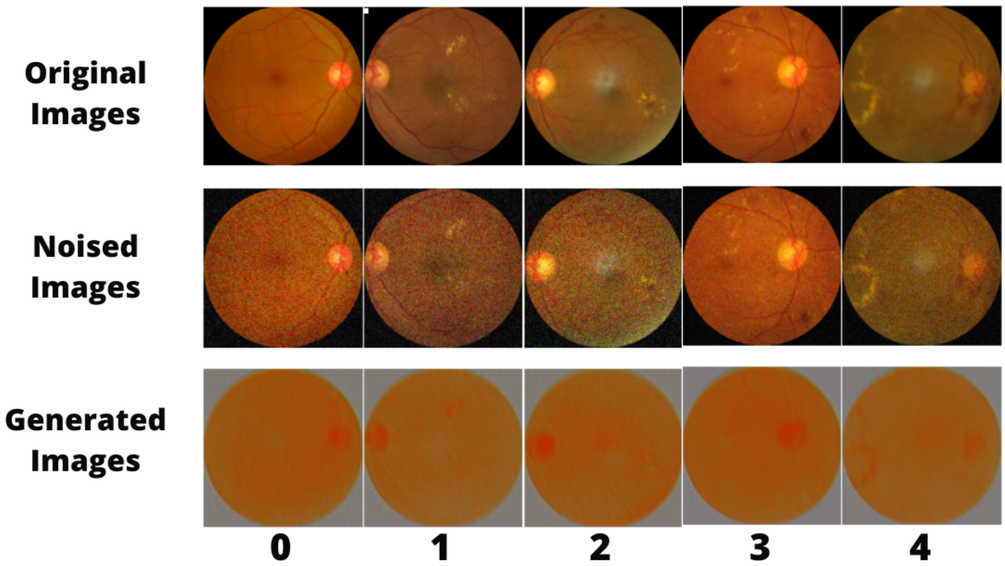


Figure 3.31: visual comparison between the generated denoised images & the original images

II.3.5 Conclusion

While the initial denoising performance of the GAN fell short of our expectations, it presents valuable insights for future improvement. The results suggest that additional training iterations or fine-tuning techniques, such as adjusting the learning rate or exploring different loss functions, may be necessary to enhance the denoising capabilities of the GAN. Furthermore, focusing on improving the D1 discriminator's ability to accurately classify noised and denoised images is crucial to achieving better denoising results.

It is important to acknowledge the limitations of the current implementation and the potential for future advancements. By refining the training process and optimizing the architecture, we believe the GAN has the potential to yield more favorable denoising outcomes, leading to improved image quality and enhanced diagnostic accuracy in the context of Diabetic Retinopathy.

Conclusion

Diabetic retinopathy is the leading cause of blindness among the mid-aged population. While it can be treated successfully in the early stages, it can cause permanent vision problems or even blindness if left undiagnosed and untreated. However, the current diagnostic methods for diabetic retinopathy face several challenges, including subjective interpretation and limited accessibility to specialized eye care services.

To address these challenges, this study focuses on the development and evaluation of an innovative GAN (Generative Adversarial Network) architecture comprising a U-Net generator and a multi-output convolutional neural network discriminator. This architecture facilitates the classification of generated images based on different degrees of the disease. Our primary goal is to improve the accuracy, efficiency, and accessibility of diabetic retinopathy screening and detection. And to achieve this objective, we developed Transfer learning based models to:

- Detect the pathology from eye fundus images: this part of our work achieved remarkable results and was presented at an international conference « Colloque sur les Objets et systèmes Connectés-COC 2023 ».
- Classify the patients eye fundus images into 3 classes and 5 classes according to the severity of the pathology: the results are promising especially for early detection which was our main objective.

these classifiers were used as a discriminator for the GAN architecture. For the generator part we proposed a multi-layer CNN model to denoise and preprocess the new images before feeding them to the discriminator.

We Achieved our defined objectives:

Perspectives

To further enhance the training of the GAN, we can leverage intensive computation resources. This could involve utilizing powerful GPUs or distributed computing frameworks to speed up the training process and optimize the model's performance.

In order to improve the robustness and generalization of the trained model, it would be beneficial to vary the training data.

It is crucial to submit the results to a specialist in order to validate the effectiveness of the generated images in facilitating the recognition of RD. The specialist can provide valuable insights and feedback on the quality and relevance of the generated images, helping to refine and optimize the model further.

It is recommended to develop a mobile application for the prediction and classification of RD. This would allow for wider accessibility and convenience, enabling healthcare professionals or even patients to use the application to detect and classify RD directly from their mobile devices. The application can leverage the trained GAN model to provide accurate predictions and assist in early detection and monitoring of the disease.

BIBLIOGRAPHIE

[1] Mustapha AATILA et al. "Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks". In: *International Journal of Computer Engineering and Data Science (IJCEDS)* 1.1 (2021), pp. 1–7.

[2] Tinuk Agustin, Ema Utami, and Hanif Al Fatta. "Implementation of data augmentation to improve performance CNN method for detecting diabetic retinopathy". In: *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*. IEEE. 2020, pp. 83–88.

[3] Wejdan L Alyoubi, Maysoon F Abulkhair, and Wafaa M Shalash. "Diabetic retinopathy fundus image classification and lesions localization system using deep learning". In: *Sensors* 21.11 (2021), p. 3704.

[4] Farrikh Alzami, Rama Arya Megantara, Ahmad Zainul Fanani, et al. "Diabetic retinopathy grade classification based on fractal analysis and random forest". In: *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*. IEEE. 2019, pp. 272–276.

[5] *An End-to-End Introduction to Generative Adversarial Networks(GANs)*. https://www.analyticsvidhya.com/blog/2021/10/an-end-to-end-introduction-to-generative-adversarial-networksgans/. Accessed: May 17, 2023.

[6] Monika Andonová et al. "Diabetic retinopathy screening based on CNN". In: *2017 International Symposium ELMAR*. IEEE. 2017, pp. 51–54.

[7] David A Antonetti, Paolo S Silva, and Alan W Stitt. "Current understanding of the molecular and cellular pathology of diabetic retinopathy". In: *Nature Reviews Endocrinology* 17.4 (2021), pp. 195–206.

[8] Martin Arjovsky and Léon Bottou. "Towards principled methods for training generative adversarial networks". In: *arXiv preprint arXiv:1701.04862* (2017).

[9] American Optometric Association. *Diabetic Retinopathy*. Accessed on juin 10, 2023. 2013. URL: https://www.aoa.org/patients-and-public/eye-andvision-problems/glossary-of-eye-and-vision-conditions/diabeticretinopathy.

[10] TR Athira and Jyothisha J Nair. "Diabetic Retinopathy Grading From Color Fundus Images: An Autotuned Deep Learning Approach". In: *Procedia Computer Science* 218 (2023), pp. 1055–1066.

[11] Michael W Berry, Azlinah Mohamed, and Bee Wah Yap. *Supervised and unsupervised learning for data science*. Springer, 2019.

[12] Gérard Biau and Erwan Scornet. "A random forest guided tour". In: *Test* 25 (2016), pp. 197–227.

[13] Ramon Casanova et al. "Application of random forests methods to diabetic retinopathy classification analyses". In: *PLOS one* 9.6 (2014), e98587.

[14] Mayank Arya Chandra and SS Bedi. "Survey on SVM and their application in image classification". In: *International Journal of Information Technology* 13 (2021), pp. 1–11.

[15] *CHASE DB1 Dataset*. https://blogs.kingston.ac.uk/retinal/chasedb1/. Last accessed: May 17, 2023.

[16] Xue-Wen Chen and Xiaotong Lin. "Big data deep learning: challenges and perspectives". In: *IEEE access* 2 (2014), pp. 514–525.

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

[18] Antonia Creswell et al. "Generative adversarial networks: An overview". In: *IEEE signal processing magazine* 35.1 (2018), pp. 53–65.

[19] Jorge Cuadros et al. *EyePACS Dataset*. http://www.kaggle.com/c/diabetic-retinopathy-detection/data. Last accessed: May 17, 2023. 2009.

[20] Etienne Decencière et al. *Messidor2 Dataset*. http://www.adcis.net/en/Download-Third-Party/Messidor.html. Last accessed: May 17, 2023. 2014.

[21] Guang Deng and LW Cahill. "An adaptive Gaussian filter for noise reduction and edge detection". In: *1993 IEEE conference record nuclear science symposium and medical imaging conference*. IEEE. 1993, pp. 1615–1619.

[22] Adrien Depeursinge, Julien Fageot, and Omar S Al-Kadi. "Fundamentals of texture processing for biomedical image analysis: a general definition and problem formulation". In: *Biomedical Texture Analysis*. Elsevier, 2017, pp. 1–27.

[23] Terrance DeVries and Graham W Taylor. "Dataset augmentation in feature space". In: *arXiv preprint arXiv:1702.05538* (2017).

[24] *DiaRetDB1 Dataset*. https://www.kaggle.com/datasets/nguyenhung1903/diaretdb1-v21. Last accessed: May 17, 2023.

[25] Muhammed Furkan Ercisli et al. "Diabetic retinopathy: mechanisms, and upcoming biomarkers". In: *Centr. Asi. J. Med. Pharm. Sci. Innov* 1.3 (2021), pp. 131–142.

[26] Mehdi Torabian Esfahani, Mahsa Ghaderi, and Raheleh Kafiyeh. "Classification of diabetic and normal fundus images using new deep learning method". In: *Leonardo Electron. J. Pract. Technol* 17.32 (2018), pp. 233–248.

[27] International Diabetes Federation. *IDF diabetes atlas*. Accessed on juin 10, 2023. 2021. URL: https://diabetesatlas.org/.

[28] Ratul Ghosh, Kuntal Ghosh, and Sanjit Maitra. "Automatic detection and classification of diabetic retinopathy stages using CNN". In: *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE. 2017, pp. 550–554.

[29] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[30] Jiuxiang Gu et al. "Recent advances in convolutional neural networks". In: *Pattern recognition* 77 (2018), pp. 354–377.

[31] Ahnul Ha et al. "Deep-learning-based enhanced optic-disc photography". In: *Plos one* 15.10 (2020), e0239913.

[32] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[33] Kaiming He et al. "Identity mappings in deep residual networks". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer. 2016, pp. 630–645.

[34] National Eye Institute. U.S. Department of Health and Human Services. *Diabetic Retinopathy*. Accessed on juin 10, 2023. 2022. URL: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-anddiseases/diabetic-retinopathy.

[35] Andrew G Howard et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv preprint arXiv:1704.04861* (2017).

[36] Garvin MK Hu Q Abràmoff MD. *rite dataset*. https://medicine.uiowa.edu/eye/rite-dataset. Last accessed: May 17, 2023.

[37] Gao Huang et al. "Densely connected convolutional networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708.

[38] Yanbo Huang. "Advances in artificial neural networks–methodological development and application". In: *Algorithms* 2.3 (2009), pp. 973–1007.

[39] Md Robiul Islam, Md Al Mehedi Hasan, and Abu Sayeed. "Transfer learning based diabetic retinopathy detection with a novel preprocessed layer". In: *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE. 2020, pp. 888–891.

[40] Md Robiul Islam, Md Al Mehedi Hasan, and Abu Sayeed. "Transfer learning based diabetic retinopathy detection with a novel preprocessed layer". In: *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE. 2020, pp. 888–891.

[41] Md Robiul Islam et al. "Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images". In: *Computers in Biology and Medicine* 146 (2022), p. 105602.

[42] Justin M Johnson and Taghi M Khoshgoftaar. "Survey on deep learning with class imbalance". In: *Journal of Big Data* 6.1 (2019), pp. 1–54.

[43] J Pradeep Kandhasamy et al. "Diagnosis of diabetic retinopathy using multi level set segmentation algorithm with feature extraction using SVM with selective features". In: *Multimedia Tools and Applications* 79 (2020), pp. 10581–10596.

[44] Sagar Suresh Karki and Pradnya Kulkarni. "Diabetic retinopathy classification using a combination of efficientnets". In: *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE. 2021, pp. 68–72.

[45] Sohier Dane Karthik Maggie. *APTOS 2019 Blindness Detection*. Last accessed on juin 6, 2023. 2019. URL: https://kaggle.com/competitions/aptos2019-blindness-detection.

[46] Sara Hosseinzadeh Kassani et al. "Diabetic retinopathy classification using a modified xception architecture". In: *2019 IEEE international symposium on signal processing and information technology (ISSPIT)*. IEEE. 2019, pp. 1–6.

[47] Harshit Kaushik et al. "Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models". In: *IEEE Access* 9 (2021), pp. 108276–108292.

[48] Salome Kazeminia et al. "GANs for medical image analysis". In: *Artificial Intelligence in Medicine* 109 (2020), p. 101938.

[49] Hager Khalil et al. "Classification of diabetic retinopathy types based on convolution neural network (CNN)". In: *Menoufia Journal of Electronic Engineering Research* 28.ICEEM2019-Special Issue (2019), pp. 126–153.

[50] Sotiris B Kotsiantis. "Decision trees: a recent overview". In: *Artificial Intelligence Review* 39 (2013), pp. 261–283.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[52] Jaykumar Lachure et al. "Diabetic retinopathy using morphological operations and machine learning". In: *2015 IEEE international advance computing conference (IACC)*. IEEE. 2015, pp. 617–622.

[53] Erik G Learned-Miller. "Introduction to supervised learning". In: *I: Department of Computer Science, University of Massachusetts* (2014), p. 3.

[54] Chongxuan Li et al. "Triple generative adversarial nets". In: *Advances in neural information processing systems* 30 (2017).

[55] Xiaogang Li et al. "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification". In: *2017 10th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*. IEEE. 2017, pp. 1–11.

[56] Yu Li et al. "Deep learning in bioinformatics: Introduction, application, and perspective in the big data era". In: *Methods* 166 (2019), pp. 4–21.

[57] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical image analysis* 42 (2017), pp. 60–88.

[58] Sharmin Majumder and Nasser Kehtarnavaz. "Multitasking deep learning model for detection of five stages of diabetic retinopathy". In: *IEEE Access* 9 (2021), pp. 123220–123230.

[59] Mehran Mehralian and Babak Karasfi. "Rdcgan: Unsupervised representation learning with regularized deep convolutional generative adversarial networks". In: *2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium*. IEEE. 2018, pp. 31–38.

[60] Mehdi Mirza and Simon Osindero. "Conditional generative adversarial nets". In: *arXiv preprint arXiv:1411.1784* (2014).

[61] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol. "A scoping review of transfer learning research on medical image analysis using ImageNet". In: *Computers in biology and medicine* 128 (2021), p. 104115.

[62] Saeed Nosratabadi et al. "Data science in economics: comprehensive review of advanced machine learning and deep learning methods". In: *Mathematics* 8.10 (2020), p. 1799.

[63] Mohammed Oulhadj et al. "Diabetic retinopathy prediction based on deep learning and deformable registration". In: *Multimedia Tools and Applications* 81.20 (2022), pp. 28709–28727.

[64] Kamal Porwal et al. *IDRiD (Indian Diabetic Retinopathy Image Dataset)*. https://idrid.grand-challenge.org/. Last accessed: May 17, 2023. 2018.

[65] Dinial Utami Nurul Qomariah, Handayani Tjandrasa, and Chastine Fatichah. “Classification of diabetic retinopathy and normal retinal images using CNN and SVM”. In: *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. IEEE. 2019, pp. 152–157.

[66] Mohaimenul Azam Khan Raiaan et al. “A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images”. In: *IEEE Access* (2023).

[67] Mihir Rao, Michelle Zhu, and Tianyang Wang. “Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy”. In: *arXiv preprint arXiv:2010.11692* (2020).

[68] Ali M Reza. “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement”. In: *Journal of VLSI signal processing systems for signal, image and video technology* 38 (2004), pp. 35–44.

[69] Sohini Roychowdhury, Dara D Koozekanani, and Keshab K Parhi. “DREAM: diabetic retinopathy analysis using machine learning”. In: *IEEE journal of biomedical and health informatics* 18.5 (2013), pp. 1717–1728.

[70] Olga Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

[71] Olga Russakovsky et al. “Imagenet large scale visual recognition challenge”. In: *International journal of computer vision* 115 (2015), pp. 211–252.

[72] Hojjat Salehinejad et al. “Recent advances in recurrent neural networks”. In: *arXiv preprint arXiv:1801.01078* (2017).

[73] Mohamed Shaban et al. “A convolutional neural network for the screening and staging of diabetic retinopathy”. In: *Plos one* 15.6 (2020), e0233514.

[74] Fatma Shaheen, Brijesh Verma, and Md Asafuddoula. “Impact of automatic feature extraction in deep learning architecture”. In: *2016 International conference on digital image computing: techniques and applications (DICTA)*. IEEE. 2016, pp. 1–8.

[75] T Shanthi and RS Sabeenian. “Modified Alexnet architecture for classification of diabetic retinopathy images”. In: *Computers & Electrical Engineering* 76 (2019), pp. 56–64.

[76] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

[77] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[78] Srikanth Tammina. “Transfer learning using vgg-16 with deep convolutional neural network for classifying images”. In: *International Journal of Scientific and Research Publications (IJSRP)* 9.10 (2019), pp. 143–150.

[79] Mingxing Tan and Quoc Le. “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.

[80] Handayani Tjandrasa et al. "Classification of non-proliferative diabetic retinopathy based on hard exudates using soft margin SVM". In: *2013 IEEE International Conference on Control System, Computing and Engineering*. IEEE. 2013, pp. 376–380.

[81] Kritika Verma and Pradeep Kumar Singh. "An insight to soft computing based defect prediction techniques in software". In: *International Journal of Modern Education and Computer Science* 7.9 (2015), p. 52.

[82] Fayziyev Voxid et al. "MACHINE LEARNING AND ITS PROSPECTS". In: *IJTIMOIY FANLARDA INNOVASIYA ONLAYN ILMIY JURNALI* 3.2 (2023), pp. 52–54.

[83] Stela Vujosevic et al. "Screening for diabetic retinopathy: new perspectives and challenges". In: *The Lancet Diabetes & Endocrinology* 8.4 (2020), pp. 337–347.

[84] Kunfeng Wang et al. "Generative adversarial networks: introduction and outlook". In: *IEEE/CAA Journal of Automatica Sinica* 4.4 (2017), pp. 588–598.

[85] M Arif Wani et al. *Advances in deep learning.* Springer, 2020.

[86] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. "A survey of transfer learning". In: *Journal of Big data* 3.1 (2016), pp. 1–40.

[87] Kele Xu, Dawei Feng, and Haibo Mi. "Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image". In: *Molecules* 22.12 (2017), p. 2054.

[88] Lin Yin et al. "Weighted median filters: a tutorial". In: *IEEE Transactions on circuits and systems II: analog and digital signal processing* 43.3 (1996), pp. 157–192.

[89] Tae Keun Yoo, Joon Yul Choi, and Hong Kyu Kim. "CycleGAN-based deep learning technique for artifact reduction in fundus photography". In: *Graefe's Archive for Clinical and Experimental Ophthalmology* 258 (2020), pp. 1631–1637.

[90] Tae Keun Yoo et al. "Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images". In: *Computer Methods and Programs in Biomedicine* 205 (2021), p. 106086.

[91] Tae Keun Yoo et al. "Deep learning can generate traditional retinal fundus photographs using ultra-widefield images via generative adversarial networks". In: *Computer Methods and Programs in Biomedicine* 197 (2020), p. 105761.

[92] Jason Yosinski et al. "How transferable are features in deep neural networks?" In: *Advances in neural information processing systems* 27 (2014).

[93] Chenyu You et al. "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)". In: *IEEE transactions on medical imaging* 39.1 (2019), pp. 188–203.

[94] He Zhang and Vishal M Patel. "Densely connected pyramid dehazing network". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2018, pp. 3194–3203.

[95] Xiaohui Zhang and Opas Chutatape. "A SVM approach for detection of hemorrhages in background diabetic retinopathy". In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* Vol. 4. IEEE. 2005, pp. 2435–2440.

[96] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. "Sim-to-real transfer in deep reinforcement learning for robotics: a survey". In: *2020 IEEE symposium series on computational intelligence (SSCI).* IEEE. 2020, pp. 737–744.

[97] Tao Zhou et al. "GAN review: Models and medical image fusion applications". In: *Information Fusion* 91 (2023), pp. 134–148.

[98] Zhi-Hua Zhou. *Machine learning*. Springer Nature, 2021.

[99] Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

ABSTRACT

Despite advancements in medical imaging technology, the interpretation of medical imagery still necessitates the expertise of specialists, which can present challenges in terms of practicality and accessibility. However, emerging technologies such as deep learning offer a promising solution to address these challenges. By leveraging deep learning algorithms, the accuracy and efficiency of diagnosis can be significantly improved, enabling faster and easier identification of various medical conditions. Among these conditions, diabetic retinopathy stands out as one that critically necessitates the advancements offered by deep learning. In this work, we exploit GANs, CNNs and Transfer learning to diagnose Diabetic Retinopathy (DR), by proposing an architecture that also allows to augment the data from real images. The experimental results obtained are very promising and a part of this work has been presented in an international conference « Colloque sur les Objets et systèmes Connectés-COC 2023 ».

RÉSUMÉ

Malgré les avancées dans la technologie d'imagerie médicale, l'interprétation des images médicales nécessite toujours l'expertise des spécialistes, ce qui peut poser des défis en termes de commodité et d'accessibilité. Cependant, les technologies émergentes telles que l'apprentissage profond offrent une solution prometteuse pour relever ces défis. En exploitant les algorithmes d'apprentissage profond, il est possible d'améliorer considérablement la précision et l'efficacité du diagnostic, permettant ainsi une identification plus rapide et plus facile de diverses pathologies médicales. Parmi ces pathologies, la rétinopathie diabétique se distingue comme étant celle qui nécessite de manière critique les avancées offertes par l'apprentissage profond. Dans ce travail, nous exploitons les réseaux génératifs antagonistes (GANs), les réseaux de neurones convolutionnels (CNNs) et l'apprentissage par transfer pour diagnostiquer la rétinopathie diabétique en proposant une architecture qui permet également d'augmenter les données à partir d'images réelles. Les résultats expérimentaux obtenus [sont très prometteurs et une partie du travail a fait objet d'une conférénce international « Colloque sur les Objets et systèmes Connectés-COC 2023 ».