FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

XAIPrivacy – XAI with differential privacy

Fábio Manuel Neves de Araújo

DISSERTATION

Mestrado em Engenharia Informática e Computação

Supervisor: João Manuel Patrício Pedrosa

Second Supervisor: João Gonçalves

January 27, 2023

XAIPrivacy – XAI with differential privacy

Fábio Manuel Neves de Araújo

Mestrado em Engenharia Informática e Computação

January 27, 2023

Abstract

Glaucoma has been a leading and rapidly growing disease in recent years. This disease affects the eye's optic nerve, causing total loss of vision. It is one of the leading causes of blindness worldwide and is characterized by degeneration associated with the death of Retinal Ganglion Cells (RGCs). Detecting it at a very early stage is essential for effective prevention. Since the manual process of making this detection has to be done for each image, it can be time-consuming. To improve Glaucoma detection, automated methods have been developed to make this detection so that detection can be made by these models and only confirmed by medical specialists. One of the big problems with these automatic systems is the lack of explainability in the decisions that have been made. Moreover, this ends up holding people back in their acceptance. A possible solution to bring us this explainability is using models based on examples. However, this process brings another problem: privacy, which is present in the biometric data of retinal images. This dissertation then intends to use GANs to make an example generation that serves as an explanation, but at the same time, that does not imply the privacy of the images. We will use two datasets, one with complete retinal images and the other with images of the optic disc area. The images of the optic disc were cut according to the masks provided in each dataset, and then the two datasets will be prepared with a resize to 256x256. During the DCGAN training, two optimizers will be used, one without differential privacy and one with differential privacy. Next, two algorithms will be used to evaluate the similarity between the original images and the images generated for both datasets, the alghoritms are CosineSimilarity and Pairwise. We will also use two algorithms for a quantitative evaluation: the Inception Score (IS) and the Fréchet Inception Distance (FID). Finally, the RetinaQualEvaluator model will be used to evaluate the quality of the generated images and a Glaucoma Classifier will be used to confirm that the images with more significant similarity have the same quality and the same Glaucoma evaluation. After training and image generation, the results obtained, have very balanced values, and the generated images are very similar to the original images. There is a small decrease when referring to the complete retinal images, it may be because they have a lot of detail and also have enough black area in the images, interfering negatively in the learning of the algorithms. In relation to the images from the optic disc area, this does not happen anymore, thus obtaining better results.

Keywords: retinography, biometric data, obfuscation, privacy preserving, image generation, XAI, GAN, DCGAN

Resumo

O glaucoma tem sido uma doença líder e em rápido crescimento nos últimos anos. Esta doença afecta o nervo óptico do olho, causando a perda total da visão. É uma das principais causas da cegueira a nível mundial e é caracterizada pela degeneração associada à morte das células Ganglionares da Retina (RGCs). A sua deteção num estado muito precoce é essencial para conseguir ser feita uma prevenção eficaz. Uma vez que o processo manual para fazer esta deteção tem de ser feita a cada imagem de forma individual e pode ser algo demorado. Para melhorar este processo, foram desenvolvidos métodos automáticos para fazer esta deteção, deste modo a deteção pode ser feita por estes modelos e apenas confirmada por médicos especialistas. Um dos grandes problemas destes sistemas automáticos é a falta de explicabilidade nas decisões que foram tomadas. E isto acaba por retrair as pessoas na sua aceitação. Uma possível solução para nos trazer esta explicabilidade, é usarmos modelos que sejam baseados em exemplos, mas este processo traz outro problema que é o da privacidade, que está presente nos dados biométricos das imagens da retina. Esta dissertação pretende então fazer uso de GANs para fazer uma geração de exemplos que sirvam de explicação, mas ao mesmo tempo que não impliquem a privacidade das imagens. Vão ser usados dois datasets, um com imagens completas da retina e o outro com imagens da área do disco óptico. As imagens do disco óptico foram recortadas de acordo com as máscaras fornecidas em cada dataset, e depois os dois conjuntos de dados foram preparados com um redimensionamento para 256x256. Durante o treino da DCGAN, serão utilizados dois optimizadores, um sem privacidade diferencial e outro com privacidade diferencial. Em seguida, serão utilizados dois algoritmos para avaliar a semelhança entre as imagens originais e as imagens geradas para ambos os conjuntos de dados, os algoritmos são o CosineSimilarity e o Pairwise. Utilizaremos também dois algoritmos para uma avaliação quantitativa: o Inception Score (IS) e o Fréchet Inception Distance (FID). Finalmente, o modelo RetinaQualEvaluator será utilizado para avaliar a qualidade das imagens geradas e um Classificador de Glaucoma será utilizado para validar que as imagens com maior similiaridade têm a mesma qualidade e a mesma avaliação de Glaucoma. Após o treino e geração de imagens, os resultados obtidos, tem valores bastante equilibrados, e as imagens geradas são muito parecidas com as imagens originais. Havendo uma pequena descida quando nos referimos as imagens completas da retina, poderá ser por terem muito detalhe e também ter bastante área preta nas imagens, interferindo negativamente na aprendizagem dos algoritmos. Em relação as imagens da zona do disco ótico, isso já não acontece obtendo assim melhores resultados.

Keywords: retinography, biometric data, obfuscation, privacy preserving, image generation, XAI, GAN, DCGAN

Acknowledgments

I want to begin by thanking those who have guided me through this journey, providing oversight and inspiration: Professor João Pedrosa and, on behalf of Fraunhofer, João Gonçalves. Who, despite my decision to postpone the dissertation one semester, never gave up and continued with me to finalize this journey of my life. I want to thank Fraunhofer for making this postponement possible, and without this opportunity, I probably would not have finished the dissertation. However, many more people contributed indirectly to this work, providing emotional support and motivation during this period in which I always thought I would not be able to conciliate everything since I had to work and also due to the fact that I was a father. Finally, I want to thank my family for their endless support, and I would like to dedicate this work to them, especially to my wife and daughter! Without you, none of this would be possible.

This work was done under the scope of the project Transparent Artificial Medical Intelligence (TAMI), co-funded by Portugal 2020 framed under the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), Fundação para a Ciência and Technology (FCT), Carnegie Mellon University, and European Regional Development Fund under Grant 45905. The authors would like to thank the Anatomical Pathology Service of the Portuguese Oncology Institute – Porto (IPO-Porto) [1]

[1] https://www.aicos.fraunhofer.pt/en/our$_{w}ork/projects/tami.html$

Contents

List of Figures

List of Tables

Abbreviations

AC-G	Angle-Closure Glaucoma
AI	Artificial Intelligence
AION	Anterior Ischemic Optic Neuropathy
AMD	Age-related Macular Degeneration
BCE	Binary Cross Entropy
CAD	Computer-Aided Diagnosis
DCGAN	Deep Convolutional Generative Adversarial Network
DL	Deep Learning
DP	Differential Privacy
DP-CGAN	Differential Privacy Conditional Generative Adversarial Network
DR	Diabetic Retinopathy
FID	Fréchet Inception Distance
GAM	Generalized Additive model
GAN	Generative Adversarial Networks
ICGA	IndoCyanine Green Angiography
IOP	IntraOcular Pressure
IS	Inception Score
LIME	Local interpretable model-agnostic explanations
LF	LipoFuscin
LRP	Layer-wise Relevance Propagation
ML	Machine Learning
NTG	Normal-Tension Glaucoma
OCT	Optical Coherence Tomography
OCTA	Optical Coherence Tomography Angiography
ONH	Optic Nerve Hypoplasia
ONH*	Optic Nerve Head
PACS	Picture Archiving and Communication System
PATE	Private Aggregation of Teacher Ensembles
PATE-GAN	Private Aggregation of Teacher Ensembles Generative Adversarial Networks
POA-G	Primary Open-Angle Glaucoma
RF-GAN	Retinal Fundus images Generative Adversarial Networks
RIGA	Retinal fundus Images for Glaucoma Analysis
RIM-ONE	Retinal IMage database for Optic Nerve Evaluation
RGC	Retinal Ganglion Cells
ROP	RetinOpathy Prematurity
RPE	Retinal Pigment Epithelium
SHAP	Shapley Additive Explanations
TAMI	Transparent Artificial Medical Intelligence
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

1.1 Context

Glaucoma has been a leading and rapidly growing disease in recent years. This disease affects the eye's optic nerve, causing complete loss of vision. It is one of the leading causes of blindness worldwide [15] and is characterized by degeneration associated with the death of Retinal Ganglion Cells (RGCs). Being able to detect this disease at an early stage is essential to get timely treatment, avoiding irreversible damage, such as vision loss. Typically the analysis of the images is done one by one manually by medical specialists. However, the manual screening of glaucoma has several disadvantages. First, the high number of images that need reporting, results in a significant burden for clinicians. Second, the variability between specialists means that the same image analysed by different specialists may have a different result. Finally, the high costs associated to this process as hospitals need to have someone to analyze the images one by one, and the process is time-consuming. Because these problems exist, the development of automatic processes for glaucoma screening, with the purpose of providing a second opinion, has been an ongoing field of research. These processes are not intended to replace doctors because the final opinion is always from the specialists. However, most of the analysis would be done automatically and reducing errors caused by fatigue, improving diagnostic accuracy, reducing inter-reader and intra-reader variability, and, of course, automating the screening process. Currently there are already some commercial solutions such as the Yanbao application and RetinaLyze.

1.2 Motivation

Despite the promising results with deep learning techniques, one of the main problems is the lack of explainability of these methods, which leads to not all people accepting them readily. There are several XAI techniques (Local interpretable model-agnostic explanations (LIME), Layer-wise relevance propagation (LRP), Grad-CAM, ExMatchina, etc.). However, one crucial technique is to show examples of similar patients to justify the decision, and we have the example of ExMatchina, which generates explainability based on examples. However, since the retina has unique biometric data, this brings a privacy problem since other patients' data will be shown which may allow for their identification. Thus it is crucial to develop methods that can generate examples that can justify a decision without allowing the identification of the patient while maintaining all other characteristics of the image.

1.3 Goals

The goals of this project are:

- The development of a Deep Convolutional Generative Adversarial Network (DCGAN) for high-quality retinal fundus imaging that does not allow the identification of the patient while maintaining the properties that lead to disease detection.
- Validation of the generated images through quantitative metrics.
- Development and validation of a framework for the retrieval of anonymized artificial examples for the explanation of an automatic glaucoma screening decision support system.

1.4 Document structure

In addition to this chapter, the preparation of the dissertation consists of five more chapters. In Chapter 2, an introduction is made about the eye and its anatomy, talking about the various parts that constitute it and its functioning, and the ocular pathologies, with a particular focus on Glaucoma. Also, this chapter introduces a section on retinal imaging and in which datasets we can obtain images, and at the end of the chapter is discussed some XAI methodes. In Chapter 3, the state of the art on artificial image generation and GANs is discussed, with a focus on artificial image generation solutions in fundus imaging. Chapter 4, contains the data preparation done on the images, the settings used to train the DCGAN, loss functions, and optimizers. It also explained how features were extracted from the pictures to measure similarity. Chapter 5, contains the metrics used to compare the training of two DCGANs, one without differential privacy, and one with. It also includes the results of the image quality and glaucoma classifier algorithms. The image quality algorithm was only used on the complete retinal images, while the glaucoma classifier was used on the pictures of the optic disc area. Finally, in Chapter 6 the conclusions of this MSc thesis are described.

Chapter 2

Glaucoma and Retinal Imaging

In this chapter, a general review of the eye is described. This review includes a brief description of the eye and its anatomy, and some existing eye diseases are presented, focusing on glaucoma. Then we have a section dedicated to imaging modalities and a summary of automatic image analysis to detect diseases.

2.1 The Eye

2.1.1 Anatomy

The human eye is a sensory organ that reacts to visible light and is the organ that allows us to see. It is part of the sensory nervous system. The human body includes five senses: sight, touch, smell, taste, and hearing. All these senses are directly related to other parts of the human body. The eye is connected to the brain and depends on it for their interpretation of what we see, in figure 2.1 as an example of the anatomy of the eye.

The eye is an extremely complex organ with multiple components that serve specific functions:

- **Choroid** This layer contains blood vessels that line the back of the eye and is situated between the retina (the light-sensitive inner layer) and the sclera (the white outer wall of the eye).

- **Ciliary Body** This structure contains muscle, and its location is behind the iris, which focuses on the lens.

- **Cornea** Is the clear front part of the eye allows light to focus and transmit (i.e., clarity and sharpness) into the eye. Corrective laser surgery alters the cornea by changing its focus.

- **Macula** Is part of the retina and contains special light-sensitive cells. These cells allow us to see fine details clearly in the center of our visual field. Their deterioration usually increases as we get older (age-related macular degeneration, or ARMD).

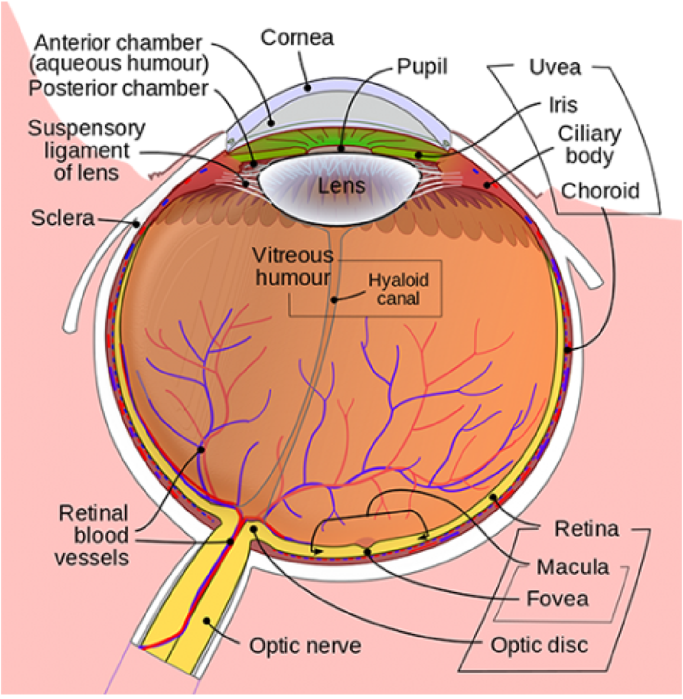


Figure 2.1: Eye anatomy. Image from [28]

- **Fovea** Is the center of the macula that allows us to have clear vision.

- **Iris** It can be identified by being the colored part of the eye. Its function is to help regulate the amount of light entering the eye. The iris closes and opens the pupil to let in less or more light, respectively. When exposed to very bright lights, it lets in less light, and when it is exposed to dim lights, it lets in more light.

- **Lens** The light rays hitting the retina are controlled by the lens. It is transparent and can be replaced if necessary. As we get older, the lens becomes more and more degraded, which leads to the need to wear reading glasses. There are intraocular lenses that are used to replace lenses clouded by cataracts.

- **Optic Nerve** Is made up of more than a million nerve fibers that transport visual messages from the retina to the brain. (In order for us to see, the eyes have to be connected to the brain, and there has to be light.) What we see is controlled by the brain, which interprets the images. The retina sees images upside down, but the brain flips the images right side up. This inversion of the images we see is similar to a camera's mirror. One of the eye conditions related to the optic nerve is Glaucoma.

- **Pupil** Is the dark part that lies in the middle of the iris. The pupil adjusts to the amount of light by changing its size (shrinks in bright light and increases in dim light). This opening and closing of light in the eye resemble the opening in most 35 mm cameras, allowing more or less light to enter.

- **Retina** is the nerve layer that lines the back of the eye. The retina senses light and creates electrical impulses that are sent through the optic nerve to the brain.
- **Sclera** Is the white outer layer of the eye, which surrounds the iris.
- **Choroid** Is the transparent and gelatinous substance that fills the central cavity of the eye.

2.1.1.1 Functioning of the eye

The way we see depends on the transfer of light. Light passes through the cornea at the front of the eye to the lens. In conjunction with the cornea, the lens helps focus the light rays on the retina located at the back of the eye. The retina cells absorb and convert the light into electrochemical impulses, which are transferred along the optic nerve reaching the brain. The operation of the eye is very similar to a camera. The shutter of a camera can close or open, depending on the amount of light needed to expose the film on the back of the camera. The eye works in the same way. The iris and pupil control the amount of light to be let into the back of the eye. The pupils grow when it is too dark, letting in more light. The lens of a camera can focus on objects far and near with the help of mirrors and other mechanical devices. The eye contains a retina that contains three cells that convert light energy into electrical energy. The rods respond to low light intensities contributing to the understanding of low resolution black and white images. In contrast, the cones respond to high light intensities contributing to the understanding of high-resolution color images. The photosensitive ganglion cells respond to all light intensities, controlling the amount of light reaching the retina, regulating and suppressing the hormone melatonin, and triggering the circadian rhythm.

2.1.2 Ocular Diseases

Due to the complexity and sensitivity of the ocular anatomy and function, several diseases can afflict the eye, having an impact on vision. These can be divided into refractive diseases and non-refractive.

Some of the most common ocular diseases are:

- **Refractive diseases**: are the most frequent eye problems. This diseases occurs between ages 40–50 years and refer to a set where there is an inadequate focusing of the images on the retina. Typically can be corrected by eyeglasses, contact lenses, or surgery. Some refractive diseases are Myopia, Hyperopia, Astigmatism, Presbyopia (or tired eyes).
- **Age-Related Macular Degeneration (AMD)**: is an eye disorder associated with aging and damages sharp and central vision. Central vision is needed to see objects in everyday tasks such as reading and driving. AMD affects the macula, the retina's central part that allows the eye to see fine details. There are two forms of AMD: wet and dry.
- **Cataract**: remains a leading cause of visual impairment worldwide. Although 90 percent of cases of this disease worldwide are referred to developing countries, its social, physical, and

economic impact remains substantial in the developed world. Cataracts are a more common disease in older people than in younger people. Surgery is often effective in restoring vision. However, it has high costs in Europe and other western countries. [35].

- **Diabetic Retinopathy**: is a major cause of preventable visual impairment and blindness in the European Region. Although most European countries have some prophylactic eye examinations for DR, they are not organized according to the principles of screening in medicine [24]. This developing disease is characterized by damage to the retina's blood vessels and to the light-sensitive tissue at the back of the eye that is necessary for good vision. DR evolves in four stages, mild nonproliferative Retinopathy (microaneurysms), moderate nonproliferative Retinopathy (blockage in some retinal vessels), severe nonproliferative Retinopathy (more vessels are blocked, leading to deprivation of the retina's blood supply leading to the growth of new blood vessels), and proliferative Retinopathy (most advanced stage). Diabetic Retinopathy usually affects both eyes.

- **Glaucoma**: is a disease that can damage the eye's optic nerve, resulting in loss of vision. Glaucoma happens when the fluid pressure inside the eye slowly increases. However, recent results confirm that it can also happen with normal fluid pressure inside the eye.

2.2 Imaging Modalities

Retinal imaging has undergone a revolution over the last 50 years to understand the eye in both health and disease. There have been significant improvements both in hardware and software. We have lasers, optics, and software for image analysis in hardware. Some optical imaging modalities such as Fundus Photography, Molecular Imaging (MI), PhotoAcoustic Microscopy (PAM), Scanning Laser Ophthalmoscopy (SLO), Fundus AutoFluorescence (FAF), Optical Coherence Tomography (OCT), and OCT Angiography (OCTA). These modalities have allowed us to visualize the pathophysiology of the retina better and have also had a major impact on medical research. These improvements in technology have resulted in earlier detection of disease, more accurate diagnosis, and better management of numerous diseases [31].

- **Fundus Photography** consists of photographing the back of an eye. The cameras used for this purpose consist of an intricate microscope connected to a camera with flash. The main structures present in a fundus photograph are the central and peripheral retina, optic disc, and macula. Fundus photography can be performed with colored filters or specialized dyes, including fluorescein and indocyanine green. The technology and models for this type of imaging have evolved quite rapidly over the last century. The equipment used to photograph the retinal fundus is quite sophisticated, and its production is quite complex because it has to meet clinical standards. Only a few manufacturers have that capability.

- **MI** techniques allow visualization of molecular processes and functional changes in living animals and humans before morphological changes occur at cellular and tissue levels. It requires high-resolution imaging, sensitive instrument detection, specific imaging agents, and endogenous molecular probes or exogenous contrast agents that link the imaging signal to a molecular probe or event.

- **PAM** Is obtained through optical excitation and ultrasonic detection. A short pulse laser illuminates and excites a target tissue, thereby inducing ultrasonic pressure waves due to specific optical absorption. The ultrasonic transducer focuses on the tissue surface recording ultrasonic signals, generating an image.

- **SLO** was first described in 1981. It makes use of a monochromatic laser with low power and is a microscopic confocal scanning technique to collect an image of the retina and optic nerve head. These images have a higher contrast when compared to photographs taken with more generic cameras, as they can reduce the effect of light scattering.

- **FAF** imaging is a non-invasive imaging modality. The primary sources are lipofuscin (LF) granules accumulated in the cells of the Retinal Pigment Epithelium (RPE). When accumulated excessively in the RPE cells, Lipofuscin granules are classified as a common pathogenic pathway in numerous retinal diseases. The significance of changes in FAF imaging can be further addressed by assessing the corresponding retinal sensitivity and response to stimuli. Severe damage to the RPE corresponds to areas of diminished autofluorescence.

- **OCT** was introduced to ophthalmology in 1991 and since then has had an excellent uptake in the ophthalmology field. Since the eye is optically accessible to visible and near-infrared light, it allows a relatively good combination of appropriate tissue penetration depth and axial resolution.

- **OCTA** is a new, noninvasive imaging technique based on OCT imaging that allows for visualizing the retinal and choroidal microvasculature without the injection of exogenous dyes. OCTA is a method of visualizing vasculature enhanced from the signal (intensity and/or phase) change caused by erythrocyte movement that arises from multiple B-scans performed at the same position. B-scans are two-dimensional images generated from several one-dimensional images. OCTA images are essentially motion-contrast images. Various algorithms have been developed for OCTA devices.

2.3 Glaucoma

Glaucoma is a major disease that has seen a considerable increase in recent years. Glaucoma affects the optic nerve of the eye. It is a disease that is usually asymptomatic in the early stages and can cause blindness, or severe vision loss, if not diagnosed and treated in a timely and appropriate manner. It is one of the leading causes of blindness worldwide and is characterized by degeneration associated with the death of Retinal Ganglion Cells (RGCs). Detection at an early stage of this disease is essential to decrease the risk of permanent vision loss. It can usually be classified into

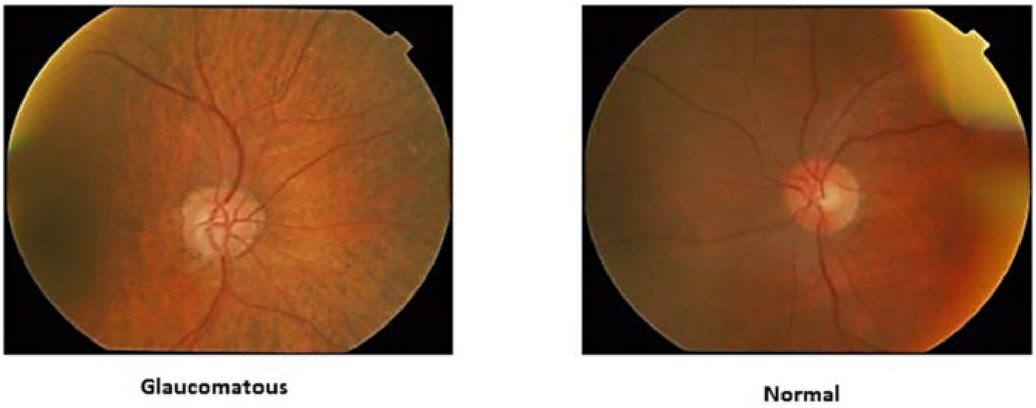


Figure 2.2: There are two retinal images in this picture: the left side has glaucoma, and the right side has no glaucoma. Image from [11]

primary open-angle Glaucoma (POA-G) and closed-angle Glaucoma (AC-G). POA-G accounts for about 90 percent of cases with this disease and is known as the most common type [11].

- **POA-G**: It is caused by the slow obstruction of the channels through which the fluid pressure inside the eye is drained, increasing eye pressure. It has a wide and open angle between the iris and cornea, develops slowly, is a condition that has no cure, and symptoms and damage are hardly detectable.
- **AC-G**: Glaucoma with this form has a problem that affects the eye's drainage angle. This means that the iris is too close to the trabecular meshwork. This disease causes the drainage angle to become blocked, which causes fluid to remain inside the eye.
- **Normal-Tension Glaucoma (NTG)**: NTG is characterized by damage to the optic nerve and vision loss despite intraocular pressure not being elevated above the average level. It is also known to result from poor blood flow to the optic nerve. It is associated with various conditions, including migraines, Raynaud's disease, and sleep apnea.
- **Congenital glaucoma**: It is a form of Glaucoma that occurs in infants and very young children due to abnormal eye drainage angle development. Symptoms of congenital Glaucoma often include light sensitivity, watery eyes, or a tendency to keep the eyes closed. The eyes may also appear larger than usual and have cloudy corneas.
- **Secondary glaucoma**: May be caused by an eye injury, inflammation, certain drugs such as steroids, and advanced cataract cases or diabetes. The type of treatment will depend on the underlying cause but usually includes medication, laser surgery, or conventional surgery.

2.3.1 Detection/Diagnosis of Glaucoma

To diagnose glaucoma, first, the retinal images are captured with the appropriate equipment. Then the images have to be analyzed one by one or several health professionals. Cup to disc ratio from retinal fundus images is an essential procedure for glaucoma detection, this can be seen in figure 2.3. This process can be time-consuming, and errors can occur, such as not detecting where it exists or where it does not exist. Not to take away the merit of the health professionals, this detection process can be aided with automatic procedures. In the end, the final opinion will always be of the responsible professional. The development of automatic processes for glaucoma screening, has been an ongoing field of research, because automatic detection is util to help doctors detect this disease [26].

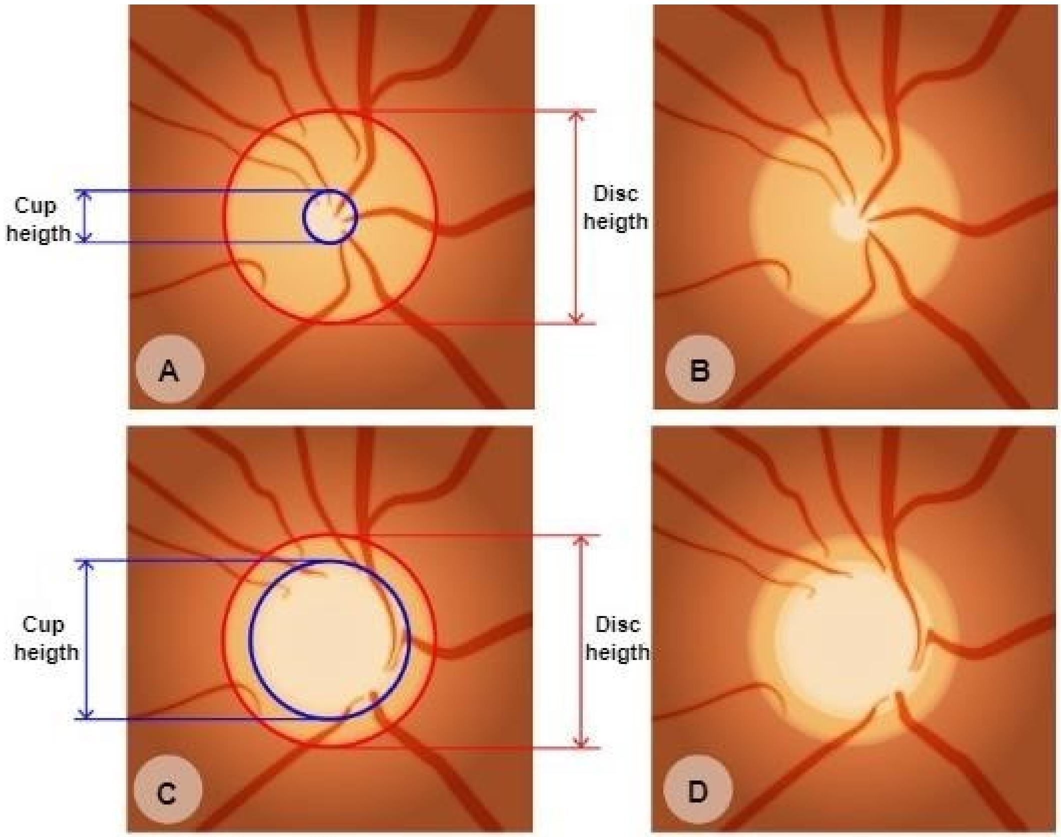


Figure 2.3: Optic disc with normal cup and increased cup caused by glaucoma: (A,B) Non-Glaucoma; (C,D) Glaucoma. Image from [41]

2.4 Automatic Retinal Imaging Analysis

Healthcare professionals usually have to analyze a massive collection of data from different body structures during their work routine, which is a tiring and challenging task with a high degree of associated responsibility. With this in mind, Computer-aided diagnosis (CAD) systems have been widely developed during the last decades to help these professionals. As the name suggests, these systems assist professionals in interpreting medical images and other types of data, providing "second opinions", allowing a faster patient triage, and sometimes acting as end-to-end solutions for an initial diagnosis of several conditions. Therefore, these systems' main goals consist of reducing the burdens caused by the intense workload of a healthcare professional, reducing errors caused by fatigue, improving diagnostic accuracy, and reducing inter-reader and intra-reader variability. Automating the screening process, allowing massive screenings for Glaucoma that were not possible before. The recent growth in the development of these systems can be justified by the successive improvements in machine learning/deep learning and computer vision fields in combination with the increasing availability of biomedical data [12].

Currently, AI in Ophthalmology is mainly focused on improving disease classification and supporting decision-making when treating ophthalmic diseases such as DR, AMD and Glaucoma. In the last couple of months, some implementations of the trending DL based Computer-aided diagnosis (CAD) systems for Glaucoma diagnosis have already reached the product state and are now available as commercial solutions.

An example is Yanbao. Yanbao App is expected to help users conveniently share high-quality glaucoma screening services using the proposed glaucoma screening algorithm based on clinical parameters. To the best of our knowledge, it is the first App specially designed for screening glaucoma. The main advantages of Yanbao App are:

- The App has been developed for smartphones which can be conveniently used at any place and at any time.
- Experiments on the public fundus database and real clinical data demonstrate that the App has good detection and classification accuracy, described in table 2.1.
- Users feedback seems quite promising in terms of real-time testing and user experience.

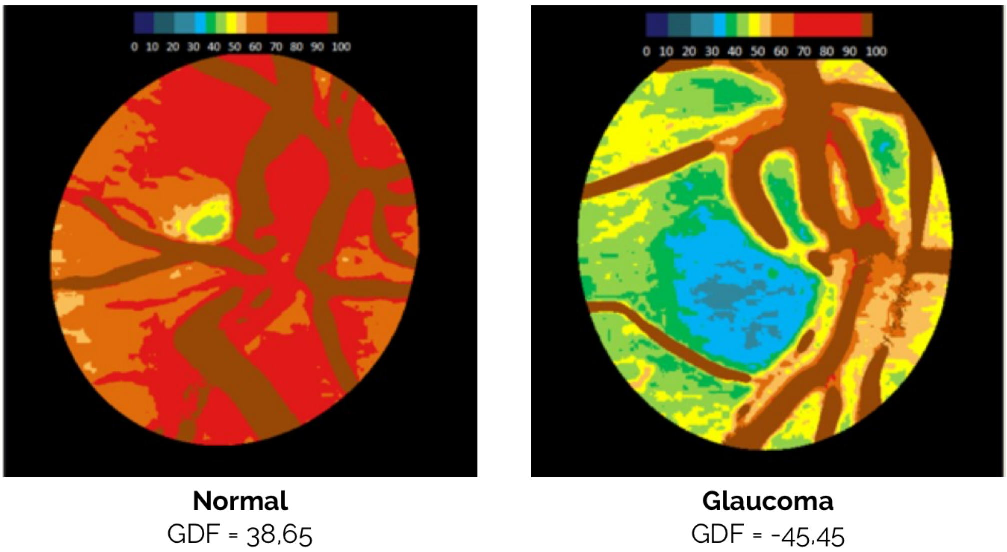
That allows the user to upload a fundus image to a server where four types of feedback are generated, CDR analysis, NRR analysis, glaucoma confidence level, and doctor diagnosis. The first three topics are reportedly returned in about 10 seconds, and the doctor's analysis is dependent on her experience that will interpret the images. In order to provide this feedback, the optic disc is first localized and used to crop the full fundus image and obtain the ROI image. Then, joint segmentation of the optic disc and cup are performed with an enhanced U-net, trained and tested on the ORIGA dataset. These segmentations are then used to perform the CDR and NRR analysis by calculating several morphological features. Afterwise, feature selection is accomplished, and the selected features are used by an SVM classifier that outputs the glaucoma confidence level. The performances are then evaluated once again on the ORIGA dataset.[21].

Classes	**Counts**	**Predictions**	**Accuracy**
Glaucoma	240	183	0.7625
Non-Glaucoma	413	316	0.7651
All Patients	653	499	0.7642

Table 2.1: Yanbao App Accuracy [20]

Another example is Retinalize. Retinalize is a screening software that aids experts conduct eye diseases screening, one of them being Glaucoma. The algorithm behind the system detects signs of eye diseases through fundus imaging analysis and can also be used as a clinical decision support system.

The RetinaLyze System [19] aims to make eye exams affordable to the general public by reducing the cost of each screening, the price, and the complexity of the equipment required to be performed eye screenings. In addition, the new RetinaLyze Glaucoma algorithm allows eye specialists, optometrists, and nurses to safely and efficiently perform glaucoma screenings. This system automatically analyzes retinal images, can assess the risk of having Glaucoma, and quickly gives a result. The process involves only the image from the fundus cameras, excluding the need for a visual field analyzer or tonometer. Figure 2.4 shows a teaser image of the RetinaLyze Glaucoma software at work assesses the level of hemoglobin in the Optic Disc, which can be used to measure damage to the Optic Nerve Head. It calculates the risk of having signs of Glaucoma. The algorithm uses only fundus images as input but achieves similar performance as visual perimetry and OCT screening methods.



Normal
GDF = 38,65

Glaucoma
GDF = -45,45

Figure 2.4: Teaser image of the RetinaLyze Glaucoma software at work. Image from [6]

More recently, the work of Leonardo et. al. [29] compare and evaluate the glaucoma classification when syntetic images with different quality scores where used during the trainning process. The methodology is based on transforming retinal fundus images to improve and degrade their quality to increase training data and evaluate diagnosis, allowing a pipeline to reject samples with lower image quality to avoid classifying these poor-quality images.

2.4.1 Public datasets

Given the need for large amounts of data in DL methods, a few datasets have been made publicly avalable, allowing to develop automatic DL methods for glaucoma detection as well as many other tasks. Some of the most relevant datasets in glaucoma screening (Table 2.2) are:

- **Retinal fundus images for glaucoma analysis (RIGA)**: A de-identified dataset of RIGA was derived from three sources. The optic cup and disc boundaries for each image were marked and annotated manually by six experienced ophthalmologists and included the cup to disc (CDR) estimates. Six parameters were extracted and assessed (the disc area and centroid, cup area and centroid, horizontal and vertical cup to disc ratios) among the ophthalmologists [7].

- **RIM-ONE Release 1**: The first version was published in 2011. The images are classified in different subsets: Normal eye, Early glaucoma, Moderate glaucoma, Deep glaucoma and Ocular hypertension [13].

- **RIM-ONE Release 2**: The images contain annotations of the optic disc boundary and a label indicating the presence of Glaucoma in each fundus image. This release is classified in two different subsets, Normal and Glaucoma, with or without suspicious [9]

- **RIM-ONE Release 3**: Two experts have segmented each image's optic disc and optic cup in ophthalmology to create the ground truth. The average segmentation is also available as the reference segmentation or gold standard [14].

- **Drishti-GS**: The images are divided into two parts as 50 images are for the training and the 51 images are for the testing phase [11].

- **ORIGA**: Contains 650 retinal images annotated by trained professionals from Singapore Eye Research Institute. A wide collection of image signs, critical for glaucoma diagnosis, are annotated [44].

- **iChallange**: Is a dataset of 1200 fundus images with ground truth segmentations and clinical glaucoma labels [34].

Name	**Number of images**	**Number of ophthalmologists**	**Year**
RIM-ONE v1 [13]	169	-	2011
RIM-ONE v2 [9]	455	-	2014
Drishti-GS [11]	101	4	2014
RIM-ONE v3 [14]	159	2	2015
RIGA [7]	750	6	2018
ORIGA [44]	650	-	2010
iChallange [34]	1200	2	2018

Table 2.2: Datasets

Chapter 3

Towards Explainable Privacy-preserving Glaucoma Screening

This chapter will present the state of the art of XAI, GANs and future directions for the development of the dissertation. It focuses on models that are targeted at retinal images.

3.1 Explainable Artificial Intelligence

Is an artificial intelligence that allows humans to interpret the results of a solution. This phenomenon is due to a combination of factors, including concerns over security and privacy, poor generalizability, trust and explainability issues, unfavorable end-user perceptions, and uncertain economic value [33]. XAI can improve the user experience of a product or service by helping end-users trust that the AI is making good decisions. It is AI in which humans can understand the results of the solution. It contrasts with the "black box" concept in machine learning, where even its designers cannot explain why an AI arrived at a specific decision. This way, XAI aims to explain the decision have Glaucoma or not and unveil the information the actions are based on. These characteristics make it possible to confirm existing knowledge to challenge existing knowledge, and to generate new assumptions.

3.1.1 XAI Methods

LRP is one of the main algorithms that aim to explain networks that use the backpropagation algorithm. LRP brings explainability to the prediction of a specific classifier at a given data point by assigning "Relevant Values" (Ri) to essential components in the input, using the topology of the trained model itself. It is used efficiently with images/videos and text where the predicted output value is used to calculate the relevance value for the lower layer neurons. How much more significant is the impact of a neuron in the forward, more significant is its relevance in the backward step. The relevance calculation follows the input where the neurons/features with more relevance are higher values than the others. As a result, the essential input neurons can be highlighted based on which final output layer is visualized. Figure 3.1 shows the flow of the relevance value calculation. Improvements in LRP are an active area of research. [40].

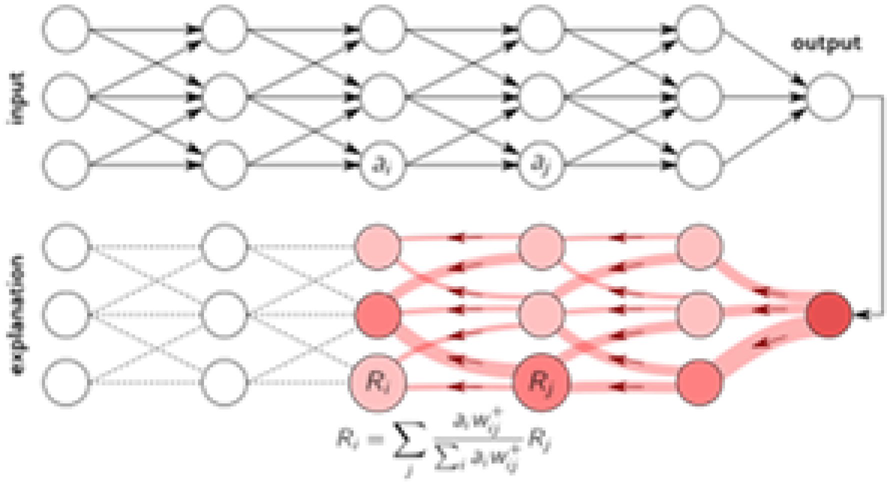


Figure 3.1: LRP model

LIME is one of the most well-known methods for bringing explainability to any classification model. It is considered inaccessible to human understanding because it is not concerned with how the classifier works. It develops a simple, interpretable alternative model, such as a linear regression model around each prediction between the input and corresponding output variables. Using a simple model allows for a better interpretation of the behavior of the classification model in proximity to the instance being predicted. It also tries to understand how the classification model works by introducing noise into a data sample's input variables and understanding how the predictions are affected. Simply put, LIME first generates random noise in the original instance. Second, it computes the similarity between the noises and the original instance. Third, it obtains the predictions for the noises. Then, it chooses a set of noises with better similarity results and calculates weights that represent the effectiveness of these noises on the original instance. A weighted linear regression model is constructed once the noises, predictions, and weights are calculated. The coefficients of this simple model will help explain how changes in the explanatory variables affect the classification outcome for the instance that wishes to be explained. LIME focuses on forming local substitution models to explain individual predictions. Therefore, it can be applied to any DL model [40].

Grad-CAM It maps all the objects in the images to a class, which is used for detection. If we have an image with several objects and only want to determine one of them, the other objects will not be considered. This procedure for ignoring the remaining objects happens because each object has a class defined. In Grad-CAM, the relevant image regions for the decision are identified by using the gradient information flowing into the last convolution layer since this layer is the last one that retains spatial information. These maps were calculated in the classification network (entitled GFI-C), and an illustration of the results with images from both classes can be found in figure 3.2. The images are overlayed with the calculated Grad-CAM maps in the figure, and blue tones indicate the area was not crucial to the classification. In contrast, redder tones indicate a significant influence of that region for the final decision. By analyzing the images, it is noticeable that the network looks to the same structures that ophthalmologists inspect, like the optic disc, cup, and retinal vessels topology [32].

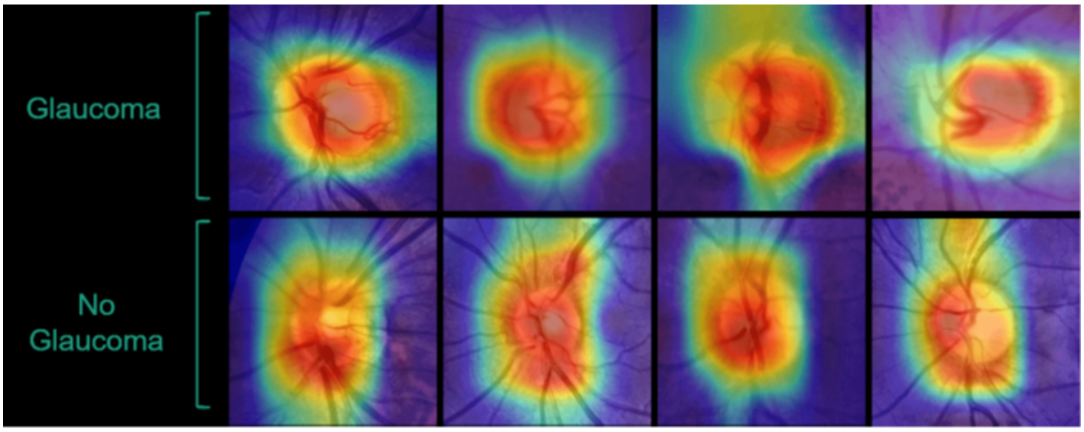


Figure 3.2: Grad-cam activation maps [39] of the GFI-C network

Methods based on examples In contrast to the above XAImethods, which map explanations onto the input space, example-based explanations project explanations across the underlying training data or other representative prototype examples. As explanation-by-example frameworks generate a set of examples as an explanation, achieving explainability is more straightforward because the examples obtained are very similar to the original ones. Generated examples are visualized in the same way an input data instance is visualized. This style of explanation has received considerable attention in recent work.

While LIME or Grad-CAM methods are input-based methods, and these methods have to try to understand the similarity between the input and the output, example-based explanations provides the nearest matching data samples from the training dataset as representative examples. These methods look very promising, and in the development of the dissertation, it will be these example-based methods that we will be explore [27]. However, the retina has unique biometric data, allowing their identification. Thus, it is crucial to develop methods that can generate examples that can justify a decision without allowing patient identification while maintaining all other image characteristics.

3.1.2 Proposed approaches for Privacy-preserving Example-based Explanations

For the development of this dissertation, two approaches can be explored. First, obfuscation of the private information will be considered. This would allow to retrieve examples from any dataset,

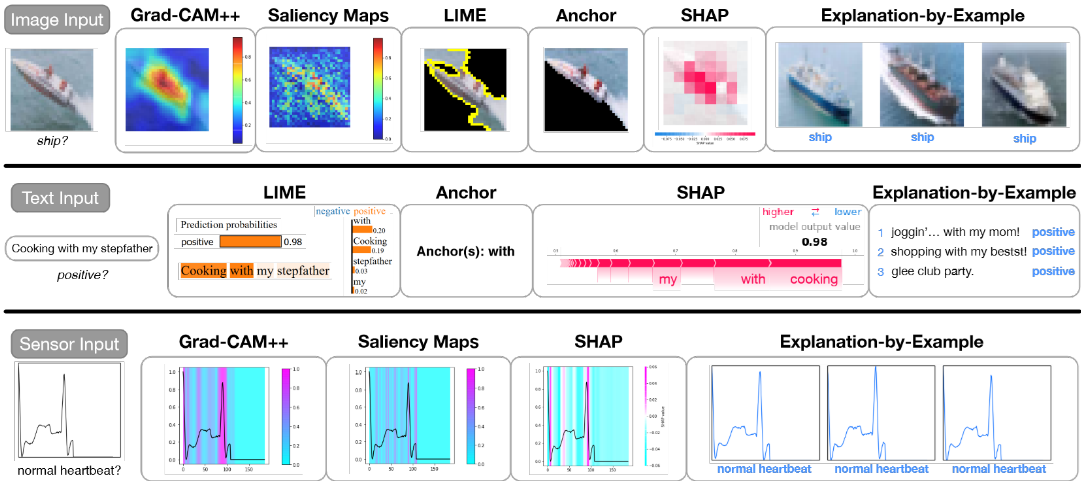


Figure 3.3: Depiction of surveyed explanation methods for image, text, and ECG input. Image from [27]

obfuscating the private data (contained in the vessel structure) while preserving the relevant clinical information (mainly the optic disc), we are talking about methods like Blurring or Blocking, these methods is presented in image 3.4. We can use several techniques in the blurring method, Average Filter, Weighted Average Filter, or Gaussian Filter. This method typically reduces details in the image, introducing some noise. Regarding blocking, it is more effective than blurring in identifying the objects that are obfuscated. Blocking is a rather negative method since it creates a conflict between the privacy of the image and the user experience. The final goal of this method is to replace the part of the image that we want to obfuscate with an object with a single color to hide the original image data completely.

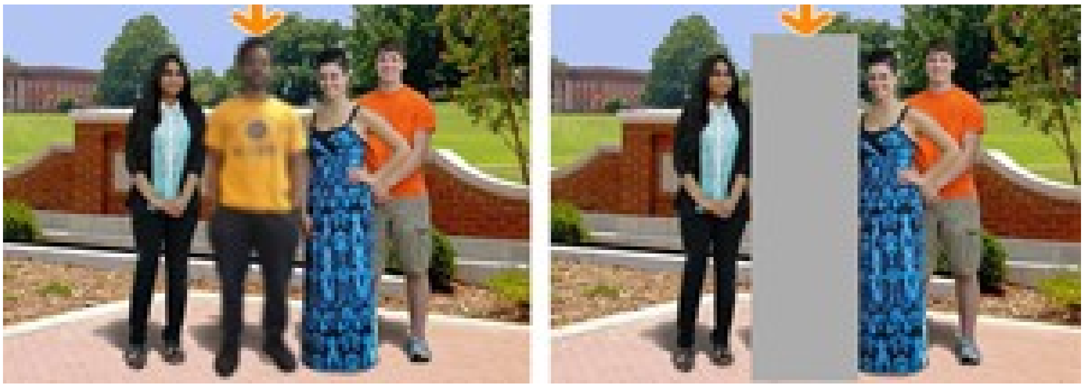


Figure 3.4: Examples of blurring and block obfuscations. Image from [30]

Another very interesting and promising technique is to remove only the blood vessels from the retinal images, which allows the obfuscation of detecting diseases such as Glaucoma not to be compromised. There are already some developments in this technique. The process of these techniques is represented in 3.6 and it works in the following way: at an early stage, we need to make a correct segmentation of the blood vessels based on the analysis of the connected components. After having a successful detection, we need to fill the vessel regions with information that does not allow the identification to create a vessel-free image 3.5 [16].

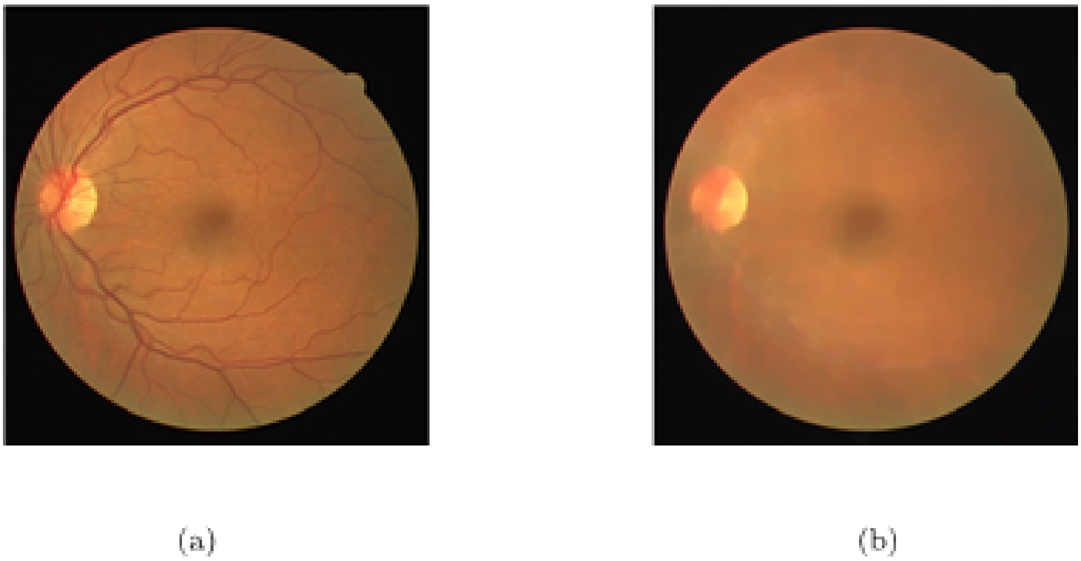


Figure 3.5: **a** Retinal Fundus RGB image, **b** Vessel removed RGB image. Image from [16]

3.2 Generative Adversarial Networks

GANs were first proposed in 2014 by Ian Goodfellow [18]. The Generative Models have gained considerable attention in unsupervised learning via a new and practical framework called GAN due to their outstanding data generation capability. These models are also known to do artificial image generation. This generation is especially used in research since the use of real images turns out to be a bit restricted, not being publicly available. Many GAN models have been proposed, and several practical applications have emerged in various computer vision and machine learning domains. Therefore, constant training is essential to achieve the best possible results. [25]. Given a set of images, generative models aim at artificially generating new images by learning the distribution of the training data. Based on unsupervised learning, generative models generate data from a vector of random numbers, called latent space. However, some models can generate images from other images. The learning phase of a generative model assures that the model creates a correct sample based on the features of the training set. The generative process of data retains value because it naturally expresses casual relations of the context of the data instead of just generalizing from mere correlations. If the training is made correctly and with the correct set, GANs can generate very realistic data.

GANs simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the

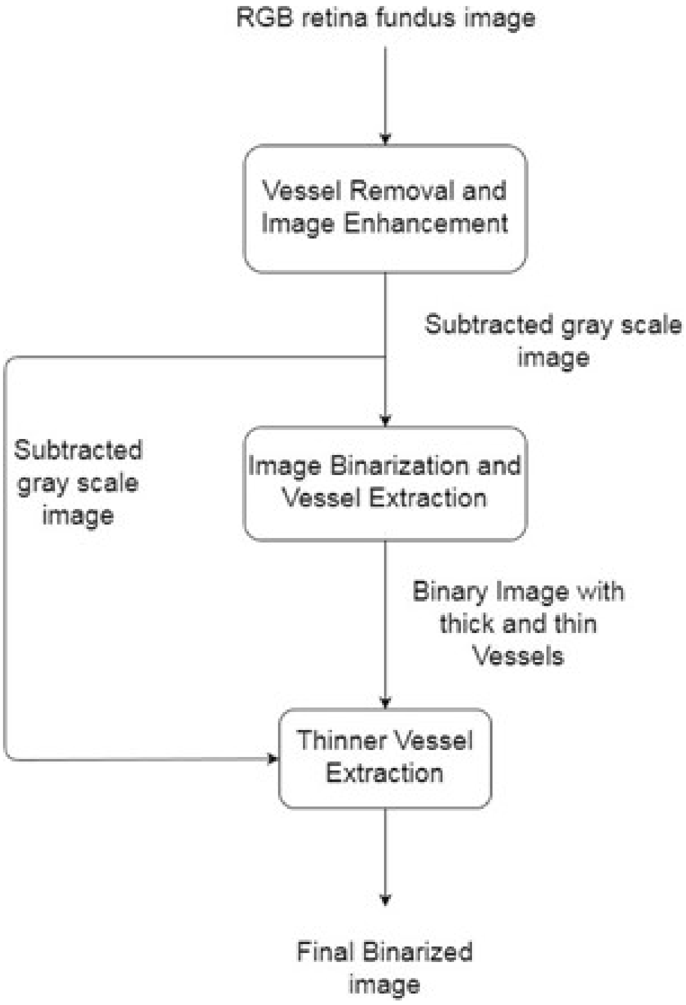

RGB retina fundus image
Vessel Removal and Image Enhancement
Subtracted gray scale image
Subtracted gray scale image
Image Binarization and Vessel Extraction
Binary Image with thick and thin Vessels
Thinner Vessel Extraction
Final Binarized image

Figure 3.6: Flow chart of the complete process of vessel segmentation. Image from [16]

training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. One of the most commonly used comparisons for GANs is a counterfeiter artist and an art expert. G, the counterfeiter, tries to replicate an artist's painting style by learning from a completed artwork, and D, the art expert, classifies the forged artwork as looking real or fake. Both G and D learn from this feedback, iteratively improving the quality of the paintings created until D can no longer distinguish a real from a fake [18].

Generator model

The generative model (presented3.8) takes as input a random vector of fixed length and generates a sample within a respective domain. The input vector is drawn randomly using a Gaussian distribution to create the generative process. After creation, the points in this multidimensional vector space will correspond to points in the problem domain, creating a compressed representation of the data distribution. This vector space is a latent space or a vector space composed of latent variables. Latent variables, or hidden variables, are essential to a domain but are not directly observable. We often refer to latent variables, or a latent space, as a projection or compression of

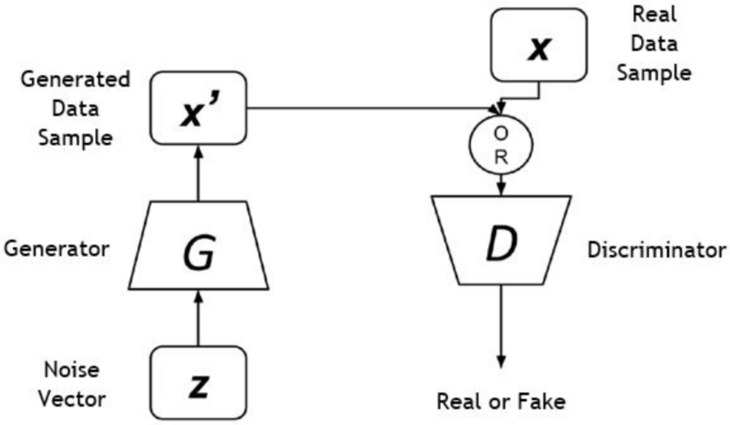


Figure 3.7: GAN representation [8]

the data distribution. A latent space provides compression or high-level concepts of the observed raw data, such as the input data distribution. In the case of GANs, the generating model is applied to points in a chosen latent space. New points extracted from the latent space can be provided to the generating model as input and used to generate new and different output examples. After creation, the generating model is maintained and used to generate new samples [17].

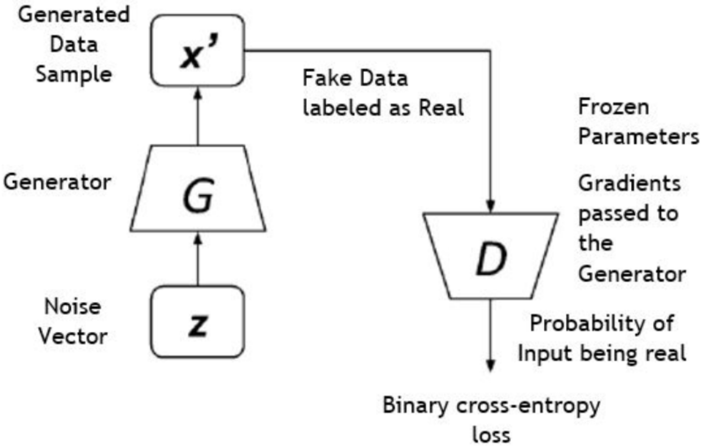


Figure 3.8: Training of a generator model [8]

Discriminator model

The discriminator model (presented in picture 3.9) uses the domain as input (real or generated) and predicts a binary result of original or false (generated). The original example comes from the training data set. The generating model produces the generated examples. The discriminating model makes a typical (and well-understood) classification. After the training process, the discriminator model is no longer used because we are interested in the generator model, which must be well trained before the discriminator model be disabled. Sometimes the generator can be repurposed once it has learned to effectively extract features from examples in the problem domain. Some or all of the feature extraction layers can be used in transfer learning applications using the same or similar input data. [36]

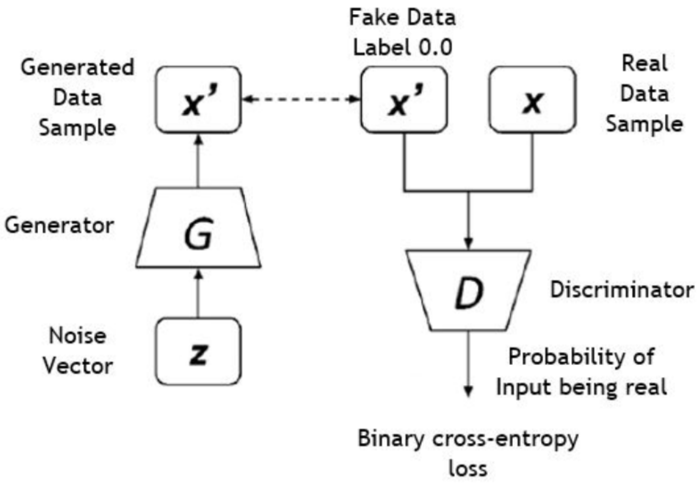


Figure 3.9: Training of a discriminator model [8]

3.2.1 Differential GANs

The generation of artificial images as examples can be considered. However, traditional GANs that allow the generation of new artificial examples have been shown to memorise details from the training dataset and thus reveal the identity of training data on generation. This has motivated the field of differential privacy GANs. These techniques are standard to protect privacy in ML models, trained with sensitive data. Differential privacy is a mathematical framework to define what level of privacy preservation we want. These methods can provide very high privacy guarantees. Several companies have adopted these methods as a standard in data protection. A flow of DP-CGAN is in image 3.10.

Some proposed models focus on differential privacy, such as DP-CGAN, PATE-GAN. One of the big problems with privacy is to be able to maintain it during the training process of the GANs.

It was in this sense that DP-CGAN was proposed. These models aim to achieve just that, to preserve the privacy of conditional GANs using DL. Differential Privacy (DP) is a technique to protect ML models' privacy. Its procedure is to cut the norm of the gradient of the loss sum of the discriminator's real and fake data, and then Gaussian noise is added to the changing gradients. DP-CGAN attempts to solve this problem of matching labels by still cutting the discriminator loss gradient on real and fake data separately, allowing better control on the model's sensitivity to real data and allowing for matching in the labels [38].

Teacher-discriminators are trained to minimize the classification loss when classifying samples as real samples or generated samples. During this step only the parameters of the teachers are updates (and not the generator). The student discriminator is trained using noisy teacher-labelled

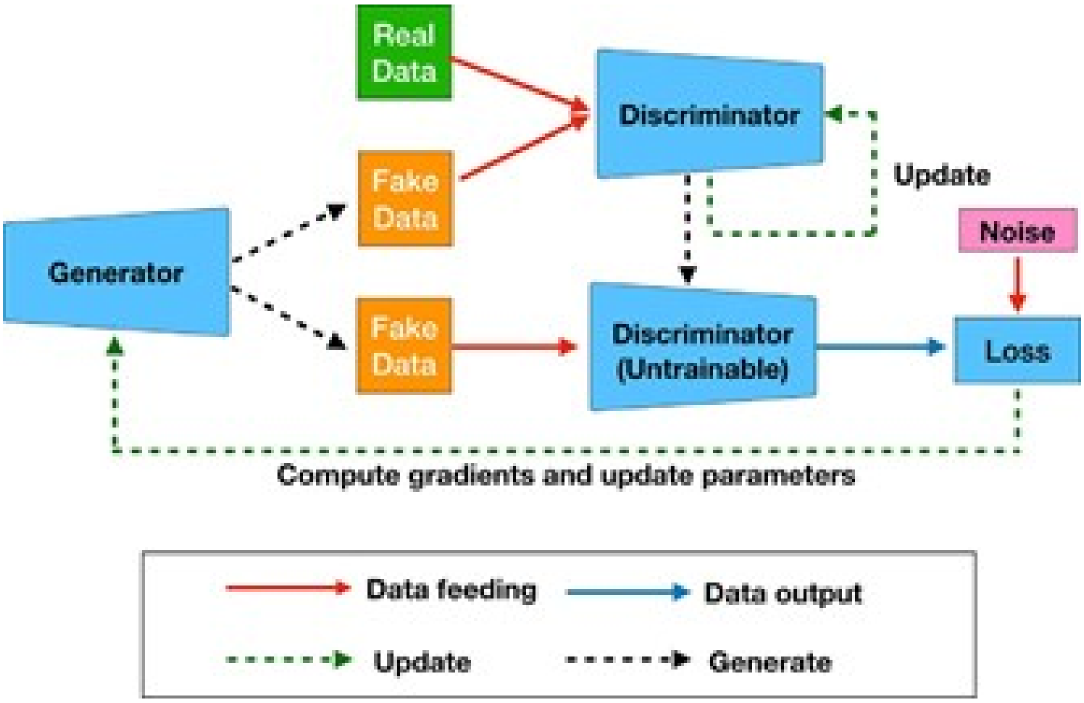


Figure 3.10: The overview of approach to achieving differentially private GANs. Image from [22]

generated samples (the noise provides the DP guarantees). The student is trained to minimize classification loss on this noisily labelled dataset, while the generator is trained to maximize the student loss. Note that the teachers are not updated during this step, only the student and the generator.

On the other hand we have Private Aggregation of Teacher Ensembles Private Aggregation of Teacher Ensembles (PATE-GAN), images 3.11 and 3.12 indicate the iterative training procedure carried out by PATE-GAN, the figures correspond to a single generator update. This model modifies the functioning of normal GANs in order to guarantee privacy. The artificial data is private as far as the original data is concerned. PATE-GANs differ in that the training process of the discriminator has been modified to be differentially private by a modified version of the Private Aggregation of Teacher Ensembles (PATE) framework. Post-processing will ensure that the generator that is trained with the discriminator with differential privacy will also be differentially private, so the data it generates will also be differentially private [43].

3.2.2 GANs directed to ophthalmology

There are several proposals for developing GANs that are targeted at the ophthalmology area. Will be referred to DCGANs, RF-GANs, etc.

Some proposals use DCGANs, but these models do not consider privacy. For disease detection models such as Glaucoma, many images are needed so that the models can be trained more

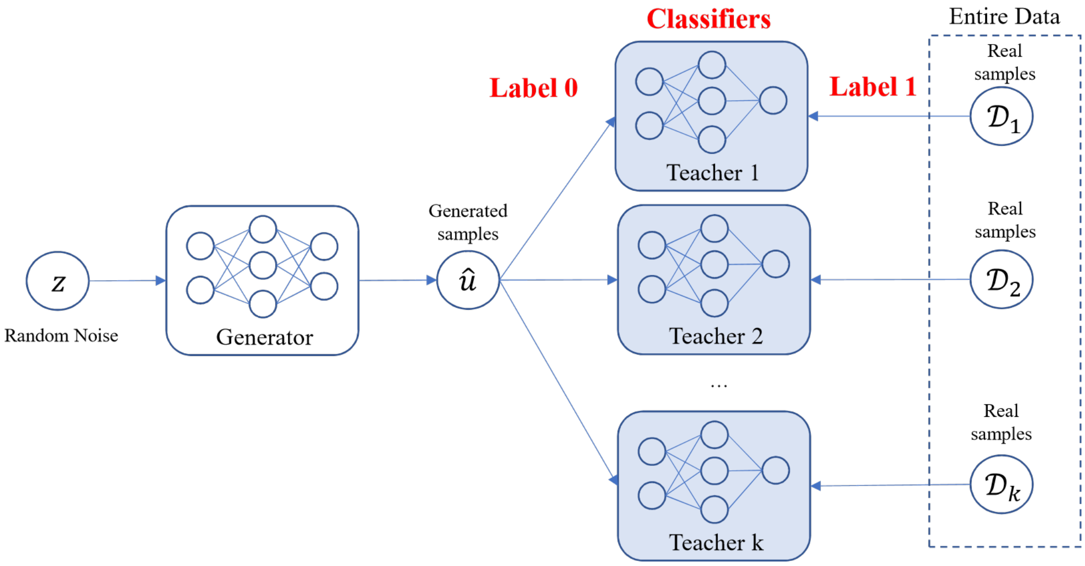


Figure 3.11: Block diagram of the training procedure for the teacher-discriminator during a single generator iteration. Image from [43]

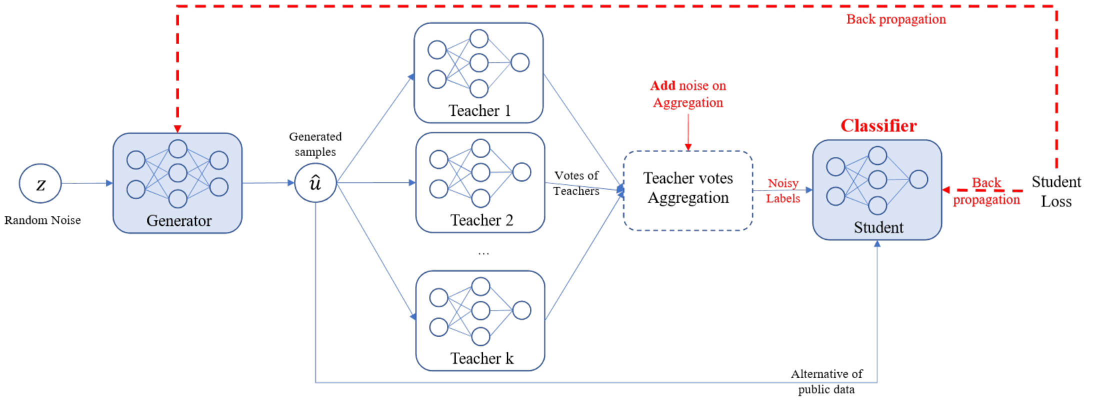


Figure 3.12: Block diagram of the training procedure for the student-discriminator and the generator. Image from [43]

efficiently, but the publicly available datasets are scarce. These models aim to combat this by generating artificial images with diseases or not but to be publicly available so that those who develop detection models can have models trained with several cases. One of these cases is a proposal in the document [37].

RF-GANs are based on GANs but focuses on the retinal fundus images. It is based on two models, RF-GAN1 and RF-GAN2. Model 1 is used to generate retinal fundus images obtained from software sources. The model is trained with the generated images and uses the training results to obtain the structural and lesion masks. While model 2 summarizes the images using the masks and disease classification labels. That is, scoring the images on how real they may appear, verifying the effectiveness of model 1 [10].

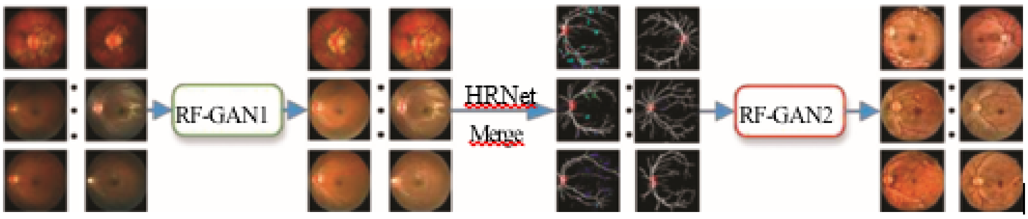


Figure 3.13: The pipeline of synthesizing retinal fundus image. Image from [10]

Chapter 4

Methods and Experiments

This chapter presents the entire flow for development of the experiments with DCGAN, used in retinography imaging for artificial image generation. The methodology adopted and the objectives of each test will be discussed in this chapter.

4.1 Data preparation

The data preparation was divided into two parts: complete retinal images were used, and images with only the optic disc area were used. These images of the optic disc alone were extracted with masks of that same area. This division allowed us to have two large datasets to train two models to understand in which situations better results are achieved.

For complete retinal images, the **Origa**, **Drishti**, **Riga** and **iChallenge** datasets were used. All images from these datasets were pre-prepared by applying a center crop followed by a resize to 256x256, some examples of images used during training 4.1, and for Optical disc area images, the **Origa**, **Acrima** and **iChallenge** datasets were used. All images from these datasets were prepared by cropping them according to the masks so that only the optical disc area was left. This was achieved, because experts segmented each image with optical disc and optical cup areas and then the same process was applied as before, a center crop followed by a resize to 256x256, some examples of images used during training 4.2.

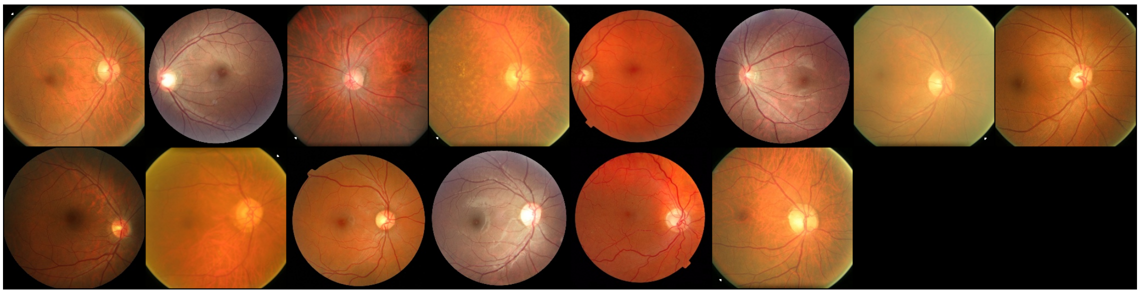


Figure 4.1: Real complete retinal images

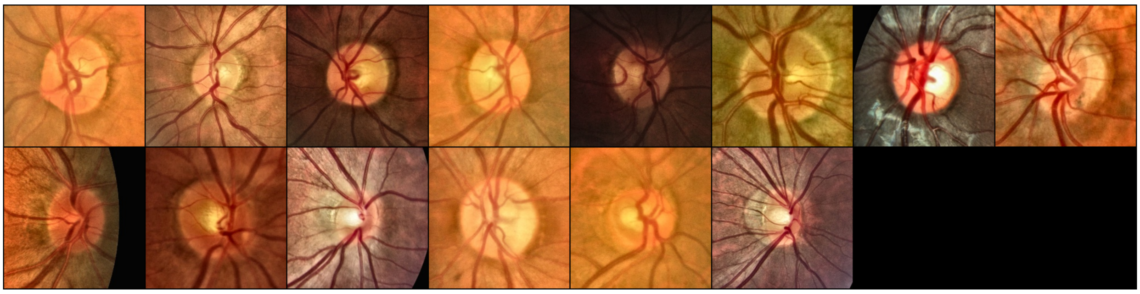


Figure 4.2: Real optical disc area images

4.2 Training configurations

Since two datasets were used, four models were also trained, one with the complete retinal images and one with the images of the optic disc area. Although the images are different, the same settings were applied for both models.The settings are:

- **Batch size during training:** 8
- **Spatial size of training images:** 256
- **Learning rate for optimizers:** 0.00001
- **Number of training epochs:** 2000
- **Size of z latent vector (i.e. size of generator input):** 100

To allow adding more diversity to the datasets, when loading the images into the model, some random transformations were applied, such as flipping horizontally, flipping vertically, and rotations.

The Generator and Discriminator models were based on the models suggested by PyTorch that appear in DCGan tutorials [5]. These models had to be adapted since they are prepared to use images with a size of 64x64. In the training of models, the objective is to use images with a size of 256x256. In the generator and the discriminator, two layers had to be added. In the generator, initially, the suggested by PyTorch is used. In the end, a layer was added to contemplate the convolution from 64 to 128 and another from 128 to 256. In the discriminator, two layers were added at the beginning to contemplate the convolution from 256 to 128 and another from 128 to 64. The image 4.3 shows the system architecture.

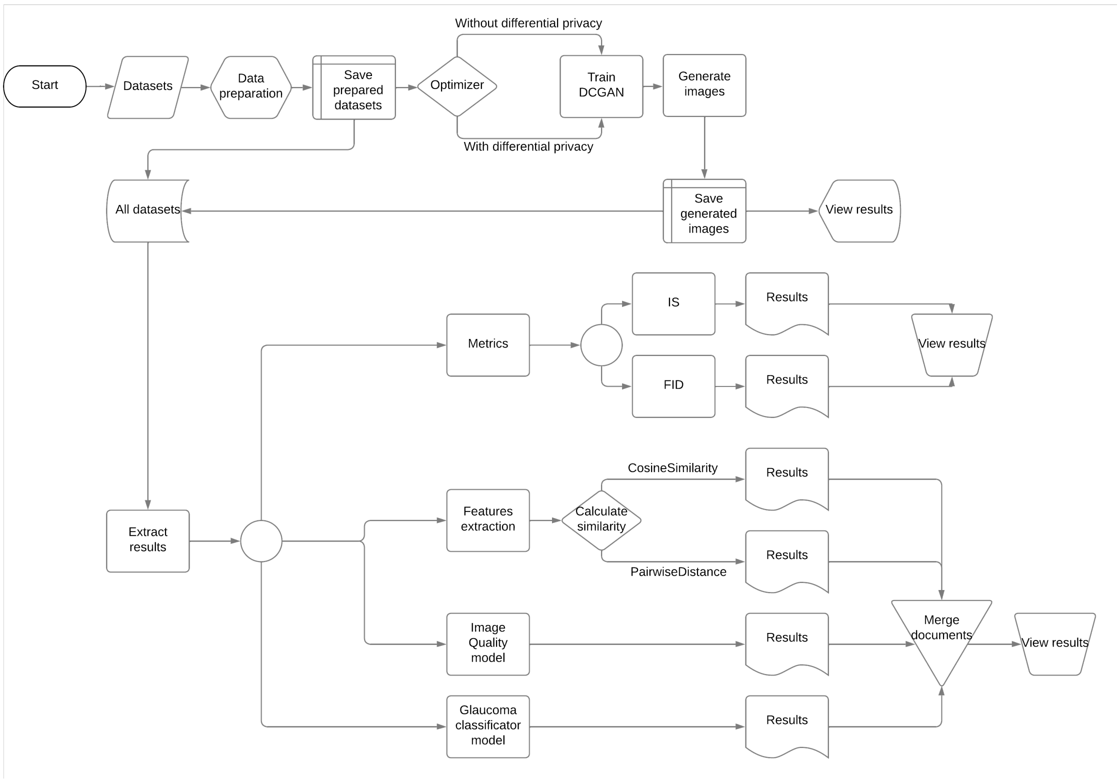


Figure 4.3: System architecture

Another change that had to be made was that when using a small batch size (8), the learning rate also had to be decreased to 0.00001.

The training for both models was performed for 2000 epochs.

Usually, when training DCGANs, a metric must be defined to be met so that the algorithm is not training indefinitely. However, these metrics are helpful when the goal is to generate images very similar to the original ones. Although, in this case, the goal is to generate images that are different from the original ones while maintaining relevant characteristics, therefore no metrics are used so as not to influence the training. Consequently the 2000 epochs were defined as the stopping method.

4.2.1 Loss Functions and Optimizers

With D and G setup, we can specify how they learn through the loss functions and optimizers. We will use the Binary Cross Entropy loss (BCELoss) function which is defined in PyTorch as:

$$l(x,y) = L = \{l_1, \cdots\cdots l_N\}^T \tag{4.1}$$

$$l_n = -\left[y_n.log\left(x_n\right) + \left(1 - y_n\right).log\left(1 - x_n\right)\right] \tag{4.2}$$

Notice how this function provides the calculation of both log components in the objective function (i.e. $log(D(x))log(D(x))$ and $log(1 - D(G(z)))log(1 - D(G(z)))$). We can specify what part of the BCE equation to use with the y input. This is accomplished in the training loop which is coming up soon, but it is important to understand how we can choose which component we wish to calculate just by changing y (i.e. GT labels).

Next, we define our real label as 0.9 and the fake label as 0. These labels will be used when calculating the losses of D and G , and this is also the convention used in the original GAN paper.

In optimizers, we set up two independent optimizers, one without differential privacy and other with differential privacy.

Without differential privacy are Adam optimizer with learning rate 0.00001 and Beta1 = 0.5. For keeping track of the generator's learning progression, we will generate a fixed batch of latent vectors that are drawn from a Gaussian distribution (i.e. fixed_noise) . In the training loop, we will periodically input this fixed_noise into G , and over the iterations we will see images form out of the noise.

With differential privacy are BlurNN. BlurNN is a pytorch-based model privacy-preserving module. This package extends optimizers in torch.optim by extra parameters for differential. The extra parameters are: **norm_bound** is a gradient cliping bound and **noise_scale** is a gaussian noise with standard deviation **noise_scale** * **norm_bound** is added to each clipped gradient privacy [1] In training are used noise_scale = 0.5 and norm_bound = 0.

4.3 Privacy Validation

The feature extraction will be used to obtain a similarity scale between original and generated images. Deep convolutional neural networks have led to a to many breakthroughs in image classification. Deep networks naturally integrate low/medium/high-level features and classifiers in an end-to-end multi-layered fashion. The "levels" of features can be enriched by the number of stacked layers (depth). Recent evidence shows that network depth and critical findings on the challenge are crucial. The ImageNet dataset explores "very deep" models, with a depth of sixteen to thirty [23]. For PyTorch, a package is already available on GitHub called Image 2 Vec (img2vec) [3]. This package includes many templates for extracting features from images. The model chosen was ResNet-18 which allows us to extract 512 features from each image that. These features are extracted from the avgpool layer. Each vector with 512 values identifies an image. The weights used in the ResNet-18 neural network were from ImageNet (`IMAGENET1K_V1`). These weights have excellent results.

After we extracted the features from each image, two methods were used to calculate the similarity between the real and the generated images. These two methods are CosineSimilarity[2] and PairwiseDistance[4]. These two methods are available in Pytorch.

- **CosineSimilarity** Returns cosine similarity between x1 and x2, computed along dim. CosineSimilarity is defined by equation

$$similarity = \frac{x1x2}{max(x1\|2x2\|2, \varepsilon)}$$

- **PairwiseDistance** Computes the pairwise distance between input vectors, or between columns of input matrices. Distances are computed using p-norm, with constant eps added to avoid division by zero if p is negative

$$dist(x,y) = \|x-y+\varepsilon * e\|_p$$

where **e** is the vector of ones and the p-norm is given by.

$$||x||p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

CosineSimilarity returns values between 0 and 1, and PairwiseDistance has no value restriction. It calculates the differences between all positions of the vectors and does the summation. Both methods have very positive results. Below are some of the results obtained and the respective images to see the similarity.

4.4 Metrics

In this section, we will discuss metrics that can be used to evaluate the quality of generated images, more specifically synthetic images, obtained by GAN models, in this case, DCGANs.

4.4.1 Inception Score

The inception score is the most widely used GAN performance metric in the literature. It uses a pre-trained initialization network as the image classification model $\mathscr{M}$ to compute

$$IS = e^{E_{\mathbf{x} \sim p_g}[KL(p_{\mathscr{M}}(y|\mathbf{x})||p_{\mathscr{M}}(y))]} \tag{4.3}$$

where $p_{\mathscr{M}}(y|\mathbf{x})$ is the label distribution of $\mathbf{x}$ that is predicted by the model $\mathscr{M}$ and $p_{\mathscr{M}}(y)$ is the marginal probability of $p_{\mathscr{M}}(y|\mathbf{x})$ over the probability p_g . A larger inception score will have $p_{\mathscr{M}}(y|\mathbf{x})$ close to a point mass and $p_{\mathscr{M}}(y)$ close to uniform, which indicates that the inception network is very confident that the image belongs to a particular ImageNet category and all categories are equally represented. A larger Inception score suggests that the generative model has both high quality and diversity [42]. To run this method, the used weights were **IMAGENET1K_V1**.

4.4.2 Fréchet Inception Distance

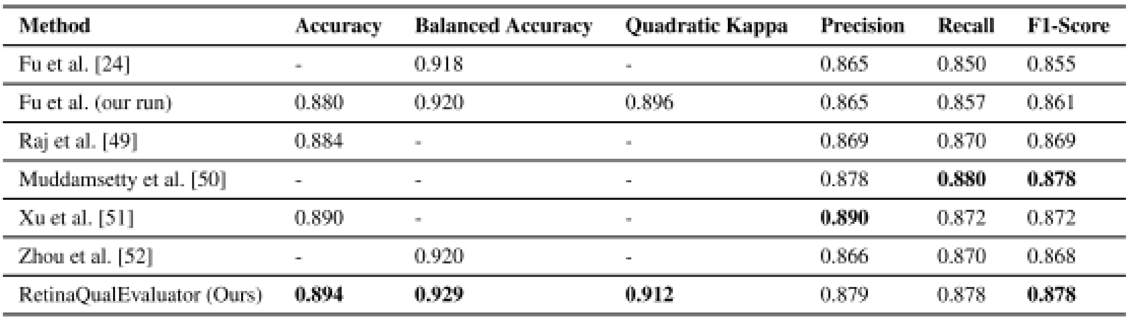
Fréchet inception distance (FID) uses a feature space extracted from a set of generated image samples by a specific layer of the inception network. Regarding the feature space as multivariate Gaussian, the mean and covariance are estimated for both the generated data and real data. FID is computed as $FID(p_r, p_g) = ||\mu_r - \mu_g||_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}})$

A smaller FID indicates better GAN performance [42].

The FID Score algorithm was used for metrics purposes, and the results were obtained for the complete retinal and optical disc area images. FID is a performance metric to evaluate the similarity between two dataset of images. It is shown to correlate well to human evaluation of image quality, and it is able to detect intra-class mode collapse.

4.5 Image quality (RetinaQualEvaluator)

Analysis and classification of retinal images to be possible and assertive, the images must have excellent quality. This quality is not always possible due to several factors, including image clarity (e.g., affected by poor camera focus or poor optics, saccadic eye movements during acquisition, cataracts, macular edema); field definition (e.g., caused by patient orientation, insufficient pupil size, or latent traces) since the image obtained must show the correct area of the retina with visible optic disc and temporal arcades. And in the case of synthetic images, this problem of image quality is very much present. The Retina Quality Evaluator (RetinaQualEvaluator) algorithm aims to distinguish between the three mentioned image quality classes (Good, Usable, and Rejected). An uncropped background image with a full FOV is required as input [29]. The accuracy of this model is 0.894 as shown in 4.4.

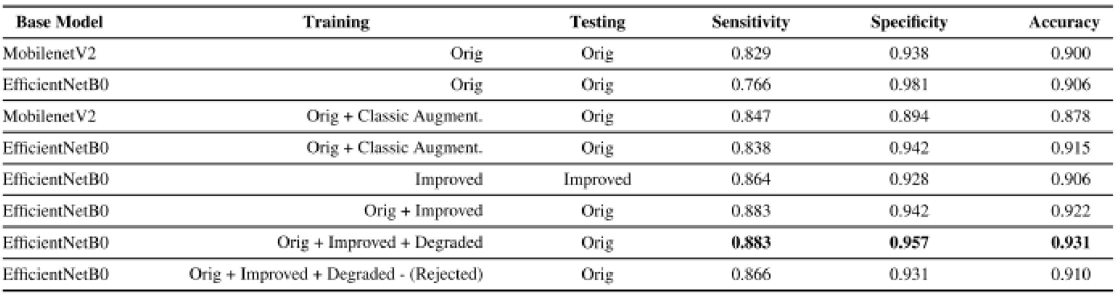


Method	Accuracy	Balanced Accuracy	Quadratic Kappa	Precision	Recall	F1-Score
Fu et al. [24]	-	0.918	-	0.865	0.850	0.855
Fu et al. (our run)	0.880	0.920	0.896	0.865	0.857	0.861
Raj et al. [49]	0.884	-	-	0.869	0.870	0.869
Muddamsetty et al. [50]	-	-	-	0.878	**0.880**	**0.878**
Xu et al. [51]	0.890	-	-	**0.890**	0.872	0.872
Zhou et al. [52]	-	0.920	-	0.866	0.870	0.868
RetinaQualEvaluator (Ours)	**0.894**	**0.929**	**0.912**	0.879	0.878	**0.878**

Figure 4.4: Summary of the results for retinal image quality assessment.

4.6 Glaucoma CADx

This model is used to classify whether a retinal image has Glaucoma. This model was trained with three different scenarios, adding different transformed sets to the original data: a) only enhanced quality images; b) improved and degraded quality images; and c) enhanced and degraded quality images, removing the transformed images classified as Rejected by the previously described RetinaQualEducative Evaluator. This model had an accuracy of 0.931 as shown in 4.5.



Base Model	Training	Testing	Sensitivity	Specificity	Accuracy
MobilenetV2	Orig	Orig	0.829	0.938	0.900
EfficientNetB0	Orig	Orig	0.766	0.981	0.906
MobilenetV2	Orig + Classic Augment.	Orig	0.847	0.894	0.878
EfficientNetB0	Orig + Classic Augment.	Orig	0.838	0.942	0.915
EfficientNetB0	Improved	Improved	0.864	0.928	0.906
EfficientNetB0	Orig + Improved	Orig	0.883	0.942	0.922
EfficientNetB0	Orig + Improved + Degraded	Orig	**0.883**	**0.957**	**0.931**
EfficientNetB0	Orig + Improved + Degraded - (Rejected)	Orig	0.866	0.931	0.910

Figure 4.5: Classification performance of Glaucoma CADx results obtained on the test set using fundus images.

Chapter 5

Results

5.1 Image generation

Image generation was done with the weights obtained after training the four models for complete retinal images and images of the optic disc area only. Two without differential privacy and other two with differential privacy.

5.1.1 Generated complete retinal images

Generated images of the model with complete retinal images can be seen on Figures 5.1 and 5.2. The first is the result of training without differential privacy, while the second used differential privacy in training. The quality of the images could be better, but it is reasonable. Some images look realistic and very representative of the original datasets. The biggest problem might be that the images have a lot of black areas, which interferes a lot with training and learning. During training, applied some random horizontal flip and vertical flip transformations to the original dataset images to have more diversity. And as we can see in the first training image with differential privacy, it was generated with two optical discs. The generated image looks realistic, but it doesn't make any sense anatomically.

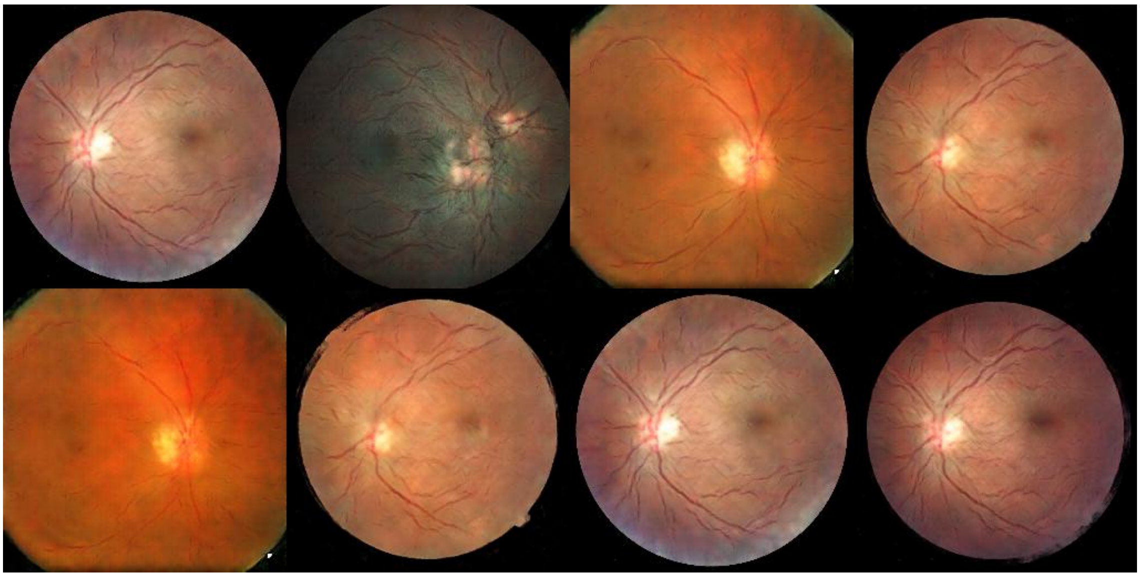


Figure 5.1: Generated Images without differential privacy

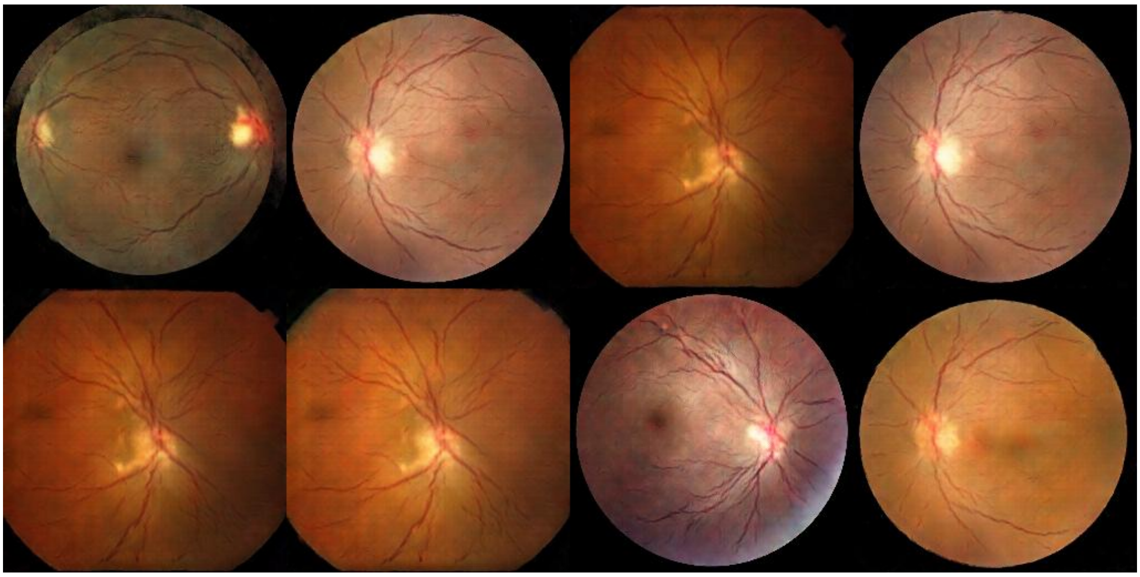


Figure 5.2: Generated Complete retinal images with differential privacy

5.1.2 Generated optical disc area images

Generated images of the model with optical disc area images can be seen on Figures 5.3 and 5.4. The first is the result of training without differential privacy, while the second used differential privacy in training. The quality of the images is quite good. There are quite a few images that look very realistic and very representative of the original datasets. In this dataset, we don't have the problem of black areas on the images, which significantly improves training and learning.

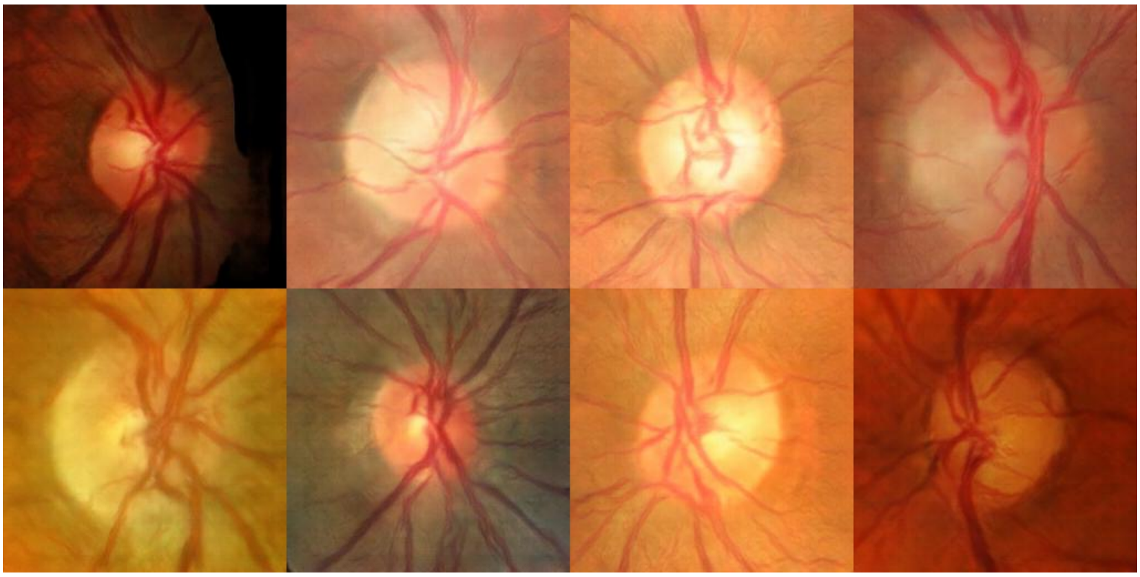


Figure 5.3: Generated Images without differential privacy

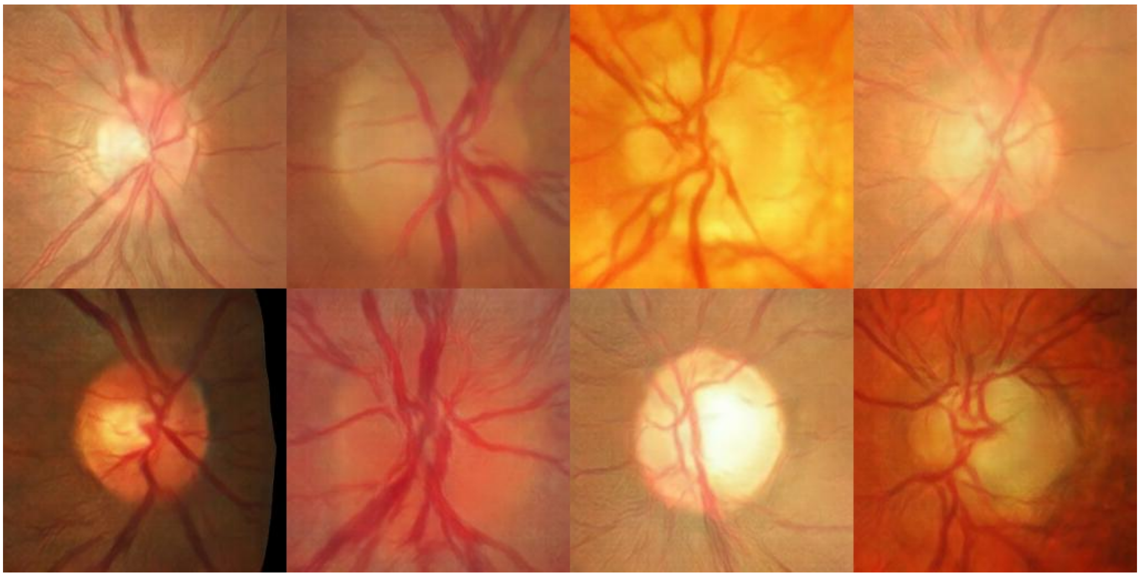


Figure 5.4: Generated images with differential privacy

5.2 Metrics

Result of used algorithms allows us to have a basis for understanding to what extent our generated images are similar to the original ones.

5.2.1 Inception Score

The Inception Score algorithm was used for metrics purposes, and the results were obtained for the complete retinal images and optical disc area images. The results are in the table5.1. When we look at the results of the entire retinal images, the average does not vary much, but the standard deviation is quite different. It is up to 5 times lower compared to the original images' results. In these images, differential privacy had better results than not using it. Since the authentic images have black areas and are very detailed, it affects the generation of the original images. Regarding the images from the optical disc area, the scenario is different. The mean and the standard deviation are closer to the values of the original images and do not vary as much. In these images, the introduction of Differential privacy has worsened the results, although it is not a very significant difference. Images generated from the optic disc area generally have better results than complete retinal images.

	Original images	**Generated images**	
		Without DP	**With DP**
Complete retinal images	1.613±0.168	1.456±0.031	1.683±0.063
Optical Disc area images	1.906±0.203	1.982±0.128	1.811±0.118

Table 5.1: Inception scrore to Complete retinal images and Optical Disc area images

5.2.2 Fréchet Inception Distance

The results are in the table 5.2. This method returns the result of the comparison between two datasets. The original dataset was compared with the images generated without the use of differential privacy and with the use of differential privacy. The difference is insignificant for the full retinal images between the images generated using differential privacy and those generated without differential privacy. For the images of the optic disc area, the scenario is inverted. The use of differential privacy has higher values when compared with the use of differential privacy, which means that when the method is without the use of differential privacy, we have better results.

	Without DP	**With DP**
Complete retinal images	113.746	109.370
Optical Disc area images	94.553	122.649

Table 5.2: FID scrore to Complete retinal images and Optical Disc area images

5.3 Similarity

5.3.1 Complete retinal images

The following tables show the results of the CosineSimilarity and Pairwise algorithms to Complete retinal images. The choice of images was made with those with more significant similarity according to the algorithm and also to have diversity in the results because sometimes an image can appear 4 or 5 times in a row with higher similarity. In the first two tables, we can see the results of the CosineSimilarity algorithm, and in the last two, the results of the Pairwise algorithm. What distinguishes these two pairs of tables is the use of differential privacy. The similarity level obtained for each algorithm can be seen on both tables. There is also the result of the image quality algorithm to see if the closest images can have the same image quality and the difference of quality between each pair of images. The result of the images generated with differential privacy tends to have worse image quality, and the distances between each pair of images are more significant.

Original image	**Generated image**	**Distance**	**Original image quality**	**Generated image quality**	**Difference**
1	627	0.94384	Normal (0.96459)	Normal (1.36426)	0.39967
2	6	0.94126	Good (0.37915)	Normal (1.4723)	1.09315
3	102	0.92127	Normal (1.38269)	Normal (1.19586)	0.18682
4	449	0.88889	Rejected (1.95263)	Normal (1.05155)	0.90108

Table 5.3: Cosine Similarity Example pairs for full images without differential privacy

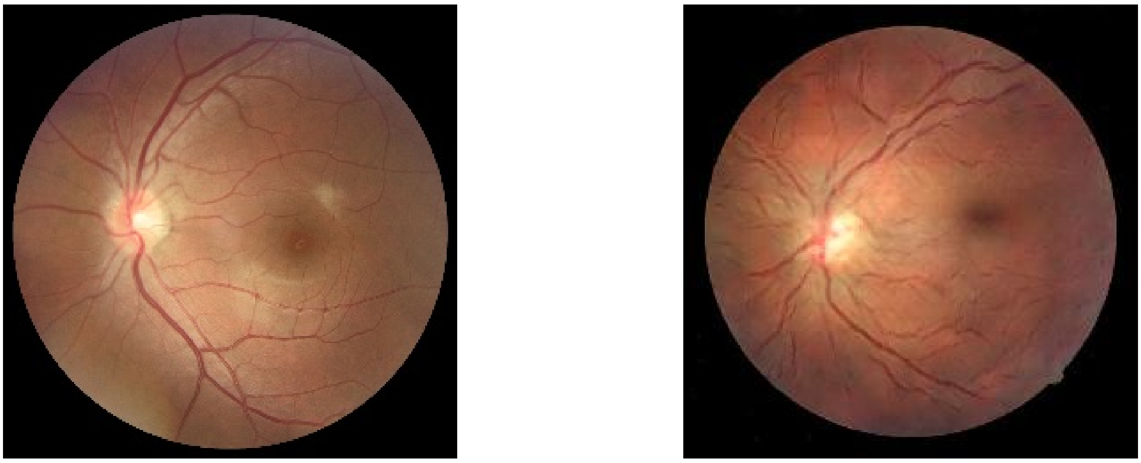


Figure 5.5: Example of real(left)-generated(right) pair of images (first row of Table 5.3). Cosine Similarity = 0.94384

Original image	**Generated image**	**Distance**	**Original image quality**	**Generated image quality**	**Difference**
1	162	0.95741	Normal (0.96729)	Normal (1.2914)	0.3241
2	31	0.95471	Good (0.00762)	Normal (1.30408)	1.29646
3	490	0.95427	Normal (0.98215)	Normal (1.03966)	0.05751
4	307	0.95253	Normal (0.6138)	Normal (1.45136)	0.83757

Table 5.4: Cosine Similarity Example pairs for full images with differential privacy

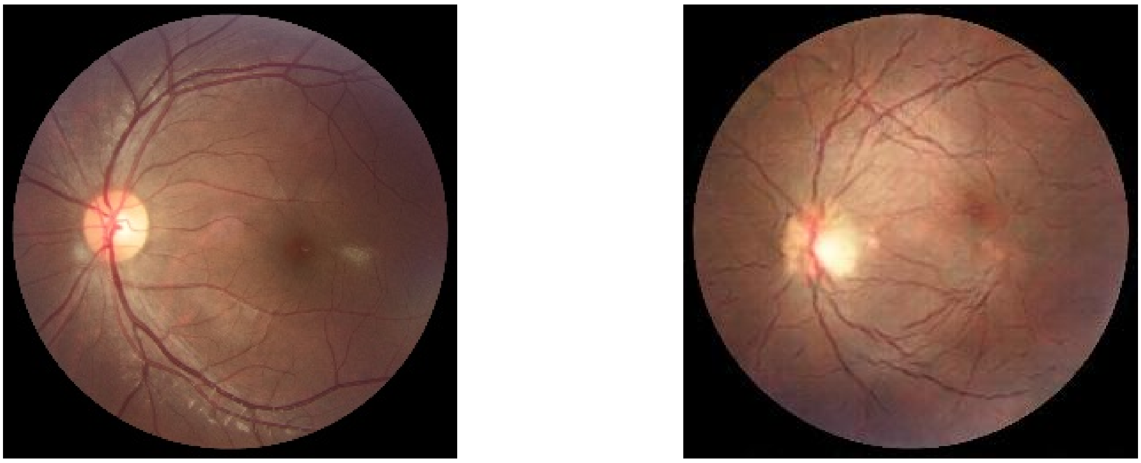


Figure 5.6: Example of real(left)-generated(right) pair of images (first row of Table 5.4). Cosine Similarity = 0.95741

Original image	**Generated image**	**Distance**	**Original image quality**	**Generated image quality**	**Difference**
1	627	128.37801	Normal (0.96459)	Normal (1.36426)	0.39967
2	304	133.81541	Normal (1.02106)	Normal (1.22763)	0.20657
3	193	193.35134	Normal (0.91226)	Normal (1.0051)	0.09283
4	129	223.42607	Rejected (1.95197)	Normal (0.99181)	0.96016

Table 5.5: Pairwise Example pairs for full images without differential privacy

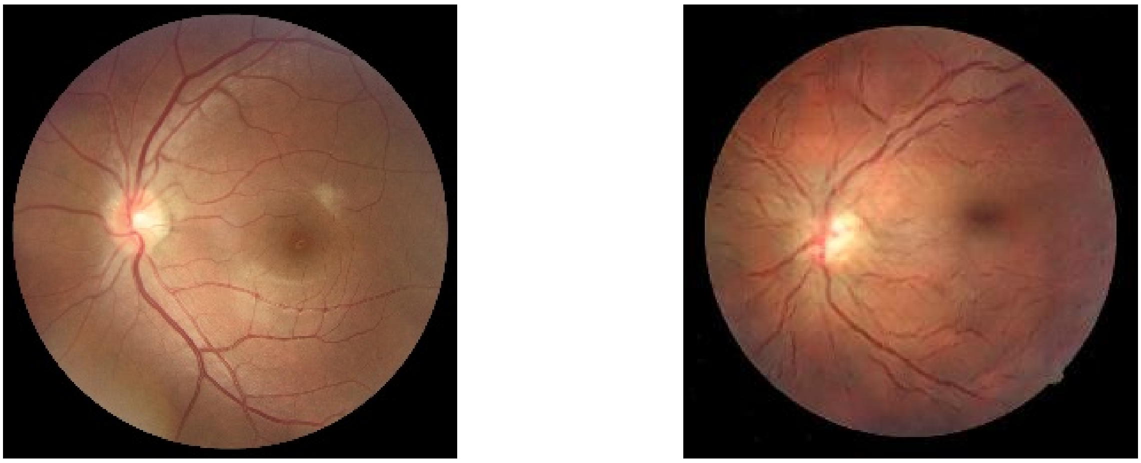


Figure 5.7: Example of real(left)-generated(right) pair of images (first row of Table 5.5). Pairwise = 128.37801

Original image	**Generated image**	**Distance**	**Original image quality**	**Generated image quality**	**Difference**
1	453	120.10165	Normal (0.74973)	Normal (1.0199)	0.27017
2	307	121.29808	Normal (0.97319)	Normal (1.45136)	0.47817
3	332	121.73177	Normal (0.6138)	Rejected (1.84479)	1.231
4	448	123.19652	Normal (0.75526)	Normal (1.05529)	0.30003

Table 5.6: Pairwise Example pairs for full images with differential privacy

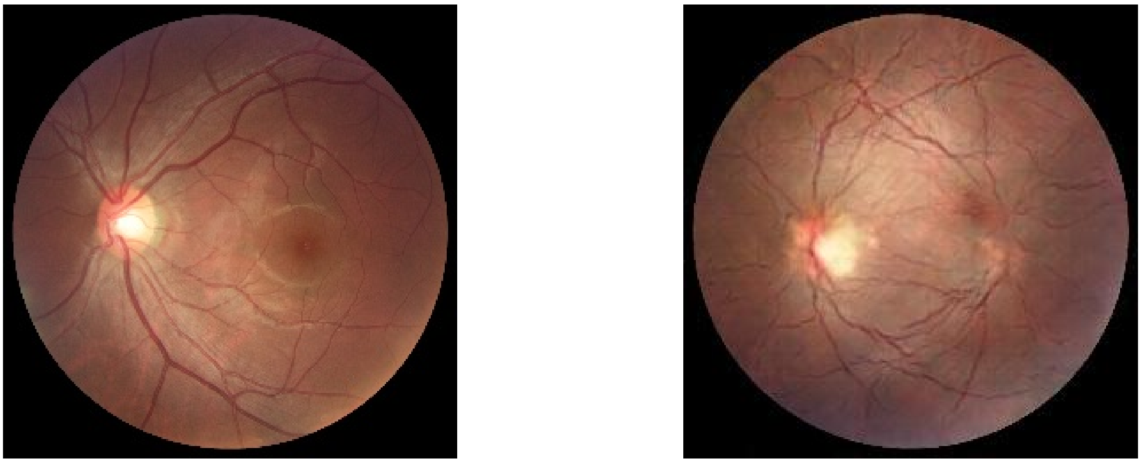


Figure 5.8: Example of real(left)-generated(right) pair of images (first row of Table 5.6). Pairwise = 120.10165

5.3.2 Optical disc area images

The following tables show the results of the CosineSimilarity and Pairwise algorithms for disc area images. The choice of images was made with those with higher similarity according to the algorithm and also to have diversity in the results because sometimes an image can appear 4 or 5 times in a row with greater similarity. In the first two tables, we can see the results of the CosineSimilarity algorithm, and in the last two, the results of the Pairwise algorithm. What distinguishes these two pairs of tables is the use of differential privacy. In all tables, you can see the level of similarity obtained, depending on the algorithm. The result of the Glaucoma classification algorithm is also presented to see if the closest images can have the same result of Glaucoma classification and the difference in rates between each pair of images. The result of the images generated with differential privacy tends to have worse image quality, and the distances between each pair of images are more significant.

Original image	**Generated image**	**Distance**	**Original glaucoma**	**Generated glaucoma**	**Difference**
1	115	0.93336	Glaucoma (0.55552)	Glaucoma (0.64043)	0.08491
2	130	0.93098	Glaucoma (0.67133)	Glaucoma (0.63962)	0.03171
3	261	0.93064	Non Glaucoma (0.38465)	Glaucoma (0.57507)	0.19042

Table 5.7: Cosine Similarity Example pairs for optical disc images without differential privacy

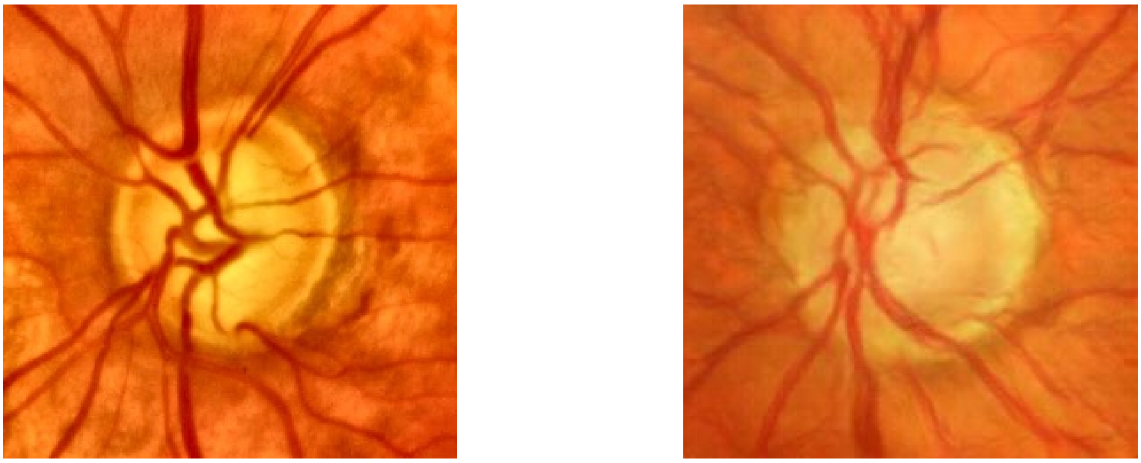


Figure 5.9: Example of real(left)-generated(right) pair of images (third row of Table 5.7). Cosine Similarity = 0.93064

Original image	**Generated image**	**Distance**	**Original glaucoma**	**Generated glaucoma**	**Difference**
1	238	0.93727	Glaucoma (0.6032)	Glaucoma (0.59223)	0.01098
2	140	0.9333	Glaucoma (0.63826)	Glaucoma (0.62611)	0.01215
3	607	0.93281	Glaucoma (0.61952)	Glaucoma (0.60145)	0.01807
4	606	0.92934	Glaucoma (0.6278)	Glaucoma (0.6243)	0.00349

Table 5.8: Cosine Similarity Example pairs for optical disc images with differential privacy

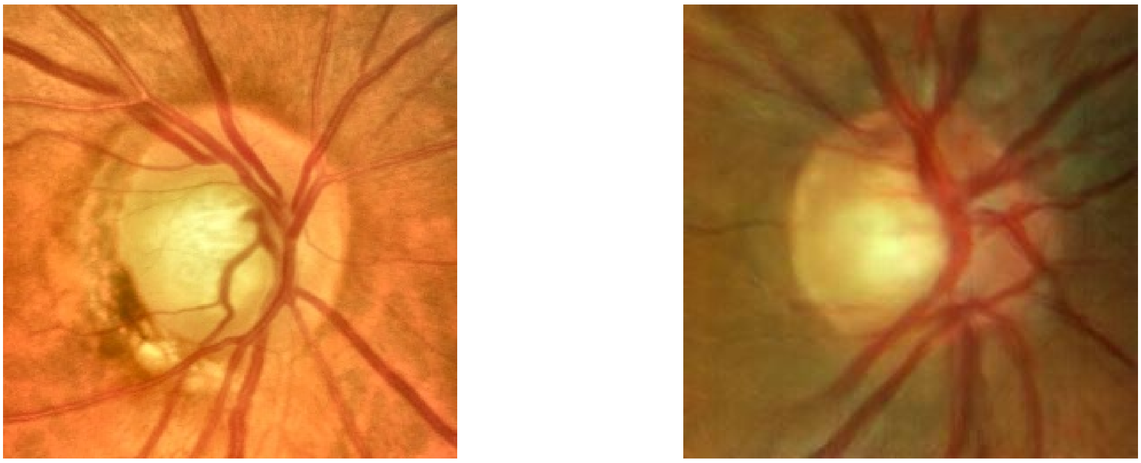


Figure 5.10: Example of real(left)-generated(right) pair of images (first row of Table 5.8). Cosine Similarity = 0.93727

Original image	**Generated image**	**Distance**	**Original glaucoma**	**Generated glaucoma**	**Difference**
1	344	137.05790	Glaucoma (0.65244)	Glaucoma (0.64682)	0.00562
2	243	140.08910	Glaucoma (0.65363)	Glaucoma (0.68209)	0.02846
3	63	144.65698	Glaucoma (0.62623)	Glaucoma (0.68062)	0.05439
4	56	146.43518	Glaucoma (0.64628)	Glaucoma (0.53443)	0.11184

Table 5.9: Pairwise Example pairs for optical disc images without differential privacy

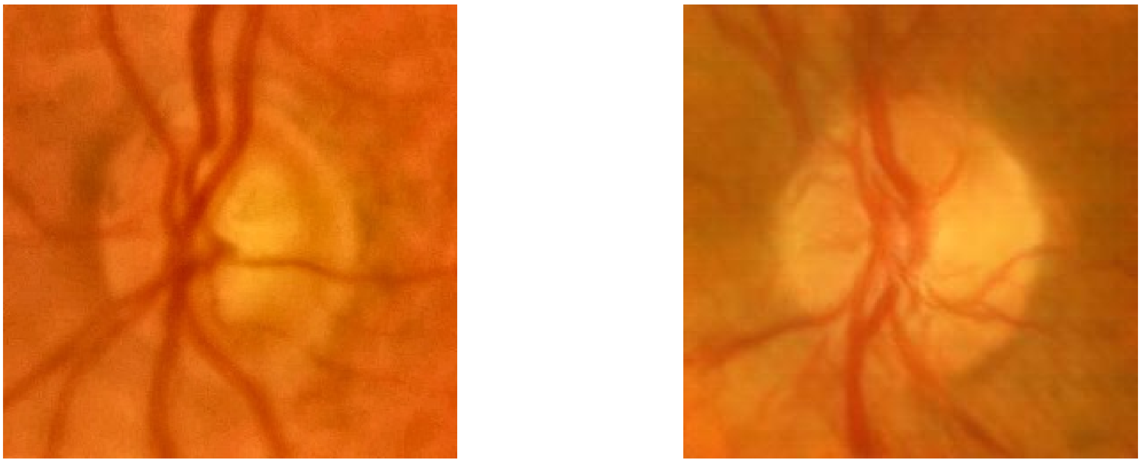


Figure 5.11: Example of real(left)-generated(right) pair of images (first row of Table 5.9). Pairwise = 137.05790

Original image	**Generated image**	**Distance**	**Original glaucoma**	**Generated glaucoma**	**Difference**
1	626	126.2094	Glaucoma (0.65363)	Glaucoma (0.62644)	0.02719
2	363	130.15157	Glaucoma (0.65363)	Glaucoma (0.57895)	0.07468
3	554	135.97733	Non Glaucoma (0.47561)	Glaucoma (0.75713)	0.28152
4	496	137.51332	Glaucoma (0.62623)	Glaucoma (0.69772)	0.07149

Table 5.10: Pairwise Example pairs for optical disc images with differential privacy

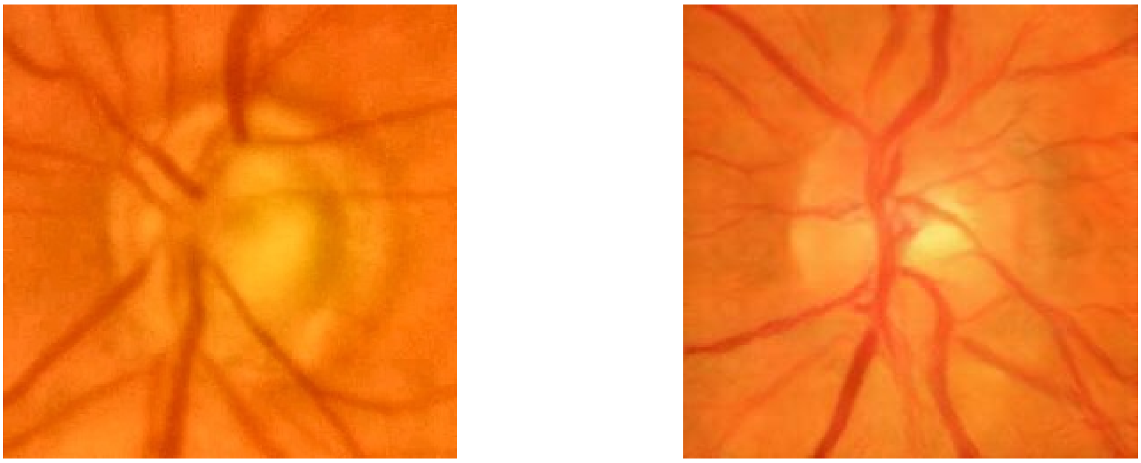


Figure 5.12: Example of real(left)-generated(right) pair of images (first row of Table 5.10). Pairwise = 126.2094

5.4 Image quality

For each original image, that is 2606 images, the five most similar images for each original image were taken, making a total of 13030 comparisons. In these comparisons, the image quality algorithm was applied. The image quality is composed of a scale with three values, Good, Normal, and Rejected. The image quality program returns a value between 0 and 2, and to adapt to the scale above it was defined that values between 0 and 0.5 is Good, between 0.5 and 1.5 is Normal and 1.5 and 2 is Rejected. This analysis is only performed for complete retinal images.

In table 5.11 we can see the total number of images classified as Good, Normal, and Rejected. Obtained these results for the original and generated images, both with and without differential privacy. The introduction of differential privacy had a significant loss with the classification by this method. Without differential privacy, although classified no images as good, many are classified as normal.

	Original images	**Generated images**	
		Without DP	**With DP**
Good	3050	0	0
Normal	7015	11256	2936
Rejected	2965	1774	10094

Table 5.11: Total image quality cases

In table 5.12 we can see the number of matches for each pair of original images and generated images with higher similarity and the same image quality result. The percentage of matches was about 50% without differential privacy and about 35% with differential privacy, concluding that differential privacy negatively influences the quality of the generated images.

	Without DP	With DP
Number of images	6701	4625
Percentage	51,42748	35,49501

Table 5.12: Image quality matches

In table 5.13 we can see the mean and standard deviation of the sum of differences of the image quality algorithm result for all pairs obtained with the highest similarity. That is, an original image was rated 1.0, and the generated one was rated 1.5, which results in a difference of 0.5. These values are for all our differences for the 13030 comparisons.

Without DP	With DP
0,59074±0,40755	0,72238±0,54955

Table 5.13: Average and std: Complete retinal images

5.5 Glaucoma classification

For each original image, that is, 2926 images, the five most similar pictures for each original image were taken, making a total of 14630 comparisons. The glaucoma classification algorithm was applied to all the obtained comparisons in these comparisons. For each original picture, it was classified whether it had glaucoma or not, and for the five most similar images as well. The results obtained were as follows.

In table 5.14 we can see the total number of images classified as Glaucoma and Non-Glaucoma. These results where obtained for the original and generated images, both with and without differential privacy. The introduction of differential privacy slightly increased the number of cases with Non-Glaucoma compared to not using it, while without differential privacy, the values are very close to the values of the original images. Overall the two methods are excellent.

	Original images	Generated images	
		Without DP	With DP
Glaucoma	13950	14154	13221
Non-Glaucoma	680	476	1409

Table 5.14: Total glaucoma and non-glaucoma cases

In table 5.15 we can see the number of matches for each pair of original images and generated images with higher similarity and the same result of the Glaucoma classification algorithm. The percentage of matches was about 92% without differential privacy and about 86% with differential privacy, so it can be concluded that both methods are excellent.

	Without DP	**With DP**
Number of images	13518	12643
Percentage	92,39918	86,41832

Table 5.15: Glaucoma classificator matches

In table 5.16 we can see the mean and standard deviation of the sum of differences of the glaucoma classification algorithm result for all pairs obtained with the highest similarity. That is, an original image was rated 0.7, and the generated one was rated 0.63, which results in a difference of 0.07. These values are for all our differences for the 14630 comparisons.

Without DP	**With DP**
0.07174±0.05361	0,61277±0,06597

Table 5.16: Average and std: Optical disc area images

Chapter 6

Conclusions

The state of the art of deep learning models for automated Glaucoma classification systems has been progressively improving. However, there are still significant limitations due to the quantity and quality of available data, lack of knowledge about the decisions made by these models, and also because the images themselves can contain the identity of the patients. This dissertation aimed to investigate and develop the artificial generation of Retina Images when trained with several datasets, introducing the feature of differential privacy. The ultimate goal is to generate retinal images similar to the original images but manage to hide the patient's identity by modifying the blood vessels. The process began by understanding the need for the diagnosis of Glaucoma and automatic analysis of retinography, followed by a review of the best datasets available. Second, the state of the art of image generation and evaluation was reviewed and researched, focusing on adverse generative networks and commonly used evaluation metrics. The contributions of this thesis are based on the DCGAN development methodology to explore the potential of architecture and evaluation methods to assess its performance successfully. The images intended to be generated can be different from the original ones to be able to hide the identity of the patients. The developed models achieved adequate performance, generating images capable of obtaining a classification of Glaucoma or even analyzing the quality of images. Although the generation was done with two groups of datasets, the images of the optical disc area achieved excellent results. In complete retinal images, there is a lot of detail and black zones that interfere a lot with learning artificial intelligence algorithms. In addition to having two datasets, two optimizers were also used, one without differential privacy and the other with differential privacy. But the optimizer with differential privacy brought us worse results, generating images with lower quality and a lower percentage of success in classifying Glaucoma. This quality and classification were made with original datasets and with the images generated with the models that do not use differential privacy and the models that do. These two methods reflect what has been said before. Metrics to calculate the similarity between the generated and original images were also extracted. Some of the best similarity ratings even reflect a high similarity between the original and synthetic images. To ensure a complete validation of the developed solutions, quantitative methods of evaluation were employed. The quantitative evaluation was ensured by the FID and IS. The results showed that these metrics, when combined, complement each other and provide a useful quantitative evaluation of GANs. Nevertheless, these metrics are known to have limitations. Overall, the development of this work was essential for understanding the needs and limitations of generative models in retinal imaging applications with differential privacy. Did it in a structured way with a thorough evaluation of the performance of the models. The architecture used, the DCGAN, although it is already an established architecture, allowed a fast development, and has shown to be efficient for developing a quality GAN capable of generating retinal images with enough similarity with the original ones.

6.1 Future work

For the qualitative evaluation, a questionnaire was developed, and added several randomly distributed images, both original and synthetic images. For each image, there are two questions, one about whether the image is real or generated, and another to classify the presence of Glaucoma with the answers Normal, Suspect, and Not classifiable. This questionnaire was done in the dissertation's final phase, so there were no results.

References

[1] Blurnn. https://github.com/ailabstw/blurnn. Accessed: 2023-01-04.

[2] Cosinesimilarity. https://pytorch.org/docs/stable/generated/torch.nn.CosineSimilarity.html#torch.nn.CosineSimilarity. Accessed: 2022-11-10.

[3] Img2vec. https://github.com/christiansafka/img2vec. Accessed: 2022-11-10.

[4] Pairwisedistance. https://pytorch.org/docs/stable/generated/torch.nn.PairwiseDistance.html#torch.nn.PairwiseDistance. Accessed: 2022-11-10.

[5] Pytorch. https://pytorch.org/tutorials/beginner/dcgan_faces_tutorial.html. Accessed: 2022-10-20.

[6] Retinalyze. https://www.retinalyze.com/post/retinalyze-glaucoma-a-revolution-in-glaucoma-screening. Accessed: 2022-02-16.

[7] Ahmed Almazroa, Sami Alodhayb, Essameldin Osman, Eslam Ramadan, Mohammed Hummadi, Mohammed Dlaim, Muhannad Alkatee, Kaamran Raahemifar, and Vasudevan Lakshminarayanan. Retinal fundus images for glaucoma analysis: the RIGA dataset. In *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, volume 10579 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 105790B, March 2018.

[8] Rowel Atienza. *Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, deep RL, unsupervised learning, object detection and segmentation, and more*. Packt Publishing Ltd, 2020.

[9] Tomaz Ribeiro Viana Bisneto, Antonio Oseas de Carvalho Filho, and Deborah Maria Vieira Magalhães. Generative adversarial network and texture features applied to automatic glaucoma detection. *Applied Soft Computing*, 90:106165, 2020.

[10] Yu Chen, Jun Long, and Jifeng Guo. Rf-gans: A method to synthesize retinal fundus images based on generative adversarial network. 2021.

[11] Omer Deperlioglu, Utku Kose, Deepak Gupta, Ashish Khanna, Fabio Giampaolo, and Giancarlo Fortino. Explainable framework for glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. *Future Generation Computer Systems*, 129:152–169, 4 2022.

[12] Hiroshi Fujita, Yoshikazu Uchiyama, Toshiaki Nakagawa, Daisuke Fukuoka, Yuji Hatanaka, Takeshi Hara, Gobert N. Lee, Yoshinori Hayashi, Yuji Ikedo, Xin Gao, and Xiangrong Zhou. Computer-aided diagnosis: The emerging of three cad systems induced by japanese health care needs. *Computer Methods and Programs in Biomedicine*, 92:238–248, 12 2008.

[13] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2011.

[14] Francisco Fumero, Jose Sigut, M Alayón, Silvia andGonzález-Hernández, and M González de la Rosa. Interactive tool and database for optic disc and cupsegmentation of stereo and monocular retinal fundus images. 06 2015.

[15] Shunxiang Gao, Qian Li, Shenghai Zhang, Xinghuai Sun, Hong Zhou, Zhongfeng Wang, and Jihong Wu. A novel biosensing platform for detection of glaucoma biomarker gdf15 via an integrated bli-elasa strategy. *Biomaterials*, 294, 2023. Cited by: 0.

[16] Ranjit Ghoshal, Aditya Saha, and Sayan Das. An improved vessel extraction scheme from retinal fundus images. 2019.

[17] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning (adaptive computation and machine learning series). *Cambridge Massachusetts*, pages 321–359, 2017.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[19] Andrzej Grzybowski and Piotr Brona. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: Idx-dr and retinalyze. *Journal of Clinical Medicine*, 10(11), 2021. Cited by: 5; All Open Access, Gold Open Access, Green Open Access.

[20] Fan Guo, Yuxiang Mai, Xin Zhao, Xuanchu Duan, Zhun Fan, Beiji Zou, and Bin Xie. Yanbao: A mobile app using the measurement of clinical parameters for glaucoma screening. *IEEE Access*, 6:77414–77428, 2018.

[21] Fan Guo, Yuxiang Mai, Xin Zhao, Xuanchu Duan, Zhun Fan, Beiji Zou, and Bin Xie. Yanbao: A mobile app using the measurement of clinical parameters for glaucoma screening. *IEEE Access*, PP:1–1, 11 2018.

[22] Chunling Han and Rui Xue. Differentially private gans by adding noise to discriminator's loss. *Computers Security*, 107:102322, 2021.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[24] Elitsa Hristova, Darina Koseva, Zornitsa Zlatarova, and Klara Dokova. Diabetic retinopathy screening and registration in europe—narrative review. *Healthcare*, 9(6), 2021.

[25] A. Jabbar, X. Li, and B. Omar. A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys*, 54, 2022.

[26] S. Jain, S. Indora, and D.K. Atal. Rider manta ray foraging optimization-based generative adversarial network and cnn feature for detecting glaucoma. *Biomedical Signal Processing and Control*, 73, 2022.

[27] Jeya Vikranth Jeyakumar, Joseph Noor, Yu-Hsi Cheng, Luis Garcia, and Mani Srivastava. How can i explain this to you? an empirical study of deep neural network explanation methods. *Advances in Neural Information Processing Systems*, 33, 2020.

[28] University of Michigan Health Kellogg Eye Center. Anatomy of the eye. Online; accessed 25-January-2022.

[29] Ricardo Leonardo, João Gonçalves, André Carreiro, Beatriz Simões, Tiago Oliveira, and Filipe Soares. Impact of generative modeling for fundus image augmentation with improved and degraded quality in the classification of glaucoma. *IEEE Access*, 10:111636–111649, 2022.

[30] Y. Li, N. Vishwamitra, B.P. Knijnenburg, H. Hu, and K. Caine. Blur vs. block: Investigating the effectiveness of privacy-enhancing obfuscation for images. volume 2017-July, pages 1343–1351, 2017. cited By 21.

[31] Yanxiu Li, Xiaobo Xia, and Yannis M. Paulus. Advances in retinal optical imaging. *Photonics*, 5, 6 2018.

[32] José Martins, Jaime S. Cardoso, and Filipe Soares. Offline computer-aided diagnosis for glaucoma detection using fundus images targeted at mobile devices. *Computer Methods and Programs in Biomedicine*, 192:105341, 2020.

[33] W.Y. Ng, S. Zhang, Z. Wang, C.J.T. Ong, D.V. Gunasekeran, G.Y.S. Lim, F. Zheng, S.C.Y. Tan, G.S.W. Tan, T.H. Rim, L. Schmetterer, and D.S.W. Ting. Updates in deep learning research in ophthalmology. *Clinical Science*, 135:2357–2376, 2021.

[34] José Ignacio Orlando, Huazhu Fu, João Barbosa Breda, Karel van Keer, Deepti R. Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, Joonseok Lee, Xiaoxiao Li, Peng Liu, Shuai Lu, Balamurali Murugesan, Valery Naranjo, Sai Samarth R. Phaye, Sharath M. Shankaranarayana, Apoorva Sikka, Jaemin Son, Anton van den Hengel, Shujun Wang, Junyan Wu, Zifeng Wu, Guanghui Xu, Yongli Xu, Pengshuai Yin, Fei Li, Xiulan Zhang, Yanwu Xu, and Hrvoje Bogunović. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59:101570, 2020.

[35] Elena Prokofyeva, Alfred Wegener, and Eberhart Zrenner. Cataract prevalence and prevention in europe: a literature review. *Acta Ophthalmologica*, 91(5):395–405, 2013.

[36] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2016.

[37] M. Smaida, S. Yaroshchak, and Y. El Barg. Dcgan for enhancing eye diseases classification. volume 2864, pages 22–33, 2021. cited By 0.

[38] R. Torkzadehmahani, P. Kairouz, and B. Paten. Dp-cgan: Differentially private synthetic data and label generation. volume 2019-June, pages 98–104, 2019. cited By 24.

[39] Ihsan Ullah, Andre Rios, Vaibhav Gala, and Susan Mckeever. Explaining deep learning models for tabular data using layer-wise relevance propagation. *Applied Sciences*, 12(1), 2022.

[40] Ihsan Ullah, Andre Rios, Vaibhav Gala, and Susan McKeever. Explaining deep learning models for tabular data using layer-wise relevance propagation. *Applied Sciences (Switzerland)*, 12, 1 2022.

[41] José E. Valdez-Rodríguez, Edgardo M. Felipe-Riverón, and Hiram Calvo. Optic disc preprocessing for reliable glaucoma detection in small datasets. *Mathematics*, 9, 9 2021.

[42] Zhengwei Wang, Graham Healy, Alan F. Smeaton, and Tomás E. Ward. Use of neural signals to evaluate the quality of generative adversarial network performance in facial image generation. *Cognitive Computation*, 12(1):13 – 24, 2020. Cited by: 23; All Open Access, Green Open Access.

[43] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*, 2019.

[44] Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, and Tien Yin Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 3065–3068, 2010.