Universidad de Montemorelos

Facultad de Ingeniería y Tecnología

Ingeniería en Sistemas Computacionales

Glaucoma Detection in Latino Population through OCT's RNFL Thickness Map using MobileNet and Inception V3

Liza Grace Olivas Escamilla

1060310

Asesor: Dr. Germán Harvey Alférez Salinas

Asesor secundario: Dr. Javier Castillo Velázquez

Montemorelos, Nuevo León, México

22 de abril de 2020

Abstract

Glaucoma is the leading cause of irreversible blindness worldwide. It is estimated that over 60 million people around the world have this disease, with only part of them knowing they have it. Timely and early diagnosis is vital to delay/prevent patient blindness. Deep learning (DL) could be a tool for ophthalmologists to give a more informed and objective diagnosis. However, there is a lack of studies that apply DL for glaucoma detection to Latino population. Our contribution is to compare the effectiveness of MobileNet and Inception V3 models to detect cases of glaucoma in Latino patients. To this end, transfer learning was used to retrain previously trained models with images of the retinal nerve fibre layer Thickness Map of Mexican patients, obtained with Optical Coherence Tomography, from a clinic in the northern part of Mexico. Specifically, the Inception V3 model showed slight better average results than the MobileNet model in the case of classifying left eye images. In average, the evaluation results for right eye images were the same for both models. The evaluation results of the MobileNet model for the left eye are: accuracy: 86%, precision: 87%, recall: 87%, and F1 score: 87%. The evaluation results of the MobileNet model for the right eye are: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%. The evaluation results of the Inception V3 model for the left eye are: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%. The evaluation results of the Inception V3 model for the right eye are: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%.

Index Terms

Glaucoma, Latino Population, Retinal Nerve Fibre Layer, Thickness Map, Optical Coherence Tomography, Deep Learning, Convolutional Neural Networks, MobileNet, Inception V3, Transfer Learning.

CONTENTS

Glaucoma Detection in Latino Population through OCT's RNFL Thickness Map using MobileNet and Inception V3

Liza G. Olivas, Germán H. Alférez and Javier Castillo

School of Engineering and Technology, Montemorelos University, Mexico

I. Introduction

A. Background

Glaucoma is a serious long-term health care problem [1]. According to the Glaucoma Research Foundation [2], over 60 million people worldwide have glaucoma. The most important consequence that may occur as a result of this condition is visual impairment, which can lead to blindness [3]. Hence the ability to detect structural loss is fundamental in the diagnosis and management of glaucoma [4]. Specialists diagnose glaucoma based on Optical Coherence Tomography (OCT) or other studies, but even among them, the diagnosis could differ. Deep Learning (DL) has been used as a tool to detect glaucoma cases since a few years back. However, as far as we know, there are not research works that compare the effectiveness of DL topologies in detecting glaucoma in Latino population. This research work is a continuation of our previous work [5]. In [5], we trained a model using the Inception V3 algorithm with fundus images of glaucomatous and non-glaucomatous eyes of Latino patients, and obtained an accuracy of 99.7%.

B. Problem Statement

Glaucoma is one of the most common causes of global irreversible blindness. Diagnosis tends to be based on subjective information, and if wrongly diagnosed, the patient could lose sight. According to [6], by the year 2040 the number of people with glaucoma worldwide will increase to 111.8 million. Most of the studies and research works on glaucoma detection using DL are based and applied on Caucasian, Asian, or mixed population [7, 8, 9, 10]. However, there is a lack of studies on Latino population. Image recognition algorithms could be used as a breakthrough approach to help ophthalmologists to diagnose glaucoma in Latino patients.

C. Justification

Glaucoma is the leading cause of irreversible blindness worldwide. Blindness from glaucoma can often be prevented with early treatment. But because it may be asymptomatic until a relatively late stage, diagnosis is frequently delayed [11, 12, 13].

OCT is an imaging technology commonly used in the evaluation of damage caused by glaucoma. Retinal nerve fibre layer (RNFL) remains the dominant parameter for glaucoma diagnosis and detection of progression [4]. Nevertheless, current tests for glaucoma detection do not provide an automatic diagnosis. They return numeric and graphical information that the ophthalmologist uses for giving a diagnosis. However, these diagnoses remain subjective. According to [14], the increasing use of Artificial Intelligence (AI) in health could be used for diagnosing glaucoma objectively. It can cause a breakthrough and a major role in screening, diagnosis, and surveillance of glaucoma. Although there is previous work on automatic detection of glaucoma via DL [7, 8, 15], there is a need for comparing the effectiveness of different DL topologies for glaucoma detection. Moreover, most of the research works on glaucoma detection with DL use databases of eye images of Caucasian, Asian, or mixed population [7, 8, 9, 10]. However, there is a lack of studies in Latino population.

D. Objectives

Our main objective is to detect glaucoma in the Latino population of patients located at the northern part of Mexico by means of using images of the OCT's RNFL Thickness Map of both eyes to retrain and evaluate MobileNet and Inception V3 models. We achieved this objective with the following specific sub-objectives:

- Obtain images from both eyes from 165 Mexican patients. These patients were screened at the Instituto de la Visión, a private clinic located in Montemorelos, NL, Mexico.
- Use transfer learning to retrain two DL models with the obtained dataset by means of MobileNet and Inception V3 topologies. For training, we used 50 images for each class: glaucomatous and non-glaucomatous (for both right and left eyes). In total, 200 images were used for training. Retraining was carried out to reduce training time and computing resources. The MobileNet and Inception V3 initial models had been trained with ImageNet[1].
- Validate the MobileNet and Inception V3 models in terms of precision, recall, and F1 score.

[1] http://www.image-net.org/

Liza Olivas is a computer science engineering student at the School of Engineering and Technology of Montemorelos University, Nuevo León, Mexico, e-mail: 1060310@alumno.um.edu.mx.

Germán H. Alferez, Ph.D., is the director of the Institute of Data Science and professor at the School of Engineering and Technology of Montemorelos University, Nuevo León, Mexico, e-mail: harveyalferez@um.edu.mx.

Javier Castillo, M.D., is an ophthalmologist and the academic director of Instituto de la Visión, Montemorelos, Nuevo León, Mexico, e-mail: oftalmo@um.edu.mx.

- Test the effectiveness of the MobileNet and Inception V3 models with 15 images for each class: glaucomatous and non-glaucomatous (for both right and left eyes), with a total of 60 images.

E. Hypothesis

By using DL, it is possible to detect glaucoma in Latino population by automatically analyzing RNFL Thickness Map images given by OCT studies.

II. Theoretical Foundation

A. Underpinnings of our Approach

Our approach is based on the following concepts (see Fig. 1).

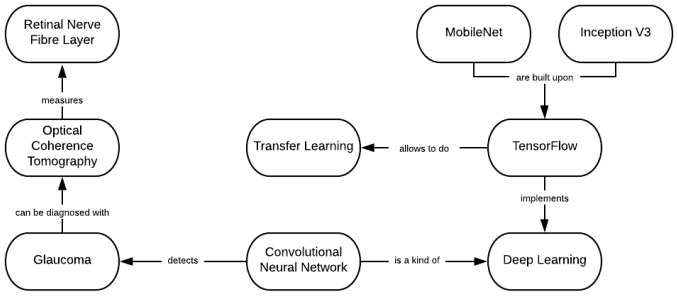

Fig. 1. Underpinnings of our approach

1) Glaucoma: Glaucoma is a term used to describe a group of disorders that have in common progressive degeneration of the optic nerve causing visual compromise and eventually blindness. Collectively, glaucoma is the leading cause of irreversible blindness worldwide [11]. The pathophysiology of glaucoma is uncertain with the different theories (mechanical and vascular) [16, 17] but the damage is the same, death of ganglion cells, with reduction in the RNFL that results in damage to the optic nerve head (ONH) with consequent visual limitation that can lead to blindness. As the balance between the production of aqueous humor in the eye and its drainage is lost, intraocular pressure (eye tone) may rise, which is the most important risk factor for damage to the optic nerve susceptible to glaucoma and may cause vision loss.

2) Retinal Nerve Fibre Layer: The RNFL consists of ganglion cell axons. Their course runs in parallel to the retinal surface; the fibers proceed to the optic disc, turn at a right angle, and exit the eye through the lamina cribrosa as the optic nerve. The fibers generally are unmyelinated within the retina. The RNFL is thickest at the margins of the optic disc, where all the fibers accumulate. The group of fibers that radiate to the disc from the macular area is called the papillomacular bundle. This important grouping of fibers carries the information that determines visual acuity [18]. The evaluation of the RNFL is one of the most important clinical examinations for diagnosing glaucoma because it becomes thinner when there is glaucoma [19, 20].

3) Optical Coherence Tomography: OCT is a noncontact, noninvasive imaging technology that uses light to create high-resolution, cross-sectional tomographic images of the retina and the ONH [21]. The OCT measures total retinal thickness, RNFL thickness, and ONH morphology [22]. The RNFL can be assessed more precisely with OCT, which directly measures the thickness of the RNFL [20].

4) Deep Learning: DL is a class of Machine Learning (ML) technique that exploits several layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification [23]. It teaches computers to do what comes naturally to humans: learn by example. In DL, a computer model learns to perform classification tasks directly from images, text, or sound. DL models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance [24].

5) Convolutional Neural Network: A Convolutional Neural Network (CNN), is a class of artificial neural networks that specializes in processing data that has a grid-like topology, such as an image. A CNN typically has three layers: a convolutional layer, pooling layer, and fully connected layer [25]:

1) The convolution layer is the core building block of the CNN. It carries the main portion of the network's computational load. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters otherwise known as a kernel, and the other matrix is the restricted portion of the receptive field.
2) The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs. This helps in reducing the spatial size of the representation, which decreases the required amount of computation and weights.
3) In the fully connected layer, the pooling operation is processed on every slice of the representation individually. Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular fully convolutional neural network. This is why it can be computed as usual by a matrix multiplication followed by a bias effect. The fully connected layer helps map the representation between the input and the output.

6) MobileNet: MobileNet is a family of mobile-first computer vision models for TensorFlow, based on depthwise separable convolutions. MobileNet algorithms are designed to effectively maximize accuracy while being mindful of the restricted resources in an on-device or embedded application. Specifically, MobileNets are small, low-latency, low-power models parameterized to meet the resource constraints of a variety of cases. They can be built upon for classification, detection, embeddings, and segmentation similar to how other popular large scale models, such as Inception, are used. All layers are followed by a batchnorm and ReLU nonlinearity with the exception of the final fully connected layer which has no nonlinearity and feeds into a softmax layer for classification [26, 27]. In [27], the creators of MobileNet present the MobileNet architecture in detail.

7) Inception V3: Inception V3 is an image recognition model that has obtained more than 78% accuracy on the ImageNet dataset. This model was developed based on [28]. The model is 42 layers deep and is made up of symmetric and asymmetric building blocks, including convolutions, average pooling, max pooling, concats, dropouts, and fully connected layers. Batchnorm is used extensively throughout the model and applied to activation inputs. Loss is computed via softmax [28, 29].

8) TensorFlow: TensorFlow[2] was developed in 2011 by the Google Brain Team. It is an open-source software library for numerical computation that uses data flow graphs. TensorFlow enables ML practitioners to do data-intensive computing with robust implementations of widely used DL algorithms. TensorFlow offers a very flexible architecture that enables to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API [30]. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that allows researchers and developers to push the state-of-the-art in ML, and easily build and deploy ML powered applications [31].

9) Transfer Learning: Transfer learning is a technique that reduces the time and computing resources used when training from scratch. Its main objective is to take advantage of data from a model that has already been trained on a related task and extract information that may be useful, and reuse it in a new model. Most often when doing transfer learning, the weights of the original model are not adjusted. Instead the final layer is removed and a new (often fairly shallow) model is trained on top of the output of the truncated model [32, 33].

B. Related Work

Tham et al. [6] studied the prevalence of glaucoma worldwide, and the projections of the disease for the years 2020 and 2040. They estimated that by the year 2020, the Latin American region will have the third largest number of persons affected by glaucoma, preceded by the Asian and African regions. In [34], the authors studied the prevalence of glaucoma in a population-based sample of Hispanic adults older than 40 years. The ocular examinations included visual acuity testing, applanation tonometry, gonioscopy, an optic disc evaluation, and a threshold visual field test; for open-angle glaucoma and angle-closure glaucoma. The study was realized in 72% of the eligible persons, with a prevalence of 1.97% of OAG (4,774 participants). They found that the prevalence of OAG in Hispanics was intermediate between reported values for Caucasians and African Americans and that the frequency of glaucoma increases more quickly with increasing age than in other ethnic groups.

Glaucoma can be detected through OCT using the RNFL thickness [14]. According to a study presented by [4], the RNFL thickness proved to be better than any other ONH parameter for discriminating between glaucomatous and non-glaucomatous eyes. In [35], the authors used the RNFL thickness deviation map for glaucoma detection. They evaluated its diagnostic performance and compared its sensitivity and specificity for glaucoma detection with circumpapillary RNFL measurement derived from the standard 3.46 mm diameter circle scan. They imaged one eye from each individual (102 normal subjects and 121 glaucoma patients) with Cirrus HD-OCT and Stratus OCT. By analyzing the RNFL thickness deviation map, the diagnostic sensitivity for glaucoma detection was improved, compared with conventional circumpapillary RNFL measurement. The reason behind it is that the RNFL thickness deviation map provides additional spatial and morphologic information of RNFL damage.

ML, in general, has proven to be a great tool for glaucoma detection [36, 37, 38, 39]. Among ML approaches, DL has been increasingly used in the last years for glaucoma detection as shown in Table 1. For instance, in our previous work [5], we used Inception V3 for glaucoma detection in Latino population. Specifically, 38 fundus images were used to retrain the model, 25 of confirmed glaucoma and 19 of non-glaucomatous eyes. After retraining, the model was tested in terms of accuracy with 6 fundus images, 3 healthy and 3 glaucomatous. The model showed a 99.7% accuracy. Unlike [5], which is focused only on one DL algorithm and on fundus images, this research work compares two different DL algorithms and uses OCT images.

In [40], the authors trained a CNN with a pretraining dataset (consisting of 4,316 OCT images), and then retrained it with an independent training dataset. The input features were 8x8 grid macular RNFL thickness and ganglion cell complex layer-thickness from SD-OCT. The area under the ROC curve (AUC) with the DL model was 93.7%.

Li et al. [9] developed a DL system for the classification of glaucomatous optic neuropathy for automated classification of color fundus images. They used a 22-layer deep convolutional neural network with 11 inception modules and tested its efficacy with 48,116 fundus images obtained by random sampling from the dataset LabelMe. This dataset was collected from almost exclusively Chinese population [41]. A total of 8,371 images were excluded from the training and validation dataset because they did not meet the selection criteria. Among the remaining 39,745 images, 8,000 were used for the validation set and 31,745 for the training set. The DL algorithm showed an AUC of 98.6%, Accuracy (ACC) of 92.9%, Sensitivity (SE) of 95.6%, and Specificity (SP) of 92.0% for classification of referable glaucomatous optic neuropathy.

In [10], the authors used a Deep Residual Learning algorithm, also known as a ResNet, for glaucoma detection in Japanese population. They trained a model with a training dataset of 1,364 eye images with OAG and 1,768 eye images of normative subjects. The authors used independent testing datasets to validate the diagnostic performance of the model. The AUC of the ResNet model was 96.5% with all images. Three residents in Ophthalmology also confirmed the presence of glaucoma in the training dataset. The AUC values obtained with the ResNet model tended to be significantly larger than those from the residents in ophthalmology.

In [7], the author used advanced DL algorithms to diagnose glaucoma based on retinal fundus images without using segmentated-based hand-crafted features. He developed the Glaucoma-Deep system, using a CNN unsupervised architec-

[2]https://www.tensorflow.org/

TABLE I
RESEARCH WORKS DONE ON DL FOR GLAUCOMA DETECTION (AREA UNDER THE ROC CURVE (AUC), SENSITIVITY (SE), SPECIFICITY (SP), PRECISION (PRC), ACCURACY (ACC))

Author	Year	Fundus/OCT	Train images	Test images	DL algorithm	AUC	SE	SP	PRC	ACC
Asaoka et al. [40]	2019	OCT	4,316	Independent	CNN (transfer learning)	93.7%	-	-	-	-
Espinoza et al. [5]	2018	Fundus	38	6	Inception V3	-	-	-	-	99.7%
Li et al. [9]	2018	Fundus	31,745	8,000	Deep CNN	98.6%	95.6%	92%	-	92.9%
Shibata et al. [10]	2018	Fundus	3,132	110	ResNet	96.5%	-	-	-	-
Abbas [7]	2017	Fundus	1,200		Deep Belief Network (DBN): Glaucoma-Deep System	-	84.50%	98.01%	84%	99%
Chen et al. [8]	2015	Fundus	99 (ORIGA), 650 (ORIGA)	551 (ORIGA), 1676 (SCES)	Deep CNN	83.1% (ORIGA) 88.7% (SCES)	-	-	-	-

ture to extract features from raw pixel intensities. The author also used a Deep Belief Network model to select the most discriminative deep features based on the training database, and a softmax linear classifier to differentiate between glaucoma and non-glaucoma. The Glaucoma-Deep system was tested on 1,200 retinal images, obtained from four datasets. The results were the following: SE of 84.50%, SP of 98.01%, ACC of 99%, and PRC of 84%.

Chen et al. [8] present a DL model based on deep CNN that is able to detect glaucoma in Asian population based on digital fundus images. The authors present a six-layer network, four convolutional and two fully connected layers. They trained their model with 99 images from the ORIGA dataset, tested it with the remaining 551 images of the same dataset, and obtained an AUC value of 83.1%. They did a second training for the SCES dataset, using the whole 650 images from the ORIGA dataset for training, and tested this model with the 1,676 images from the SCES dataset. They obtained an AUC value of 88.7%.

DL has been used to detect glaucoma obtaining good results. However, there is a lack of research work focused on Latino population. Moreover, although there is one research work [15] done on comparing different DL topologies for glaucoma detection, it is not based on Latino population.

III. RESULTS

A. Methodology

The IBM Foundational Methodology for Data Science is focused on data science projects, hence the decision to follow this methodology on this research work. This methodology consists of ten stages that form an iterative process for using data to uncover insights [42]. These stages are described in the context of this research work.

1) Problem Understanding: The problem was understood as stated in Section I.B.

2) Analytic Approach: Two DL image recognition topologies, MobileNet and Inception V3, were chosen as the analytical approaches to classify the OCT images in two classes: glaucomatous and non-glaucomatous.

Inception V3 was chosen because of its high accuracy and relatively small size compared to other deep neural network models [43]. MobileNet was chosen based on the evaluation results presented in [27], where its creators present and compare MobileNet to other topologies, such as VGG16 and GoogleNet. MobileNet showed a similar accuracy to other state-of-the-art topologies while being considerably smaller in size and less compute intensive. Moreover, transfer learning can be applied to MobileNet and Inception V3, which can reduce the time and computing resources used when training from scratch.

3) Data Requirements: The required data consist of images of the RNFL Thickness Map that is part of the OCT scan result of glaucomatous eyes and non-glaucomatous eyes. Images must be in RGB, with an input size of 224 x 224 pixels for MobileNet and 299 x 299 pixels for Inception V3.

4) Data Collection: The OCT scan files were collected from a Zeiss OCT machine at the Instituto de la Visión in PDF format, and classified and organized in folders by an expert into two classes: 1) glaucomatous, and 2) non-glaucomatous. This classification was carried out for both eyes. With these images, we got a dataset of 333 files in total.

5) Data Understanding: At this point, the dataset was conformed by PDF files as shown in Figure 2. Each one of these files contains the following structure: the top part shows the information of the patient and the exam. Then, there are five images corresponding to information about the right eye, five images corresponding to information about the left eye, and in between, a table and four charts comparing them. Specifically, for both eyes, Section A in Figure 2 shows the RNFL Thickness Map; Section B shows the RNFL Deviation Map; Section C and D show the Extracted Horizontal Tomogram and the Extracted Vertical Tomogram, respectively; and Section E shows the RNFL Circular Tomogram. The table in Section F displays measures corresponding to both eyes. Section G shows the Neuro-Retinal Rim Thickness, Section H shows the RNFL Thickness, Section I shows the RNFL Quadrants, and Section J shows the RNFL Clock Hours.

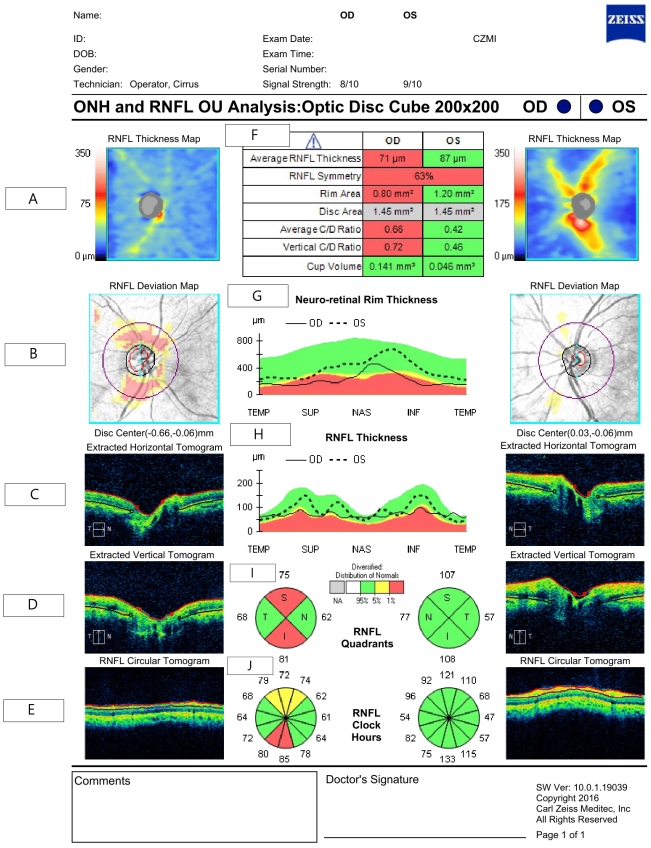

Fig. 2. Sample of an OCT scan file

6) Data Preparation: Since this research work is focused on RNFL Thickness Map images, the PDF files from the previous step were cropped to obtain only the RNFL Thickness Map of the corresponding eye out of the files (see section A in Figure 2). The cropped images were saved as JPG files. This action was carried out for images in both classes, glaucomatous and non-glaucomatous, until completing the final dataset. For instance, Figure 3 shows the images of non-glaucomatous left and right eyes, and Figure 4 shows the images of glaucomatous left and right eyes.

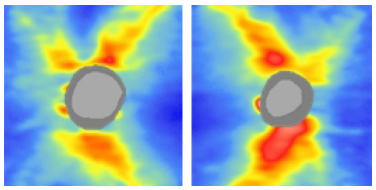

Fig. 3. Non-glaucomatous left and right eyes, respectively

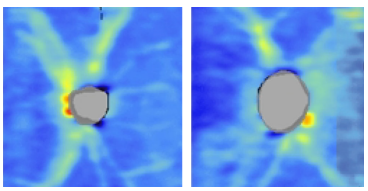

Fig. 4. Glaucomatous left and right eyes, respectively

Originally, 333 images were collected for left and right eyes. However, some of them were damaged. An image was considered as damaged when it had black spots in which data was missing. Therefore, to train the networks with the MobileNet and Inception V3 algorithms in the best way possible, the damaged parts of the image were cut out when it was possible, as shown in Figure 5. In the cases where the black spots occupied a large area of the image, as in Figure 6, the image was cut off from the training and testing datasets.

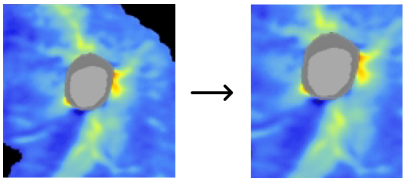

Fig. 5. Cut out image

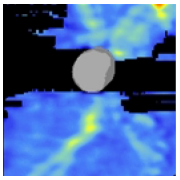

Fig. 6. Cut off image from the dataset

After cutting out and cutting off images, there were different amounts of images per class in both eyes. Specifically, there were 52 images of glaucomatous left eyes, 50 images of non-glaucomatous left eyes, 51 images of glaucomatous right eyes, and 57 images of non-glaucomatous right eyes. In order to have the same number of images per class, images were removed from the classes with more images to balance the number of images per class. Therefore, in the training stage for each eye, 50 images were used per class. The classification was made in the prediction stage, using 15 images per class, different than the ones used in the training stage, to avoid overfitting (see Table 2 and Table 3). In total, 260 images were used in the experiments.

TABLE II
IMAGES USED IN THE EXPERIMENTS (LEFT EYE)

Class / Stage	Glaucomatous	Non-glaucomatous
Training	50	50
Prediction	15	15
Total number per class	65	65
Total number of images	**130**	

TABLE III
IMAGES USED IN THE EXPERIMENTS (RIGHT EYE)

Class / Stage	Glaucomatous	Non-glaucomatous
Training	50	50
Prediction	15	15
Total number per class	65	65
Total number of images	**130**	

7) Modeling: Four models were generated, two trained with MobileNet, one for the left eye and one for the right eye, and another two trained with Inception V3. For instance, the command to retrain the model for left-eye images with MobileNet is presented in Listing 1.

```
1   python -m scripts.retrain \
2   --bottleneck_dir=mobile_v1_100_224/OS/tf_files/bottle
3   necks \
4   --how_many_training_steps=500 \
5   --model_dir=mobile_v1_100_224/tf_files/models/ \
6   --summaries_dir=mobile_v1_100_224/OS/tf_files/train
7   ing_summaries \
8   --output_graph=mobile_v1_100_224/OS/tf_files/retrain
9   ed_graph.pb \
10  --output_labels=mobile_v1_100_224/OS/tf_files/retrain
11  ed_labels.txt \
12  --architecture="mobilenet_1.0_224" \
13  --image_dir=IDV/OS/Train
```

Listing 1. Command to retrain the MobileNet algorithm with left-eye images

In the first line, the retrain script from the TensorFlow Hub repository[3] is executed. This script retrains the top layer in the model, which is capable of recognizing specific classes of images. Lines 2 and 3 create the directory where the bottleneck files are going to be stored. A bottleneck is the layer just before the final output layer that actually does the classification. Every image is reused multiple times during the training stage. Therefore, calculation operations in the layers behind the bottleneck for each image take a significant amount of time. Since these lower layers of the network are not modified, their outputs can be cached and reused. These outputs are the "bottleneck files" that are stored. In case of rerunning the script, these files are reused.

In line 4 the number of training steps is declared. Line 5 indicates the directory where the model is stored. In lines 6 and 7 the training summaries directory is created. This directory stores the training progress reports that Tensorboard is monitoring. Lines 8-11 indicate the directories where the output labels and graphs are going to be sent to. Line 12 indicates the architecture to be used. In the case of MobileNet, the "mobilenet_1.0_224" parameter is used. In the case of Inception V3, the parameter "inception_v3" is used . Finally, line 13 indicates the directory of the training dataset.

In lines 2, 5, 6, 8, and 10, the directory "mobile_v1_100_224" changes to "inceptionv3" when using Inception V3. In lines 2, 6, 8, 10, and 13, the directory "OS" changes to "OD" when training with right-eye images.

B. Results

This section covers the evaluation stage of the methodology followed in this research work. Specifically, it shows the results obtained with the MobileNet and Inception V3 models.

The MobileNet and Inception V3 models were retrained using Python v3.6.9, TensorFlow v1.14.0, and Anaconda v2019.07. A personal laptop with an Intel(R) Core(TM) i5-6200U CPU and 12 GB RAM was used for the training and evaluation of the models. Both models were retrained using the retrain script with the training images dataset. For the testing stage,15 different images for each eye (15 images for the right eye and 15 for the left eye) were used per class. Table 4 and 5 show the confusion matrices of both eyes with MobileNet. Table 6 and 7 show the confusion matrices of both eyes with Inception V3, both having equal results.

[3]https://github.com/tensorflow/hub/blob/master/examples/image_retraining/retrain.py

TABLE IV
BINARY-CLASS CONFUSION MATRIX OF MOBILENET (LEFT EYE)

Actual \ Prediction	Glaucomatous	Non-glaucomatous
Glaucomatous	12	3
Non-glaucomatous	1	14

TABLE V
BINARY-CLASS CONFUSION MATRIX OF MOBILENET (RIGHT EYE)

Actual \ Prediction	Glaucomatous	Non-glaucomatous
Glaucomatous	14	1
Non-glaucomatous	2	13

TABLE VI
BINARY-CLASS CONFUSION MATRIX OF INCEPTION V3 (LEFT EYE)

Actual \ Prediction	Glaucomatous	Non-glaucomatous
Glaucomatous	13	2
Non-glaucomatous	1	14

TABLE VII
BINARY-CLASS CONFUSION MATRIX OF INCEPTION V3 (RIGHT EYE)

Actual \ Prediction	Glaucomatous	Non-glaucomatous
Glaucomatous	13	2
Non-glaucomatous	1	14

Tables 8-11 show the evaluation results of the classification model in terms of precision, recall, and F1 score.

TABLE VIII
RESULTS OBTAINED FROM THE EVALUATION OF MOBILENET (LEFT EYE)

Class	Precision	Recall	F1 Score
Glaucomatous	0.82	0.93	0.87
Non-glaucomatous	0.92	0.80	0.86
Average	**0.87**	**0.87**	**0.87**

TABLE IX
RESULTS OBTAINED FROM THE EVALUATION OF MOBILENET (RIGHT EYE)

Class	Precision	Recall	F1 Score
Glaucomatous	0.93	0.87	0.90
Non-glaucomatous	0.88	0.93	0.90
Average	**0.90**	**0.90**	**0.90**

TABLE X
RESULTS OBTAINED FROM THE EVALUATION OF INCEPTION V3 (LEFT EYE)

Class	Precision	Recall	F1 Score
Glaucomatous	0.88	0.93	0.90
Non-glaucomatous	0.93	0.87	0.90
Average	**0.90**	**0.90**	**0.90**

TABLE XI
RESULTS OBTAINED FROM THE EVALUATION OF INCEPTION V3 (RIGHT EYE)

Class	Precision	Recall	F1 Score
Glaucomatous	0.88	0.93	0.90
Non-glaucomatous	0.93	0.87	0.90
Average	**0.90**	**0.90**	**0.90**

C. Discussion

In the results in Tables 8-11, the MobileNet and Inception V3 algorithms show promising results for the two classes with images of both eyes. The Inception V3 model showed slight better average results than the MobileNet model in the case of classifying left eye images. In average, the evaluation results for right eye images were the same for both models.

IV. CONCLUSIONS AND FUTURE WORK

In this research work, the MobileNet and Inception V3 algorithms were used to create two classification models for glaucoma detection of Latino population. Specifically, the MobileNet and Inception V3 models were retrained for both eyes, with a total of 200 images. For the testing stage, 60 images were used, and the average results were satisfactory in both models. Specifically, the Inception V3 model showed slight better average results than the MobileNet model in the case of classifying left eye images. In average, the evaluation results for right eye images were the same for both models. The evaluation results of the MobileNet model for the left eye were: accuracy: 86%, precision: 87%, recall: 87%, and F1 score: 87%. The evaluation results of the MobileNet model for the right eye were: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%. The evaluation results of the Inception V3 model for the left eye were: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%. The evaluation results of the Inception V3 model for the right eye were: accuracy: 90%, precision: 90%, recall: 90%, and F1 score: 90%. As future work, we expect to extend this research work with a larger dataset of glaucoma and non glaucoma OCT images of Latino patients, obtained from different eye care centers in Mexico. This model will be deployed in a clinical environment. Also, we expect to improve it based on the feedback from ophthalmologists who use it to make classifications.

REFERENCES

[1] M. S. Berlin, "Method for treating glaucoma," English, pat. US9820883B2, Nov. 2017. [Online]. Available: https://patentimages.storage.googleapis.com/c6/78/39/2710a8cad38909/US9820883.pdfx.

[2] Glaucoma Research Foundation, *January is glaucoma awareness month*, Jan. 2020. [Online]. Available: https://www.glaucoma.org/news/glaucoma-awareness-month.php.

[3] M. I. Azcona-Cruz, M. Ríos-Lobo, and S. Amador-Jiménez, "Glaucoma: Aspectos relevantes para la detección oportuna," *Salud y Administración*, vol. 2, no. 4, pp. 23–35, Jan. 2015, ISSN: 2448-6159. [Online]. Available: http://www.unsis.edu.mx/revista/doc/vol2num4/A3_Glaucoma.pdf.

[4] I. I. Bussel, G. Wollstein, and J. S. Schuman, "OCT for glaucoma diagnosis, screening and detection of glaucoma progression," *British Journal of Ophthalmology*, vol. 98, no. Suppl II, pp. ii15–ii19, Dec. 2014. DOI: 10.1136/bjophthalmol-2013-304326.

[5] M. A. Espinoza, G. H. Alferez, and J. Castillo, "Predicción de Glaucoma Mediante Redes Neuronales Convolucionales," Proceedings of the 2018 International Conference on Health Informatics and Medical Systems (HIMS 2018), Las Vegas, NV, USA, 2018.

[6] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, Nov. 2014, ISSN: 0161-6420/14. DOI: 10.1016/j.ophtha.2014.05.013. [Online]. Available: https://www.aaojournal.org/article/S0161-6420(14)00433-3/fulltext.

[7] Q. Abbas, "Glaucoma-Deep: Detection of glaucoma eye disease on retinal fundus images using deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 41–45, 2017. DOI: 10.14569/ijacsa.2017.080606. [Online]. Available: https://thesai.org/Downloads/Volume8No6/Paper_6-Glaucoma_Deep_Detection_of_Glaucoma_Eye_Disease.pdf.

[8] X. Chen, Y. Xu, D. W. Kee Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 715–718. DOI: 10.1109/EMBC.2015.7318462. [Online]. Available: https://ieeexplore.ieee.org/document/7318462.

[9] Z. Li, Y. He, S. Keel, W. Meng, R. T. Chang, and M. He, "Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs," *Ophthalmology*, vol. 125, no. 8, pp. 1199–1206, Aug. 2018. DOI: 10.1016/j.ophtha.2018.01.023.

[10] N. Shibata, M. Tanito, K. Mitsuhashi, Y. Fujino, M. Matsuura, H. Murata, and R. Asaoka, "Development of a deep residual learning algorithm to screen for glaucoma from fundus photography," *Scientific Reports*, vol. 8, no. 14665, Oct. 2018. DOI: 10.1038/s41598-018-33013-w. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6168579/.

[11] J. L. Wiggs and L. R. Pasquale, "Genetics of glaucoma," *Human Molecular Genetics*, vol. 26, no. R1, R21–R27, May 2017. DOI: 10.1093/hmg/ddx184. [Online].

Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6074793/.

[12] K. Boyd, *What is Glaucoma?* Accessed: 08/28/19, American Academy of Ophthalmology, Aug. 2019. [Online]. Available: https://www.aao.org/eye-health/diseases/what-is-glaucoma.

[13] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma," *JAMA*, vol. 311, no. 18, p. 1901, 2014. DOI: 10.1001/jama.2014.3192.

[14] D. S. Ting, L. Peng, A. V. Varadarajan, P. A. Keane, P. M. Burlina, M. F. Chiang, L. Schmetterer, L. R. Pasquale, N. M. Bressler, D. R. Webster, M. Abramoff, and T. Y. Wong, "Deep learning in ophthalmology: The technical and clinical considerations," *Progress in Retinal and Eye Research*, 2019. DOI: 10.1016/j.preteyeres.2019.04.003.

[15] K. R. Patel, G. Lee, M. Durbin, M. Wall, P. Artes, and J. Flanagan, *Deep neural network based glaucoma detection using rnfl thickness map*, Poster 1470 - A0154, https://medicine.uiowa.edu/eye/content/arvo-2019, Apr. 2019. [Online]. Available: https://www.zeiss.com/content/dam/Meditec/ref_master/events/arvo/posters/patel_arvo_glaucoma_detection_using_rnfl_thickness_map_v3.pdf.

[16] American Academy of Ophthalmology, *Theories of Glaucomatous Optic Nerve Damage*. [Online]. Available: https://www.aao.org/bcscsnippetdetail.aspx?id=f19571e0-b8d7-4679-a2c5-fd30a9b016c1.

[17] S. S. Ahmad, "Controversies in the vascular theory of glaucomatous optic nerve degeneration," *Taiwan Journal of Ophthalmology*, vol. 6, no. 4, pp. 182–186, Dec. 2016. DOI: 10.1016/j.tjo.2016.05.009.

[18] L. A. Remington, "Chapter 4 - retina," in *Clinical Anatomy and Physiology of the Visual System (Third Edition)*, L. A. Remington, Ed., Third Edition, Saint Louis: Butterworth-Heinemann, 2012, pp. 61 –92, ISBN: 978-1-4377-1926-0. DOI: https://doi.org/10.1016/B978-1-4377-1926-0.10004-9. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781437719260100049.

[19] J. W. Shin, M. Seong, J. W. Lee, E. H. Hong, and K. B. Uhm, "Diagnostic ability of retinal nerve fiber layer thickness deviation map for localized and diffuse retinal nerve fiber layer defects," *Journal of Ophthalmology*, vol. 2017, pp. 1–9, 2017. DOI: 10.1155/2017/8365090. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5259680/#B1.

[20] T. Shaarawy, M. Sherwood, R. Hitchings, and J. Crowston, *Glaucoma*. Elsevier, 2015, ISBN: 9780702051937. DOI: 10.1016/c2011-1-04562-9. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/retinal-nerve-fiber-layer.

[21] Z. Burgansky-Eliash, G. Wollstein, T. Chu, J. D. Ramsey, C. Glymour, R. J. Noecker, H. Ishikawa, and J. S. Schuman, "Optical coherence tomography machine learning classifiers for glaucoma detection: A preliminary study," *Investigative Opthalmology & Visual Science*, vol. 46, no. 11, p. 4147, 2005. DOI: 10.1167/iovs.05-0366.

[22] M. N. Menke, G. T. Feke, and C. L. Trempe, "OCT measurements in patients with optic disc edema," *Investigative Opthalmology & Visual Science*, vol. 46, no. 10, p. 3807, 2005. DOI: 10.1167/iovs.05-0352.

[23] L. Deng, "Deep learning: Methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014. DOI: 10.1561/2000000039.

[24] MathWorks, *What is Deep Learning? 3 Things You Need To Know*, English, MathWorks. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html.

[25] M. Mishra, *Convolutional Neural Networks, Explained*, English, Oracle, Mar. 2019. [Online]. Available: https://www.datascience.com/blog/convolutional-neural-network.

[26] A. G. Howard, *Mobilenets: Open-source models for efficient on-device vision*, Jun. 2017. [Online]. Available: https://ai.googleblog.com/2017/06/mobilenets-open-source-models-for.html.

[27] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 17, 2017. arXiv: http://arxiv.org/abs/1704.04861v1 [cs.CV].

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Dec. 2, 2015. arXiv: https://arxiv.org/abs/1512.00567v3 [cs.CV].

[29] Google Cloud, *Advanced guide to inception v3 on cloud tpu*, English, Google, Jan. 2019. [Online]. Available: https://cloud.google.com/tpu/docs/inception-v3-advanced.

[30] G. Zaccone, M. R. Karim, and A. Menshawy, *Deep Learning with TensorFlow: Explore neural networks with Python*. Packt Publishing, 2017, ISBN: 978-1-78646-978-6.

[31] TensorFlow, *Why TensorFlow*. [Online]. Available: https://www.tensorflow.org/.

[32] ——, *What is transfer learning?* [Online]. Available: https://www.tensorflow.org/js/tutorials/transfer/what_is_transfer_learning.

[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org.

[34] H. A. Quigley, "The prevalence of glaucoma in a population-based study of hispanic subjects," *Archives of Ophthalmology*, vol. 119, no. 12, p. 1819, 2001. DOI: 10.1001/archopht.119.12.1819.

[35] C. K. Leung, S. Lam, R. N. Weinreb, S. Liu, C. Ye, L. Liu, J. He, G. W. Lai, T. Li, and D. S. Lam, "Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography," *Ophthalmology*, vol. 117, no. 9, pp. 1684–1691, 2010. DOI: 10.1016/j.ophtha.2010.01.026.

[36] P. Murtagh, "Current applications of machine learning in the screening and diagnosis of glaucoma: A system-

atic review and meta-analysis," *International Journal of Ophthalmology*, vol. 13, no. 1, pp. 149–162, 2020. DOI: 10.18240/ijo.2020.01.22.

[37] L. S. Shigueoka, J. P. C. de Vasconcellos, R. B. Schimiti, A. S. C. Reis, G. O. de Oliveira, E. S. Gomi, J. A. R. Vianna, R. D. dos Reis Lisboa, F. A. Medeiros, and V. P. Costa, "Automated algorithms combining structure and function outperform general ophthalmologists in diagnosing glaucoma," *PLOS ONE*, vol. 13, no. 12, B. Mortazavi, Ed., e0207784, 2018. DOI: 10.1371/journal.pone.0207784.

[38] M. Christopher, A. Belghith, R. N. Weinreb, C. Bowd, M. H. Goldbaum, L. J. Saunders, F. A. Medeiros, and L. M. Zangwill, "Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression," *Investigative Opthalmology & Visual Science*, vol. 59, no. 7, p. 2748, 2018. DOI: 10.1167/iovs.17-23387.

[39] S. J. Kim, K. J. Cho, and S. Oh, "Development of machine learning models for diagnosis of glaucoma," *PLOS ONE*, vol. 12, no. 5, B. Liu, Ed., e0177726, May 2017. DOI: 10.1371/journal.pone.0177726. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5441603/.

[40] R. Asaoka, H. Murata, K. Hirasawa, Y. Fujino, M. Matsuura, A. Miki, T. Kanamoto, Y. Ikeda, K. Mori, A. Iwase, N. Shoji, K. Inoue, J. Yamagami, and M. Araie, "Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images," *American Journal of Ophthalmology*, vol. 198, pp. 136–145, 2019. DOI: 10.1016/j.ajo.2018.10.007.

[41] M. Christopher, A. Belghith, C. Bowd, J. A. Proudfoot, M. H. Goldbaum, R. N. Weinreb, C. A. Girkin, J. M. Liebmann, and L. M. Zangwill, "Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs," *Scientific Reports*, vol. 8, no. 1, 2018. DOI: 10.1038/s41598-018-35044-9.

[42] J. B. Rollins, *Metodología Fundamental para la Ciencia de Datos*, Jun. 2015. [Online]. Available: https://www.ibm.com/downloads/cas/WKK9DX51.

[43] A. Canziani, A. Paszke, and E. Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications," May 24, 2016. arXiv: http://arxiv.org/abs/1605.07678v4 [cs.CV].