

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier

Dual Machine-Learning system to aid Glaucoma Diagnosis using disc and cup feature extraction.

JAVIER CIVIT-MASOT¹, MANUEL J. DOMÍNGUEZ-MORALES¹, SATURNINO VICENTE¹ AND ANTON CIVIT.¹

¹Robotics and Technology of Computers Lab (www.rtc.us.es), University of Seville (Spain). School of Computer Engineering, Avenida Reina Mercedes, s/n, 41012, Seville, Spain

Corresponding author: Javier Civit-Masot (e-mail: jcivit@atc.us.es).

ABSTRACT Glaucoma is a degenerative disease that affects vision, causing damage to the optic nerve that ends in vision loss. The classic techniques to detect it have undergone a great change since the intrusion of machine learning techniques into the processing of eye fundus images. Several works focus on training a convolutional neural network (CNN) by brute force, while others use segmentation and feature extraction techniques to detect glaucoma. In this work, a diagnostic aid tool to detect glaucoma using eye fundus images is developed, trained and tested. It consists of two subsystems that are independently trained and tested, combining their results to improve glaucoma detection. The first subsystem applies machine learning and segmentation techniques to detect optic disc and cup independently, combine them and extract their physical and positional features. The second one applies transfer learning techniques to a pre-trained CNN to detect glaucoma through the analysis of the complete eye fundus images. The results of both systems are combined to discriminate positive cases of glaucoma and improve final detection. The results show that this system achieves a higher classification rate than previous works. The system also provides information on the basis for the proposed diagnosis suggestion that can help the ophthalmologist to accept or modify it.

INDEX TERMS Glaucoma, Ensemble networks, Medical diagnostic aids, medical imaging, explainable AI.

I. INTRODUCTION

THE term **Glaucoma** is used for a group of progressive neuropathies that affects vision (mostly bilateral) and is characterized by loss of retinal ganglion cells and damage to the optic nerve head, causing loss of the visual field and, finally, blindness [1]. Moreover, it is one of the main causes of irreversible visual damage and blindness worldwide (second leading cause in Europe).

There are two main types of glaucoma whose causes are well known:

- Open-angle glaucoma (OAG): the most common form of glaucoma (at least 90% of all glaucoma cases). It is caused by the slow clogging of the drainage canals, resulting in increased eye pressure. “Open-angle” means that the angle where the iris meets the cornea is as wide and open as it should be (also called primary or chronic glaucoma).
- Angle-closure glaucoma (ACG): It is caused by blocked

drainage canals, resulting in a sudden rise in intraocular pressure. It is also called acute glaucoma or narrow-angle glaucoma and, unlike OAG, ACG is a result of the angle between the iris and the cornea closing.

There are other glaucoma types like normal-tension glaucoma (the optic nerve is damaged even though the eye pressure is not very high, so it is not well known why this damage is produced), congenital glaucoma (occurs in babies when there is incorrect or incomplete development of the eye’s drainage canals during the prenatal period), and several variants of OAG and ACG. Most of them are not well studied and their causes are not defined yet.

Both main types (OAG and ACG) are usually evaluated separately, obtaining a mean prevalence worldwide of 1.96% for OAG and 0.69% for ACG according to [2]. In any case, the progression from diagnosis to at least unilateral blindness is above 1% per year [3]. Figure 1 shows the difference between a healthy eye and an eye with glaucoma in fundus

images. The images come from the DRISHTI dataset.

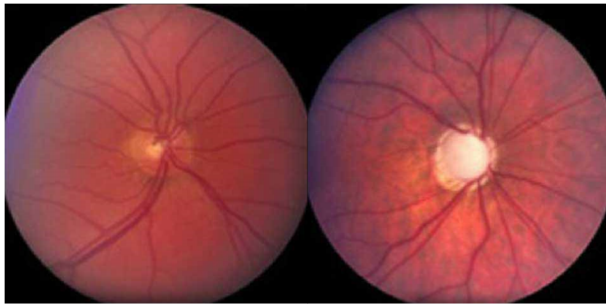


FIGURE 1: healthy eye (right) versus eye with glaucoma (left).

As the population continues to age, the number of glaucoma patients worldwide is expected to reach 111.8 million in 2040 [4]. In addition, the global disease burden of blindness and visual impairment due to glaucoma has been shown to be significantly associated with a decrease in quality of life, physical functioning and mental health [5]–[7]. Figure 2 shows how glaucoma affects the patient vision and its clear disabling effects.



FIGURE 2: Patient vision with and without glaucoma. Images from National Eye Institute.

In order to determine if a patient has glaucoma or not, many advances have appeared since the initial clinical tests. The classic glaucoma detection mechanisms are based on the verification of five key factors: tonometry, ophthalmoscopy, perimetry, gonioscopy and pachymetry. Depending on the results of each exam independently, the doctor may continue performing the others or not.

- Tonometry [8]: measures the pressure within the eye. The range for normal pressure is between 12 and 22 mmHg. Most glaucoma cases are diagnosed with pressure exceeding 20 mmHg; however, some people can have glaucoma at pressures between 12–22 mm Hg, thus, other exams may be required.
- Ophthalmoscopy [9]: helps to examine the shape and color of the optic nerve for glaucoma damage. If the intraocular pressure (IOP) is not within the normal range or if the optic nerve looks unusual, more exams are needed.
- Perimetry [10]: visual field test that produces a map of the complete field of vision. This test will help to

determine whether your vision has been affected by glaucoma.

- Gonioscopy [11]: helps to determine whether the angle where the iris meets the cornea is open and wide (a possible sign of open-angle, chronic glaucoma) or narrow and closed (a possible sign of angle-closure or acute glaucoma).
- Pachymetry [12]: measures the thickness of the cornea. Corneal thickness has the potential to influence eye pressure readings; so, with this measurement, it is easier to understand the IOP readings.

As detailed above, there is not a unique procedure to diagnose glaucoma because of the different eye characteristics of each patient. Moreover, those exams have to be interpreted by the doctor before making a diagnosis.

Currently, glaucoma diagnosis techniques based on medical image analysis are gaining popularity over more classic tests. In these cases several features of the retinal structure need to be observed: the optic nerve head (ONH), cup, peripapillary atrophy, retinal nerve fiber layer, etc. In a fundus image, the ONH is a bright and rounded area, and there is a smaller, rounded area inside the ONH called a cup. Peripapillary atrophy appears as a crescent which coincides with the area outside the ONH. The retinal nerve fiber layer is also located outside the ONH, which has white striated textures. In general, to identify the features needed to diagnose a glaucoma using computer vision, two types of techniques can be applied: segmentation and feature extraction.

Several of the most popular methods involve processes of localization and segmentation, such as thresholding [13] and active contours [14]. Additional methods have been developed, such as fuzzy c-means [15]. These methods often misclassify the ONH area; so, to overcome this inconvenient, morphological operations usually need to be applied.

On the other hand, several methods for ONH texture feature extraction have been developed as this is the main feature used to detect glaucoma. Wavelet and higher-order spectra (HOS) methods are the most popular methods for feature extraction. Some works use the discrete wavelet transform (DWT) and HOS applied with a support vector machine (SVM) classifier like [16]. Others like [17] use Principal Component Analysis (PCA) to reduce the features obtained using DTW.

The works mentioned above were only some examples of the evolution in glaucoma detection techniques. However, the study of medical images in general has experienced a great progress with the inclusion of Machine Learning systems capable of automatically extracting the necessary characteristics to make a correct diagnosis [18].

These systems require a dataset made up of several images corresponding to glaucoma patients and healthy patients (all of them previously labeled by a professional). Using this knowledge, neural network-based systems are able to automatically analyze those images and extract the characteristics necessary to help diagnosing glaucoma.

These systems require other steps: a preprocessing stage, the correct choice of the network architecture, a training stage (that sometimes requires supervision), among others.

Despite all this, the results of using neural networks with medical images in several works (not only for glaucoma diagnosis) are better than the ones obtained by classical diagnostic systems [19], [20]. Thus, based on these premises, this work consists of using Machine Learning techniques applied to medical images of the fundus to obtain an aid system for glaucoma diagnosis.

In a previous work, this research group started combining multiple publicly available datasets (RIM-One V3, DRISHTI and DRIONS) and, using segmentation techniques in a cloud-based service, results demonstrated the feasibility of applying these mechanisms to glaucoma detection [21].

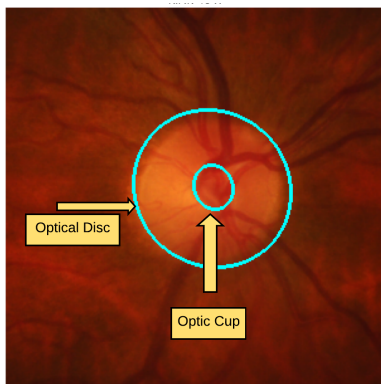


FIGURE 3: Eye fundus image: optical disc and optical cup are distinguished.

Although the results obtained were promising, the system was trained using disc and cup segmentation independently (see Figure 3). So, in order to obtain better results and trying to mainly reduce the false negative classification, a more complex system is needed.

Some papers have used several deep learning networks in parallel to improve the results that would be obtained by using a single network implementation. As an example [22] combines the results of five CNNs to improve the results on the LSVRC-2010 ImageNet training set. This technique has also been applied to glaucoma identification obtaining interesting results [23]. However, all these approaches use a set of CNNs to obtain the same type of results (e.g. the patient has or does not have Glaucoma) and then obtain a combined result by some sort of final voting.

A completely different approach is based on the segmentation of the optic disc and cup. There are several methods that can help predict glaucoma from the segmented disc and cup data in fundus images. First, the ratio between the diameters of the optic disc and cup, known as cup to disc ratio (CDR), is a very useful predictor for Glaucoma. Additionally the order of the widths of the different borders (inferior, superior, temporal and nasal- ISTN) can be used too. Several works have implemented deep learning approaches to segment optic

disc and cup in order to be able to estimate the CDR or use the ISTN approach. An important problem of these approaches is that, in a few cases, they produce segmentation results with shapes that are not compatible with the ophthalmological knowledge that requires these shapes to be similar to ellipsoids.

In this paper we use an ensemble approach to glaucoma prediction but, instead of using several convolutional networks to directly predict glaucoma, we use the following approach:

- We segment cup and disc using a generalized U-Net. [24] to calculate the CDR as a glaucoma predictor.
- we use RANSAC [25] to find out if the predicted shapes are similar enough to ellipses.
- We use transfer learning on a MobileNet V2, pretrained with weight from the imageNet 1K challenge to directly predict glaucoma.
- We combine all our result to provide the ophthalmologist with a Glaucoma likelihood score.

So, based on these previous approaches and the works done by this research group, this work combines a dual convolutional neural network (CNN) to classify discs and cups (see Figure 3) using data augmentation and feature extraction (extracting physical and positional features), with a classification system based on a pre-trained CNN with transfer learning techniques.

This feature extraction technique is combined with the eye fundus classification CNN for glaucoma detection in a novel work that obtains a diagnosis aid system with results better than previous works.

The rest of the paper is divided in the following way: first, the dataset and the system's architecture are described in the Materials and Methods section, including the different implemented stages. Next, the results obtained after the training process in Keras and their evaluation are detailed and explained in the Results and Discussion section. Finally, conclusions are presented.

II. MATERIALS AND METHODS

This section presents the dataset used for training the Machine-Learning system in this work, as well as the global architecture of the system implemented to diagnose glaucoma based on the properties of the disc and the cup.

A. DATASET

The database used in this work combines two publicly available datasets: RIM-One V3 and DRISHTI. This is the one used in a previous work [21], and it is important to continue with this combination in order to compare the results obtained in this work with the ones obtained before.

Both datasets provide labels indicating if the images correspond to a patient with glaucoma or not. The labeling process includes the supervised evaluation of each of the dataset samples by a professional in the field. Thus, this professional certifies that each of the images from the datasets corresponds

to a patient with glaucoma or a healthy patient. Works that perform cup and disc segmentation also need the ophthalmologists to manually perform this segmentation and, thus, provided also the labeled images indicating the ground truth for the disc and cup areas.

The DRIONS dataset used in previous studies is not useful in this case as it does not provide segmentation data for the cup which is essential in our case. That is why, in this work, it is not included.

DRISTI-GS dataset from Aravind Eye Hospital, Madurai (India), is made up of 101 color fundus images labeled for both disc and cup; and RIM-ONE dataset from the University of La Laguna is composed of 151 images also labeled for disc and cup.

Figure 4 shows an image from each dataset and makes clear that, even though both provide good quality data for segmentation the characteristics of images from both datasets are significantly different.

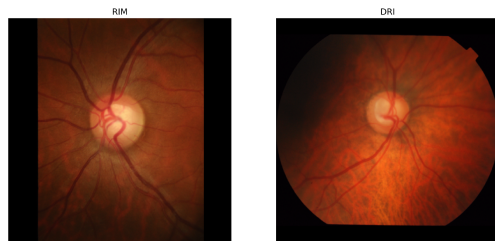


FIGURE 4: Images from RIM-ONE and DRISHTI datasets.

In our work we use 75% of the images from each dataset for training and the remaining 25% of the images for validating the results. However, static (offline) and dynamic (online) data augmentation stages are included in the system's architecture, so the total number of images used for training and testing is much higher than in the original datasets. This can be observed in Table I.

TABLE I: Dataset summary

Dataset	Images	Images after D.A.	Train (75%)	Test (25%)
RIM-ONE	149	6980	5235	1745
DRISHTI	101	2380	1785	595
TOTAL	250	9360	7020	2340

The first column shows the number of images that are provided in those public datasets, the second column indicates the final amount of images used after data augmentation processes and, finally, the other two columns present the number of images used for training and testing purposes, respectively.

B. SYSTEM ARCHITECTURE

Once the problem we want to solve in this work and the datasets used to train the machine-learning system are detailed, it is very important to describe the full system architecture used for training and classification.

Our approach is based on two subsystems whose results are finally combined to produce a diagnosis assistance report for the ophthalmologist. The first subsystem is based on two generalized U-Net based stages to segment the disc and cup plus a feature extraction post-processing stage. The second subsystem is based on a MobileNet V2 [26] network used for direct fundus image classification. There is also a final fusion stage to produce the blended results as a report to assist the ophthalmologist in her or his diagnosis process.

The full system implemented and trained in this work is presented as a graphical abstract in Figure 5 for the first subsystem, and in Figure 6 for the second subsystem. Both figures show all the steps implemented for training and testing each subsystem.

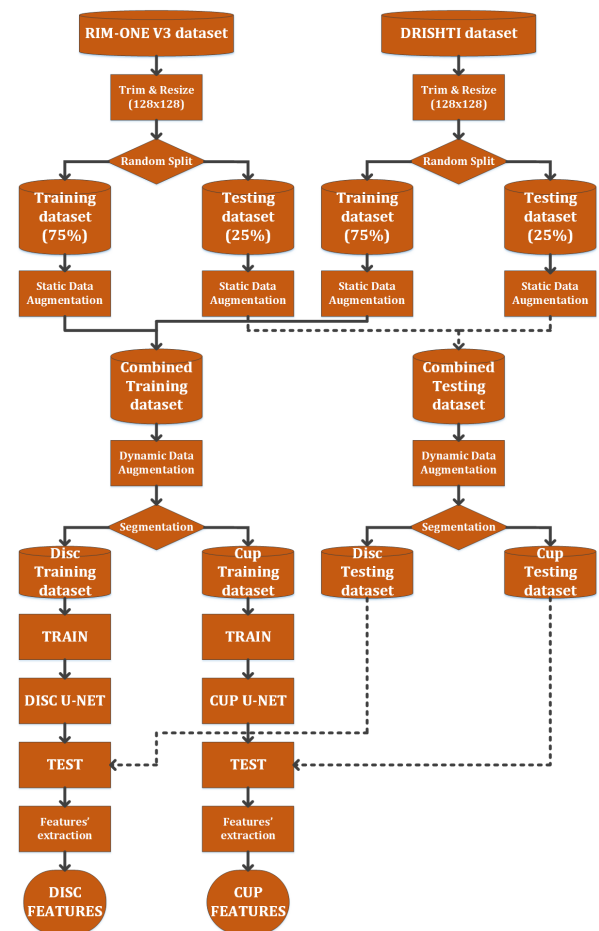


FIGURE 5: First subsystem. Disc and Cup segmentation subsystem.

Several stages can be appreciated in those figures for both subsystems, from the preprocessing stages to the final evaluations. However, results obtained from both subsystems are finally combined in the diagnosis aid tool, and this can be observed in Figure 7.

Next, both subsystem are detailed step by step.

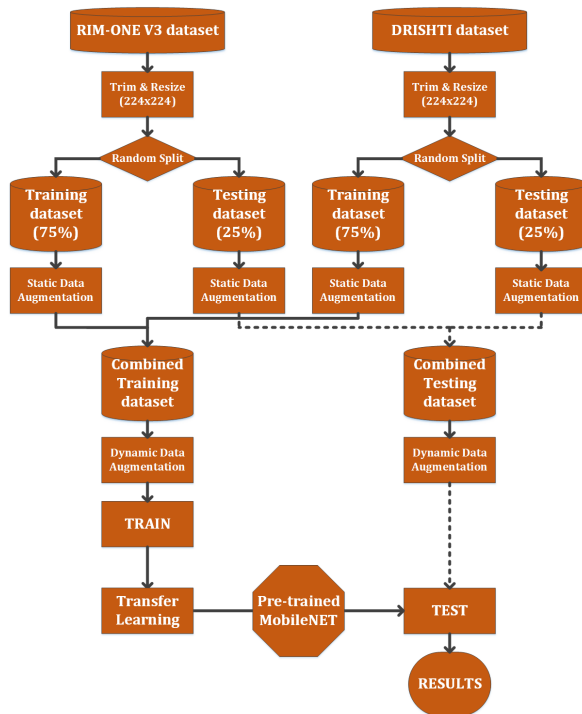


FIGURE 6: Second subsystem. Eye fundus image classification.

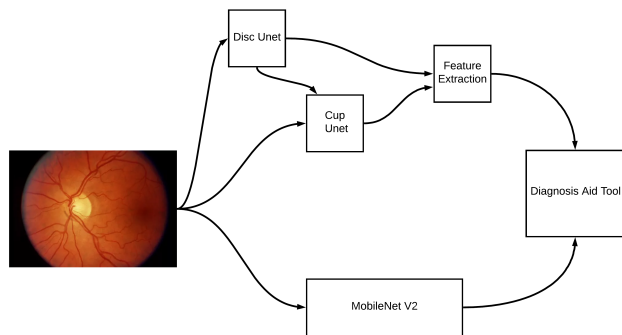


FIGURE 7: Diagnosis Tool Architecture.

1) Segmentation Subsystem

The first subsystem has been named as "segmentation subsystem" as it uses the segmentation process to train two independent systems for disc and cup features' extraction. The different stages implemented for this subsystem are detailed below.

a: Pre-processing

In order to be able to use the dataset images in the segmentation subsystem we need to:

- Perform image trimming to remove borders
- Resize images to the subsystem input size. We use 128×128 images for the segmentation subsystem. We use resampling using pixel area relation for image size reduction.

- Perform contrast limited adaptive histogram equalization.

b: Static (Offline) Data Augmentation

When we are training the system we perform static data augmentation on the training fraction of the combined dataset. This process consists in producing images with modified brightness or contrast parameters. Our data augmentation approach is loosely based on [27].

c: Dynamic (Online) Data Augmentation

When training we use image generators to perform on the fly data augmentation. This process performs moderate zooming and rotations. It is important to understand that glaucoma diagnosis is related to the orientation of the segmented image and, thus rotations should be limited to small angle values (below 15 degrees).

d: Segmentation Network

To segment the disc and the cup from fundus images we use a generalized U-net architecture and train it using Google cloud TPUs. U-net is widely used fully convolutional network that has been widely used for medical image segmentation. This part of the architecture is fully described in [28]. In our case we are using a 6 level network with 64 channels in the first descending stage and a layer channel increment ratio (IR) of 1.1. This model has less than 2.5M trainable parameters and produces good results for both segmentation cases. Although the model has one more stage than the original U-net and the same number of channels in the first layer the reduction of the IR from 2 to 1.1 has decreased the number of parameters from 138M to less than 2.5M. The proposed U-Net implementation block diagram is shown in Figure 8.

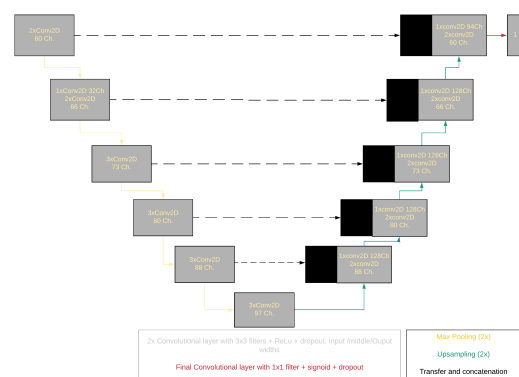


FIGURE 8: Generalized U-Net architecture.

e: Training

Our network is implemented as a recursive function in Keras 2-3-0-tf under Tensorflow 2.2.0. We use 120 image samples as this size is suitable for training using TPUs, GPUs or even CPUs. We use Adam optimizer with dynamically variable

learning rates (between $1e-3$ and $2e-4$) and perform the training process during 100 epochs.

f: Post-processing

It is quite common that some segmentation results are not acceptable to ophthalmologists the main reasons for this are the following:

- The cup and the disc should always be always a single connected region.
- The shape of both regions should be approximately elliptical.
- The size of the optical disc is similar in images captured with the same instrument.

To solve the first problem, in the few cases where segmentation produces multiple regions we select only the one with the largest area. In these cases we decrease the certainty score for the ophthalmologist.

Next we have to establish the similarity between the segmented area and an ellipse. Initially we tried approaches based on the ellipse Hough transform [29] obtaining poor results for our scenario. An approach fitting an ellipse model using Random sample consensus (RANSAC) produces much better results and facilitates the calculation of an ellipse similarity score. When this score is below a certain threshold we also decrease the certainty score.

As a last post filtering stage we penalize those cases where the size of the optic disc is outside a 4 standard deviation interval centered on the disc size mean. This interval is specific for each acquisition instrument. In our case all the image in each dataset have been captured with the same instrument.

2) Direct Classification Subsystem

The other subsystem implemented in this work has been named "direct classification subsystem" as it trains a classical CNN without any segmentation process, so the full images are used to train the system by "brute force". The different stages implemented for this subsystem are detailed in order below.

a: Pre-processing

In order to be able to use the dataset images in the classification subsystem we apply process the images in the same way as for the segmentation subsystem but resize images to the 224×224 images for the classification subsystem.

b: Static Data Augmentation

When we are training the system we perform static data augmentation on the training fraction of the combined dataset. This process consists in producing images with modified brightness or contrast parameters and is very similar to the approach used for the segmentation subsystem.

c: Dynamic Data Augmentation

Static augmentation has proven sufficient in this case and no further improvement was obtained when enabling the dynamic augmentation component.

d: Classification Network

Initially we implemented the classification network using a vgg16 [30] pretrained with the ImageNet 1K challenge [31] weights. This network has been successfully used by other researchers [23] for fundus image classification. This network is relatively large (about 15M parameters) and, thus would make future embedded implementations of our proposed system very difficult. There are, however, newer more efficient alternatives that can lead to similar performance figures. In our case we decided to base our implementation in MobileNet v2. This network is much lighter (less than 2.5M parameters) thus making the embedded implementation of our system feasible. The accuracy of this network on the ImageNet challenge is very similar to that of VGG16, however, its accuracy density, i.e. the accuracy divided the number of parameters is an order of magnitude higher [32]. For our system we remove the top layers of the original MobileNet V2 and add a final classifier based on an average pooling layer whose output is flattened and fetched to an 80 node dense layer, a dropout stage and a final 2 node layer to distinguish between the two required classes.

At the top of the model we include an average pooling layer, a dense layer with 64 nodes and dropout and a final dense layer with 2 nodes to classify our two classes. This can be seen in Figure 9.

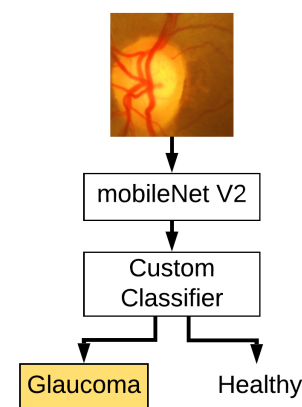


FIGURE 9: Classification subsystem.

e: Training

Our network is implemented as a recursive function in Keras 2-3-0-tf under Tensorflow 2.2.0. We use 64 image batches as this size is suitable for training using TPUs, GPUs or even CPUs. We use a RMSprop optimizer with initial $1e-3$ learning rate with decay and perform training for 50 epochs. This has proven suitable as we are just training the last stages

of the Mobilenet V2 network pretrained with ImageNet plus the additional classifier network.

Once both systems obtains information independently, these outputs may be fused in order to obtain the final output of the diagnosis aid tool (as shown in Figure 7). This fusion is detailed in the next subsection.

C. DATA FUSION AND REPORT GENERATION

The final objective of our system is to help the ophthalmologist in his or her diagnosis. Most Machine learning assistance tool are "oracle based" in the sense that they provide a diagnosis with, in the best case a probability estimation on the reliability of the result.

To be widely accepted by the medical community it is necessary to provide some explanation on the basis on which the result is obtained [33]. Our system does not pretend to be a full-flagged glaucoma diagnosis assistance tool but it provides the physician with:

- The result of the classification subsystem with the assigned probability.
- The result of the segmentation subsystem with the associated calculated CDR.
- The accuracy of the ellipse form matching post-processing stage to let the physician know if the forms of the obtained disc and cup are similar to what should be expected.
- The likeness that the size of the disc is correct.

All these aspects are shown in simplified form in Figure 10.

III. RESULTS AND DISCUSSION

In this section we will use Figure 10 that shows all the intermediate images and data produced by our system to explain the obtained results and compare them with those from other sources.

A. SEGMENTATION SUBSYSTEM

We compare our Disc and Cup segmentation results with other works that use Deep learning based segmentation and use the same fundus image data sets. We have to take into account two important distinguishing features of our work:

- We want to be independent from the specific characteristics of the capture device and, thus, we train with a combined dataset while the compared works train and test independently with each specific dataset.
- We want our system to be very lightweight to be able to implement it in an embedded system in the future.

In Table II we show the Dice coefficient scores for disc and cup segmentation from Sevastopolsky [34] who uses a very light U-Net and provides results for RIM ONE. We also include results Zilly et al. [35] who use a three-layer CNN including sophisticated pre and postprocessing and apply it independently to both data sets. Al-Bander [36] uses a heavily modified, dense U-Net and provides results for both

data sets. Shankaranarayana [37] uses a residual U-Net and provides results for RIM ONE.

We can see, that even though we train with a mixed dataset and use very light segmentation networks our results are fully in line with those obtained by other researchers with heavier networks who train and test specifically with each dataset.

After post-processing by preforming the ellipse conversions we see that the mean values are practically identical which shows that ellipse based approximation is a very good option to codify the disc and cup shapes. Only in very few cases the ellipse extracted from the segmented cup or disc differs significantly from the segmentation provided by the U-net. The RANSAC fitter gives us enough information to signal this cases very easily. When this happens we include this information in the final report for the physician so that he or she knows that the segmentation result has less confidence in this case. An example where the extracted ellipse does not fit well enough with the segmentation data is shown in Figure 11. This specific case (image G8 from the RIM ONE Dataset) is correctly identified as a glaucoma subject both by segmentation and direct classification. It is clear that, although the predicted ellipse is not as large as the correct result the calculated CDR (0.6) is enough to classify the image as coming from a Glaucoma patient.

Regarding the parameter calculation block we estimate the CDR by the relation between the height of the cup bounding box to the height if the cup image. As we cut the cup image to the Disc bounding box, plus a 10% margin on each border which we take into account, this corresponds to the vertical CDR which is the most widely use CDR version. The typical values of this parameters are 0.65 ± 0.13 for glaucoma patients and 0.39 ± 0.15 for healthy individuals [38]. Thus a value between 0.52 and 0.54 seems the most adequate for discriminating both cases. Experimentally we find that the value 0.52 produces the best results with our datasets.

Once we have calibrated our CDR based classifier we can analyze the specificity (recall) and sensitivity of our approach. There are many works that segment the optic disc and cup using many different technologies, however, only a few try to use the segmentation data to do real glaucoma predictions. [39] was one of the first researcher to compute the sensitivity and specificity of their glaucoma predictions using an approach that mixed a morphology based CDR calculation with vessel segmentation in different regions. They reported very good result but based their work in only 15 test cases. On the other hand, [40] used a Stochastic Watershed transformation approach to segmentation with a much larger dataset and obtains a specificity value above 70% with a sensitivity over 60%. When we consider only our segmentation subsystem we get a specificity over 90% with a sensitivity over 75%. It should be clarified that, in a diagnostic assistance tool, which tries to help in a diagnostic but, not to completely carry out automatic diagnosis the main problem are type II error (i.e. false negatives) where a patient with glaucoma is identified as healthy. Sensitivity, the probability that a person with glaucoma is detected as such, is

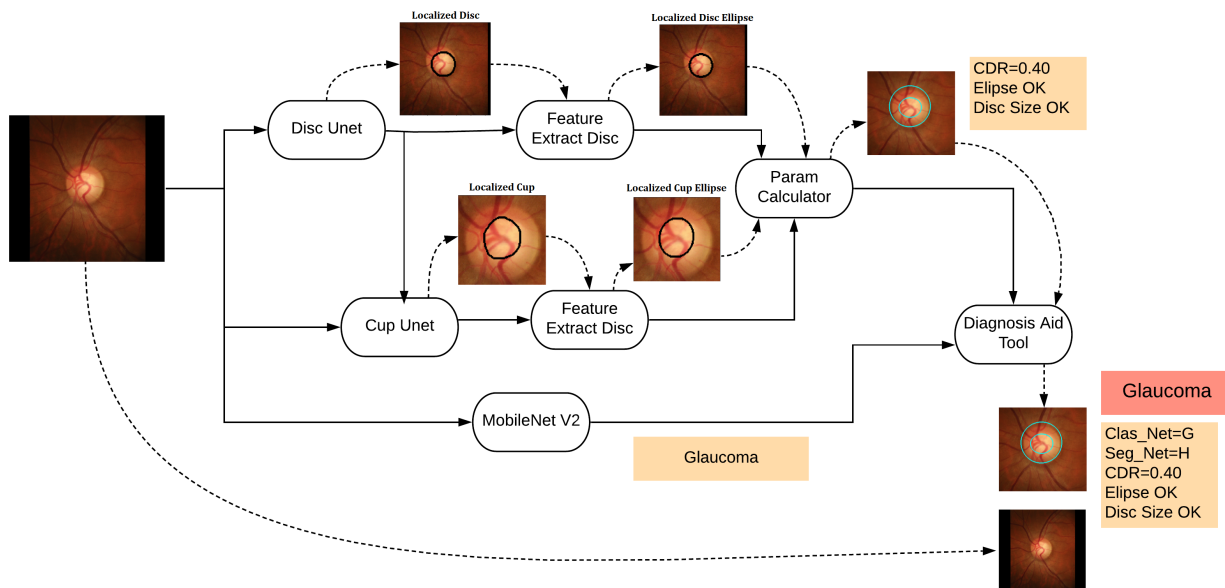


FIGURE 10: System diagram with intermediate data and reports.

TABLE II: Disc and Cup Dice Coefficients

Author	Method	Cup RIM-ONE	Disc RIM-ONE	Cup Drishti	Disc Drishti
Al-Bander [36]	DenseNet	0.69	0.90	0.83	0.95
Sevastopolsky [34]	Modified U-Net	0.82	0.94	-	-
Zilly et al. [35]	Ensemble learning CNN	-	-	0.87	0.97
Shankaranarayana et al. [37]	Fully convolutional adversarial net	0.94	0.98	-	-
This work	Generalized U-Net	0.84	0.92	0.89	0.93

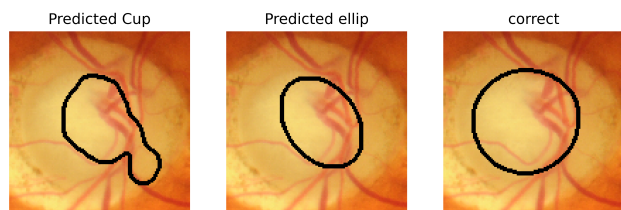


FIGURE 11: Case where ellipse feature extraction has low confidence.

more important than specificity which is the probability that a healthy patient is detected as such.

Table III condenses the sensitivity and specificity data for CDR based diagnosis tools. In the table Se stands for sensitivity and Sp for specificity.

TABLE III: CDR based methods sensitivity and specificity.

	Sp	Se
CDR+vessel [39]	0.80	1.00
Watershed [40]	0.73	0.60
Generalized U-net	0.93	0.76

In Figure 12 we can see the normalized confusion matrix for the U-Net based classifier. We can see that approximately a quarter of the glaucoma cases are classified as healthy using

this approach.

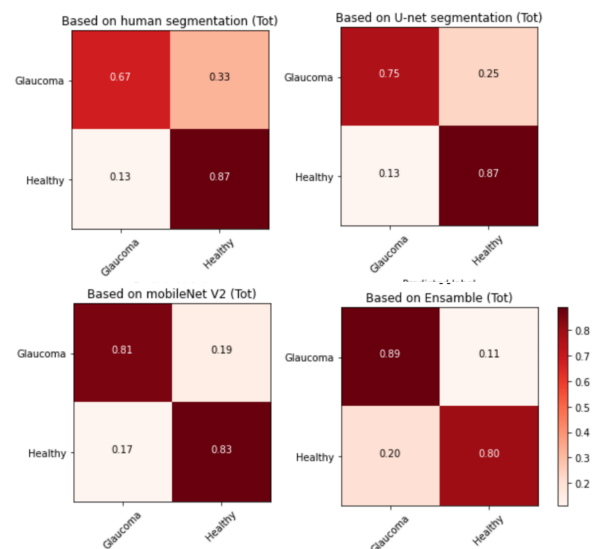


FIGURE 12: RIM-one confusion matrices.

The ROC curve for the U-Net classifier is shown in figure 13. The area under the curve is 0.91 which is better than the

results obtained for in [40] for the CDR based classifier.

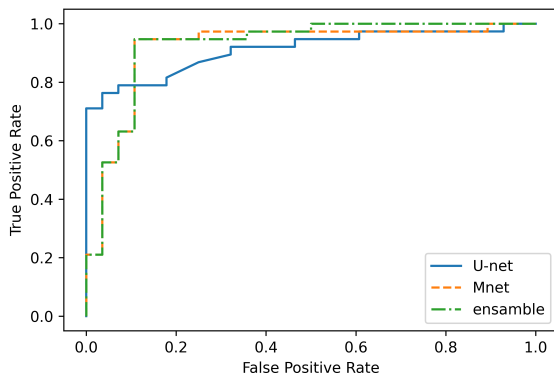


FIGURE 13: ROC for Glaucoma Class.

B. CLASSIFICATION SUBSYSTEM.

Our classification subsystem is based on the very lightweight MobileNet V2. In table IV we compare our system with several classifiers implemented using different networks in [40]. The comparison networks were VGG16 [30], ResNet50 [41] and Xception [42].

We can see that our results are comparable, specially regarding sensitivity, with those obtained by implementations that require at least 20 times more computing performance [32].

TABLE IV: CNN based classifiers Specificity and sensitivity

Network	AUC	Acc	Sp	Se	GF
VGG16	0.96	0.89	0.88	0.90	15
ResNet50	0.96	0.90	0.89	0.91	15
Xception	0.96	0.89	0.85	0.93	10
MobileNetV2	0.93	0.86	0.82	0.89	0.5
Ensamble	0.96	0.88	0.86	0.91	1.5

In Figure 12 we can see the normalized confusion matrix for the MobileNet V2 based classifier. We can see that under 20% of the glaucoma cases are classified as healthy using this approach.

The ROC curve for the MobileNet classifier is shown in figure 13. The area under the curve is 0.93 which is somewhat inferior to other possible alternatives. We have to consider, however, that the implementation is much lighter and that it is designed to be part of an ensemble that is planned to deliver good results as a combined network.

C. ENSEMBLE NETWORK

In table IV we have included also the specificity and sensitivity of the network combining both the U-Net CDR based pattern extraction classifier and the MobileNet direct classifier. It should be clarified that, as our aim is to build a diagnostic assistance tool and, thus we want to avoid false negatives as much as possible our voting scheme decides that a patient is a glaucoma candidate whenever any of the

two networks indicates so. This option improves mainly the network sensitivity. In our case we get a sensitivity of 0.91 which is fully in line with the best available alternatives implemented with a much higher computational cost.

In Figure 12 we can see the normalized confusion matrix for the Ensemble based classifier. We can see that only about 10% of the glaucoma cases are classified as healthy using this approach. This has improved the false negative rate very significantly in comparison with the individual networks that compose the ensemble.

The ROC curve for the Ensemble classifier is shown in Figure 13. In a typical ROC curve construction we modify the threshold on the probability of the result belonging to the analyzed class. In our type of ensemble we have two thresholds that can be chosen independently. Thus, to construct the curve we can chose a strictly increasing function that establishes the relation between the classifiers thresholds. In figure 13 we see an example where we use a linear relation to tie both thresholds. Changing this function to a non-linear relation we can obtain almost any curve that is under the union of the curves for both classifiers. The AUC value provided in IV is an upper limit on the possible values of AUCs for ROC curves that we could construct for the ensemble.

D. REPORTING TOOL

Medical image processing will experiment a breakthrough when ML based diagnostic assistance tools became widely available and accepted in medical daily practice. A problem regarding the adoption of systems is their lack of understandability for the medical professional. This fact has been highlighted by several recent articles, (e.g. [43] [44]) which emphasize the importance of visible (as opposed to black-box) approaches to machine learning based diagnostic assistance. We do not claim that our tool is a full flagged explainable glaucoma diagnosis aid prototype. However we have done an important effort to provide the ophthalmologist with additional data to be able to judge the validity of the proposed diagnosis. This data (see 10 includes information on the adequacy of the size of the segmented disc, the adequacy of the shapes of the disc and the cup, the calculated CDR and the probability of the decision for the direct classification subsystem. We also always provide the initial and the segmented fundus images.

It is clear that understanding our report requires more training than understanding an 'Oracle based' glaucoma or healthy diagnosis but it also gives the physician, who is responsible for the diagnostic decision, much more information on which to base his or her decision.

IV. CONCLUSIONS

Deep learning based diagnosis aid tools are going to be part of the physician daily life in the near future. In this paper we show that a lightweight tool which can be implemented in an embedded system can provide results at the same level of other tools that require much higher computing performance.

The describe tool is based on an ensemble based in two subsystems using completely different technologies. One of the subsystems is a segmentation based network plus a feature extraction post processing stage. The second subsystem is based on a very lightweight last generation classification network which is able to provide the same level of performance as other more traditional heavier networks.

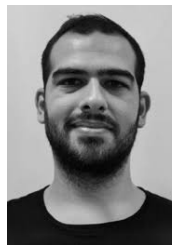
A very important part of our system is the reporting tool which combines the output of both networks and provides the physician with enough data to understand the system's diagnosis proposal and, thus, be able to use it adequately in his or her own final decision.

There are plenty of possibilities for expanding this work and using it to build a useful medically acceptable diagnostic assistance tool. First we would need to train the ensembles with more data coming from public and private datasets. Including a second lightweight classification subsystem (possibly based on EfficientNet [45]). This would improve the reliability and sensitivity of the results even further.

REFERENCES

- [1] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma: a review," *Jama*, vol. 311, no. 18, pp. 1901–1911, 2014.
- [2] H. A. Quigley and A. T. Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *British journal of ophthalmology*, vol. 90, no. 3, pp. 262–267, 2006.
- [3] A. Hommer, "Glaucoma and generics in europe," *OphthaTherapy*, vol. 6, no. 1, pp. 31–36, 2019.
- [4] 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: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.
- [5] L. Quaranta, I. Riva, C. Gerardi, F. Oddone, I. Floriano, and A. G. Konstas, "Quality of life in glaucoma: a review of the literature," *Advances in therapy*, vol. 33, no. 6, pp. 959–981, 2016.
- [6] J. L. Yip, D. C. Broadway, R. Luben, D. F. Garway-Heath, S. Hayat, N. Dalzell, P. S. Lee, A. Bhaniani, N. J. Wareham, K.-T. Khaw, et al., "Physical activity and ocular perfusion pressure: the epic-norfolk eye study," *Investigative ophthalmology & visual science*, vol. 52, no. 11, pp. 8186–8192, 2011.
- [7] E. W. Chan, P. P. Chiang, J. Liao, G. Rees, T. Y. Wong, J. S. Lam, T. Aung, and E. Lamoureux, "Glaucoma and associated visual acuity and field loss significantly affect glaucoma-specific psychosocial functioning," *Ophthalmology*, vol. 122, no. 3, pp. 494–501, 2015.
- [8] M. F. Armary and R. E. Sayegh, "The cup/disc ratio: The findings of tonometry and tonography in the normal eye," *Archives of Ophthalmology*, vol. 82, no. 2, pp. 191–196, 1969.
- [9] G. Wollstein, D. F. Garway-Heath, and R. A. Hitchings, "Identification of early glaucoma cases with the scanning laser ophthalmoscope," *Ophthalmology*, vol. 105, no. 8, pp. 1557–1563, 1998.
- [10] G. Trope, M. Eizenman, and E. Coyle, "Eye movement perimetry in glaucoma," *Canadian journal of ophthalmology. Journal canadien d'ophtalmologie*, vol. 24, no. 5, pp. 197–199, 1989.
- [11] G. Gorin, "Developmental glaucoma*: A concept based on correlation of gonioscopic findings with clinical manifestations," *American journal of ophthalmology*, vol. 58, no. 4, pp. 572–580, 1964.
- [12] M. Iester, M. Mete, M. Figus, and P. Frezzotti, "Incorporating corneal pachymetry into the management of glaucoma," *Journal of Cataract & Refractive Surgery*, vol. 35, no. 9, pp. 1623–1628, 2009.
- [13] A. Septiarini, A. Harjoko, R. Pulungan, and R. Ekantini, "Optic disc and cup segmentation by automatic thresholding with morphological operation for glaucoma evaluation," *Signal, Image and Video Processing*, vol. 11, no. 5, pp. 945–952, 2017.
- [14] M. C. V. S. Mary, E. B. Rajasingh, J. K. K. Jacob, D. Anandhi, U. Amato, and S. E. Selvan, "An empirical study on optic disc segmentation using an active contour model," *Biomedical Signal Processing and Control*, vol. 18, pp. 19–29, 2015.
- [15] N. E. A. Khalid, N. M. Noor, and N. M. Ariff, "Fuzzy c-means (fcm) for optic cup and disc segmentation with morphological operation," *Procedia Computer Science*, vol. 42, no. C, pp. 255–262, 2014.
- [16] M. R. K. Mookiah, U. R. Acharya, C. M. Lim, A. Petznick, and J. S. Suri, "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features," *Knowledge-Based Systems*, vol. 33, pp. 73–82, 2012.
- [17] N. Annu and J. Justin, "Classification of glaucoma images using wavelet based energy features and pca," *Int J Sci Eng Res*, vol. 4, no. 5, pp. 1369–1374, 2013.
- [18] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2017.
- [19] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.
- [20] S.-L. Jhuo, M.-T. Hsieh, T.-C. Weng, M.-J. Chen, C.-M. Yang, and C.-H. Yeh, "Trend prediction of influenza and the associated pneumonia in taiwan using machine learning," in *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 1–2, IEEE, 2019.
- [21] J. Civit-Masot, F. Luna-Perejon, S. Vicente-Diaz, J. M. R. Corral, and A. Civit, "Tpu cloud-based generalized u-net for eye fundus image segmentation," *IEEE Access*, vol. 7, pp. 142379–142387, 2019.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [23] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J. M. Mossi, and A. Navea, "Cnns for automatic glaucoma assessment using fundus images: an extensive validation," *Biomedical engineering online*, vol. 18, no. 1, p. 29, 2019.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [25] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.
- [27] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," *arXiv preprint arXiv:1906.11172*, 2019.
- [28] J. Civit-Masot, A. Billis, M. Dominguez-Morales, S. Vicente-Diaz, and A. Civit, "Multidataset incremental training for optic disc segmentation," in *Proceedings of the 21st EANN (Engineering Applications of Neural Networks)*, Springer-Nature, 2020.
- [29] N. Guil and E. L. Zapata, "Lower order circle and ellipse hough transform," *Pattern Recognition*, vol. 30, no. 10, pp. 1729–1744, 1997.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [32] S. Bianco, R. Cadene, L. Celona, and P. Napolitano, "Benchmark analysis of representative deep neural network architectures," *IEEE Access*, vol. 6, pp. 64270–64277, 2018.
- [33] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
- [34] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 618–624, 2017.
- [35] J. Zilly, J. M. Buhmann, and D. Mahapatra, "Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 28–41, 2017.
- [36] B. Al-Bander, B. Williams, W. Al-Nuaimy, M. Al-Tae, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and

- cup in colour fundus for glaucoma diagnosis,” *Symmetry*, vol. 10, no. 4, p. 87, 2018.
- [37] S. M. Shankaranarayana, K. Ram, K. Mitra, and M. Sivaprakasam, “Joint optic disc and cup segmentation using fully convolutional and adversarial networks,” in *OMIA 2017, Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 168–176, Springer International Publishing, 2017.
- [38] S. MacIver, D. MacDonald, and C. L. Prokopich, “Screening, diagnosis, and management of open angle glaucoma,” *Canadian Journal of Optometry*, vol. 79, no. 1, pp. 5–71, 2017.
- [39] J. Nayak, R. Acharya, P. S. Bhat, N. Shetty, and T.-C. Lim, “Automated diagnosis of glaucoma using digital fundus images,” *Journal of medical systems*, vol. 33, no. 5, p. 337, 2009.
- [40] A. Y. D. Pinto, Machine learning for glaucoma assessment using fundus images. PhD thesis, Universitat Politècnica de València, 2019.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *European conference on computer vision*, pp. 630–645, Springer, 2016.
- [42] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- [43] W. Knight, “The dark secret at the heart of ai,” *Technology Review*, vol. 120, no. 3, pp. 54–61, 2017.
- [44] K. Y. Michael, J. Ma, J. Fisher, J. F. Kreisberg, B. J. Raphael, and T. Ideker, “Visible machine learning for biomedicine,” *Cell*, vol. 173, no. 7, pp. 1562–1565, 2018.
- [45] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” *arXiv preprint arXiv:1905.11946*, 2019.



JAVIER CIVIT-MASOT received the degree in biomedical engineering from the Universities of Seville and Malaga, Spain, in 2018. He is currently pursuing the master's degree in intelligent systems for energy and transport with the Higher Polytechnical School, Seville. He is currently with COBER SL, an SME focused on medical instruments and AI-based diagnosis aids. He has also collaborated with the Medical Physics Laboratory, Aristotle University of Thessaloniki, Greece. His current research interests include deep learning algorithms for diagnostic support, medical instrumentation design, and finite element analysis of intestinal flow.



SATURNINO VICENTE-DÍAZ received the master's degree in computer science and the Ph.D. degree from the University of Seville, Spain, in 1996 and 2001, respectively. Since 2010, he has been an Associate Professor with the University of Seville and also the Vice Dean of the Technical School of Computer Engineering. He has been a Researcher of the Robotics and Computer Technology Group, since 1996. He has authored and coauthored more than 40 articles in refereed international journals and conferences in the fields of robotics, accessibility, embedded systems, and bio-inspired systems. He has participated in more than 20 research projects and contracts, such as the EU projects FLEX, CAVIAR, and CARDIAC.



ANTON CIVIT received his master's degree in physics (electronics) and his Ph.D. from the University of Seville, Spain, in 1984 and 1987. After working for Hewlett-Packard, he joined the University of Seville, where he is currently a Full Professor of computer architecture and head of the Robotics and Computer Technology Research Group. He has authored several papers and research reports on neuromorphic engineering, computer architecture, rehabilitation technology, and robotics. He has also been a member of the European Commission eAccessibility expert group and responsible for several EU projects in the areas of Health, Neuromorphic Engineering and eAccessibility. He is currently associate editor for *Frontiers in Neuroscience*. His current research interests include deep learning based medical diagnosis, medical instrumentation and robotics.

...



MANUEL DOMÍNGUEZ-MORALES received the B.S. degree in Computer Science Engineering in 2008, a M.S. degree in Software Engineering in 2009, a M.S. degree in Computer Engineering in 2014 and the Ph.D. degree on Industrial Informatics in 2014, from the University of Seville (Spain). From 2009 to 2014 he has worked as Assistant Professor at the Computer Architecture and Technology Department of the University of Seville and, since 2014, he works as Postdoc Researcher and Lecturer in the same department. His research fields are focused on computer vision, image processing, signal processing, embedded systems, programmable hardware design, computer architecture, robotics and e-Health.