IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.
Digital Object Identifier 10.1109/ACCESS.2017.DOI

Deep Transfer Learning Strategy to Diagnose Eye-Related Conditions and Diseases: An Approach Based on Low-Quality Fundus Images

GABRIEL D. A. ARANHA[1], and RICARDO A. S. FERNANDES[1], (Member, IEEE)
[1]Graduate Program in Computer Science, Federal University of São Carlos, São Carlos, SP 13565-905 Brazil
Corresponding author: GABRIEL D. A. ARANHA (e-mail: gabriel.aranha@estudante.ufscar.br).

ABSTRACT Data from the World Health Organization indicate that billion cases of visual impairment could be avoided, mainly with regular examinations. However, the absence of specialists in basic health units has resulted in a lack of accurate diagnosis of systemic or asymptomatic eye diseases, increasing the cases of blindness. In this context, the present paper proposes an ensemble of convolutional neural networks, which were submitted to a transfer learning process by using 38,727 high-quality fundus images. Next, the ensemble was tested with 13,000 low-quality fundus images acquired by low-cost equipment. Thus, the proposed approach contributes to advance the state-of-the-art with a novel deep transfer learning strategy, which is more suitable and feasible to be applied by emerging and under-developing countries. From low-quality images, it was able to reach accuracies between 87.4%, 90.8%, 87.5%, 79.1% to classify cataract, diabetic retinopathy, excavation and blood vessels, respectively.

INDEX TERMS Convolutional neural network; deep learning; eye-related conditions; fundus images; transfer learning.

I. INTRODUCTION

Considering the gradual increase in life expectancy, eye complications are prevalent and, commonly, the diseases related to blindness are asymptomatic. According to the World Health Organization (WHO), billion vision impairment cases could be prevented with appropriate treatment and regular examinations [1]. A study from WHO [2] estimates that 80 million people have had a total loss of vision in the world until the end of 2020, wherein almost 90% of them live in emerging countries [3]. Moreover, in emerging and under-developing countries, there are a considerable number of preventable blindness cases, being predominant in the low-income class [4].

From this scenario, clinical decision support (CDS) systems are promising to improve and streamline diagnostics [5], mainly when the amount of information exceeds the limit of human analysis and perception. This way, these systems support clinicians in their decisions based on evidence from the history of diagnoses and registered treatments, mitigating clinical failures, such as suggested in [6]. Additionally, in emerging and under-developing countries, ophthalmology pathologies are aggravated in emergency situations where only a general practitioner is usually present, as well as the access to robust diagnostic equipment is limited in the public health system [7].

Currently, most of the CDSs are based on machine learning algorithms, more specifically on convolutional neural networks (CNNs). According to the review published by [8], the increase use of CNNs to classify pathologies observed in the retina was due to its capability of pattern recognition in large image datasets and the possibility of using pre-trained models with a transfer learning strategy. As mentioned by [9], transferring the base knowledge from a large dataset to a specific domain commonly contributes to improve the model's performance.

Following the above-mentioned context, this paper presents a CNN-based ensemble to classify eye conditions and diseases on real images provided by Brazilian research institutes (the São Paulo Vision Institute and the Federal University of São Paulo). The proposed approach consists of processing images of the ocular fundus and use pre-trained CNN models (VGG16 architecture). These models employ a transfer learning strategy to better adapt to different images' qualities and eye conditions, that are: cataract (operable or not), referable diabetic retinopathy (present or not), excavation (abnormal or not) and blood vessels (abnormal or not).

Different from the approaches found in the literature, the present paper contributes to advancing the state-of-the-art in terms of: (i) validating the proposed transfer learning strategy by recognizing eye-related conditions and diseases in low-quality images (produced by lower-cost equipment); (ii) using high-quality images (obtained by high-cost equipment from referral hospitals) only to train the predictive models; and (iii) reaching results comparable to the state-of-the-art, even using low-quality images. From these contributions, the proposed transfer of learning strategy can improve the service of the public health system, especially when considering a realistic scenario of basic health units from emerging and under-developed countries.

The remainder of the paper is organized as follows. Section II addresses a background review on the use of machine learning algorithms to classify eye-related diseases. Next, the proposed approach is described in section III. The results and discussions are presented in section IV. At last, section V provides the conclusions.

II. BACKGROUND REVIEW

The studies on the use of machine learning algorithms to classify eye-related diseases from images gained notoriety in the last decade [10]–[16]. Next, the approaches that advance the state-of-the-art are presented in a chronological fashion way, focusing on those that consider cataract, diabetic retinopathy (DR), excavation, and blood vessels, since they are the diseases/conditions treated in the present paper.

A. CATARACT

Cataract is a common eye disease responsible to cause vision distortion and blindness if not operated. In this way, some researches were proposed to determine the cataract presence and its grade. An automatic feature learning was proposed by [17] to grade the severity of nuclear cataracts from slit-lamp images. This task was performed through filters (using a CNN) and recursive neural networks. Another feature learning procedure was proposed by [18], where a Wavelet Transform and a sketch method were employed to extract features from fundus images. In order to identify the cataract (presence or absence) and grade it (mild, moderate or severe), a multilabel discriminator was used, which reached a maximum accuracy of 90.9% in a dataset with only 445 images.

In [19], the authors extracted features using the Wavelet Transform, sketch and texture methods. From these features, an ensemble based on support vector machines (SVM) and multilayer Perceptron (MLP) neural networks was used to classify the cataract. The average accuracy obtained was about 93.2% to identify the disease and 84.5% to grade it.

A pre-trained CNN with transfer learning strategy was considered by [20] to extract features. In the sequence, an SVM performs cataract classification. The authors employed public datasets and the labeling task was supported by ophthalmologists. However, this data curation process was not explained in details. The proposed approach was able to reach an average accuracy of 92.91%.

A novel deep learning model, named as CataractNet, was proposed by [21] to identify cataract in a dataset composed of 1,130 fundus images (augmented to 4,746 images). The model obtained accuracies higher than 98%, being the best literature result. However, it is important to mention that, although the authors perform a comparative analysis, it cannot be considered adequate since the data are not the same.

It is worth mentioning that specifically on the cataract identification/classification, a benchmark is not a trivial task given the need for a public dataset properly curated. Even so, it is expected that predictive models are able to achieve accuracies above 90%, as also shown in [18]–[20].

B. DIABETIC RETINOPATHY

DR affects the retinal blood vessels of individuals and, for this reason, some researches were addressed to classify its grade. However, the identification of referable DR (presence or absence) is of great importance, as its early-stage detection allows preventive measures to be taken.

In [22], the authors proposed a comparison between the Naïve Bayes and SVM to classify DR. Although the Naïve Bayes achieved an accuracy of 83.37% against 64.91% obtained by the SVM, the dataset used was composed of only 300 images.

One of the most cited studies in the field was proposed by [23], where the authors trained a CNN Inception-v3 using data from EyePACS and Indian hospitals to classify DR. However, the model was validated in both EyePACS and Messidor-2 public datasets that were curated by ophthalmologists. The results presented specificity between 87.0% and 93.9%, while the sensitivity was between 96.1% and 98.5%.

Retinal images from multiethnic populations were classified by [24] using a deep learning model, that was validated with a dataset composed of 494,661 images. This model obtained sensitivity and specificity of 90.5% and 91.6%, respectively.

Another approach based on the CNN Inception-v3 was proposed by [25], where the authors also used the EyePACS and Messidor-2 datasets. They demonstrated that, without data curation, it was not possible to obtain the same classification rates as [23]. Results between 71.2% and 83.6% were reached for specificity, while the range between 68.7% and 92.0% were reached for sensitivity.

A visualization tool and deep learning models were proposed by [26] to support decisions in terms of referable DR. The models were trained and validated using 48,116 retinal images. To visualize the meaningful features for clinicians, the authors applied a CNN-independent adaptive kernel technique based on a sliding window to crop images and produce a feature map. Good results were obtained (considering a confidence interval of 95%), reaching 96% of true positive cases.

In order to verify the performance of a CNN-based ensemble (composed of Resnet50, Inception-v3, Xception, Dense121, and Dense169), the authors of [27] seek to classify five stages of DR, considering the EyePACS dataset. The approach was evaluated through accuracy, recall, specificity, precision, and f1-score, obtaining 80.8%, 51.5%, 86.7%, 63.8%, and 53.7%, respectively.

InceptionResNet-v2 CNNs were used by [28] to determine the retina status (healthy or pathological) and to classify DR (healthy, non-referable or referable). A Gaussian filter was used to preprocess images and a margin max technique was deployed to improve the models' sensitivities. In this sense, they were able to respectively obtain sensitivity values close to 80% and 98% for pathological and healthy retinas.

From the above-presented studies, it is also notable a difficult to establish an adequate comparison, since they used different datasets, not detailed data curation and evaluation metrics.

C. EXCAVATION

Another form of diagnosis and staging of ocular pathologies is the analysis of the optic nerve. Typically, this kind of diagnosis aims to validate if the diameter of the optic disc, also known as excavation, is more extensive than average, denoting a possible disease like glaucoma.

The first glaucoma detection study reported in the literature using CNN [29] developed a six-layer architecture, being the first four convolutional and the remaining two fully-connected. Furthermore, the output of the last fully-connected layer was provided to a softmax classifier trained in the pathology domain. Images from the ORIGA (650 images) and SCES (1,722 images) datasets were used. Results demonstrated an area under the curve (AUC) of 82.3% and 86%, respectively.

Employing two stages, the authors of [30] firstly located and extracted the optic nerve from the retina with a region-based CNN. In the sequence, the images were classified (healthy or glaucoma) using another deep Learning model. The approach was validated in the ORIGA dataset, reaching an AUC of 87.4%.

Focusing only on glaucoma disease, the study proposed by [31] presented a transfer learning approach, combining it with data augmentation and uncertainty sampling. The authors used a dataset composed of 4,933 fundus images, which was divided into training and validation subsets. A deep learning model achieved an AUC of 99.5%, sensitivity of 98.0% and specificity of 91% for individuals with and without glaucoma.

A deep neural network approach was proposed by [32] to segment the optic nerve region and classify the image as pathological (glaucoma) or healthy. This model used pixel-level features obtained from the segmentation process and image-level features extracted from the images. Two public datasets were used for validation, REFUGE and DRISHTI-GS, whose results in 97.60% and 94.74% of AUC, respectively.

D. BLOOD VESSEL

The blood vessel analysis in fundus images is of paramount importance to support the identification of eye-related and systemic (e.g. hypertension and ischemic stroke) diseases. There are a considerable number of studies on the segmentation of the eyes' vascular structure, since it is commonly used as a procedure before classification [33].

A K-Nearest Neighbors-based approach was proposed by [34] to classify arteries or veins, which used segmented images as inputs. It was extracted the structural profile of the vein or artery, color, and textures. They used the DRIVE dataset. The results demonstrated an average accuracy of 92.3%.

In the scope of detecting glaucoma using blood vessel segmentation, the authors of [35] developed a method for classifying the severity of the pathology. As a first stage, they extracted features from the vessel images, which were used as inputs to a hybrid model composed of an Adaptive Neuro-Fuzzy Inference System (ANFIS) and an SVM. The sensitivity, specificity, and accuracy were calculated to verify the model's performance, reaching 85.7%, 92%, and 97.77%, respectively. Information about the dataset is scarce.

Aiming at classifying non-specific eye conditions that can cause blindness, the work of [36] employed an Inception-v3 that has in its composition an attention network indicating the most prone pixels to represent discriminating information. The images were labelled in grades (from 0 to 2), where the grade 0 represents greater severity. The dataset used for the model was the EIARG1, with 120 images instituted with the premise of denoting the severity of ocular anomalies. However, as it is an atypical approach, only the network models created by the authors were compared, whose best result was the f1-score equals to 0.92, obtained by the Inception-v3 model.

Having an unsupervised model, the authors of [37] proposed an approach that classifies arteries and veins that only takes advantage of the local contrast between blood vessels and the background to the surroundings. A graph was used to represent the vascular structure of the retina. Thus, a multilevel threshold was applied to the graph to properly classify the images. Next, the arteriole-venular ratio was calculated. This approach was applied to the INSPIRE and DRIVE datasets, obtaining accuracies of 96% and 80% respectively.

III. DEEP TRANSFER LEARNING-BASED APPROACH

The deep transfer learning approach proposed in the present paper uses high-quality and low-quality fundus images produced by a Brazilian reference vision institute. These images were labelled by ophthalmologists and storage in a centralized way to be processed and used to retrain an ensemble-based model (composed of CNNs with VGG16 ImageNet architecture). Primarily, on a machine learning server, only high-quality images were used to generate the baseline model. Thereafter, a retrain was made only using low-quality images based on the knowledge acquired in the previous model. Once generated, the model can be made available to public health units, where low-quality images would be given as inputs to the model. A general overview is depicted on Fig. 1, which could be used by other emerging and/or under-developing countries.

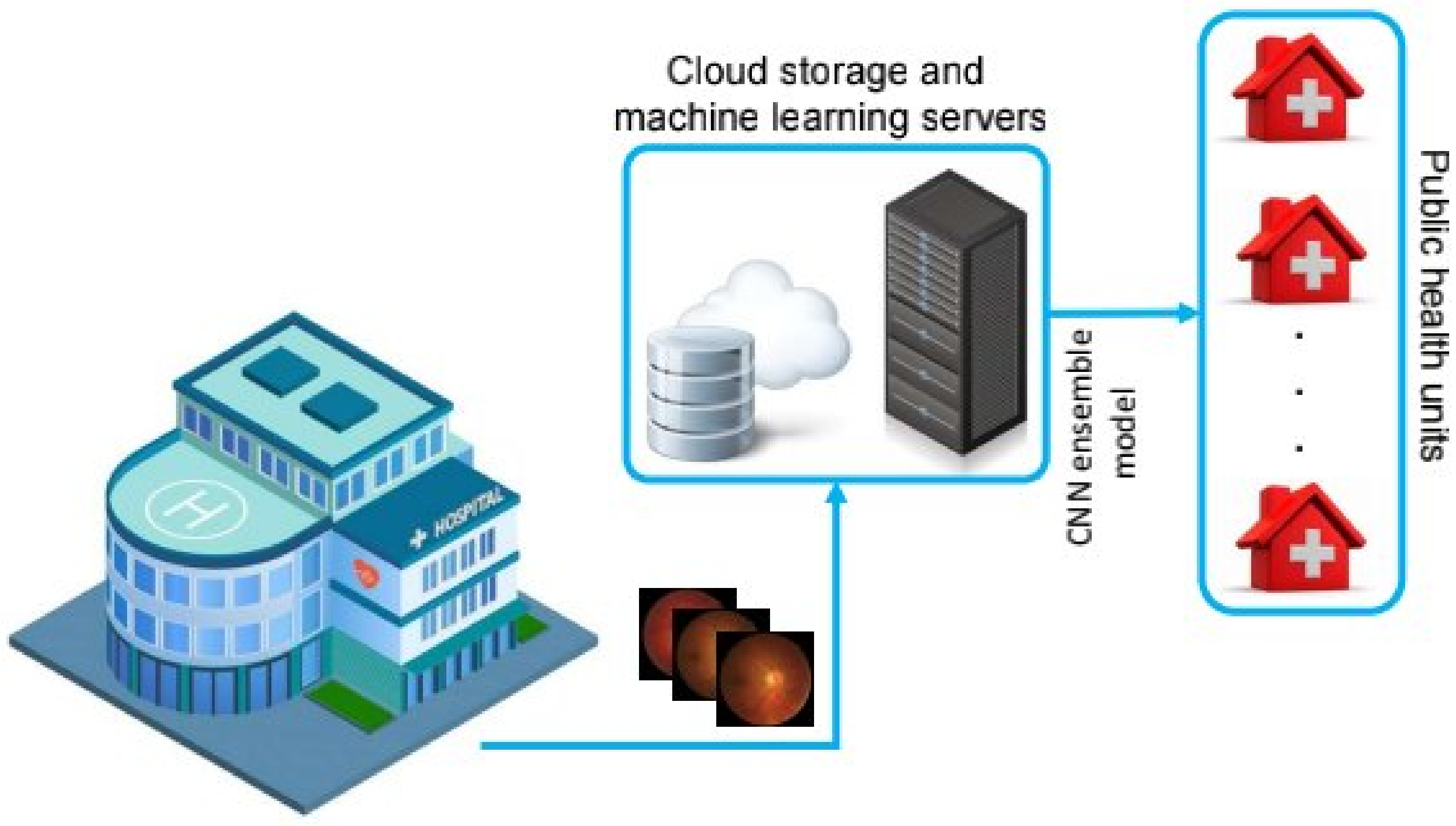

FIGURE 1. General overview of the proposed approach when employed by public health units.

On the cloud storage and machine learning server side, the images were firstly collected and preprocessed (subsection III-A). In the sequence, the CNNs VGG16 were retrained (from ImageNet) in the domain of high-quality fundus images (subsection III-B), and then a novel transfer learning strategy was applied by considering low-quality images produced by low-cost equipment (subsection III-C). Finally, the ensemble model composed of CNNs was tested using only low-quality fundus images and, for this purpose, some performance metrics were considered (subsection III-D).

A. DATA COLLECTION AND PROCESSING

The high-quality images used to retrain the CNNs VGG16 come from the diagnosis base of the São Paulo Vision Institute, which were collected from 2018 to 2020. These images were generated by two equipment, that are: (i) Canon CR2 – Phelcom Eyer; and (ii) Topcon NW100. On the other hand, the low-quality images acquired by public health units were generated using a 3Nethra Classic. All images were labeled by five ophthalmologists with great expertise in the area (working in different periods and units). Thus, the images were classified into: cataract (operable or not operable); referable DR (presence or absence); excavation (abnormal or normal); and blood vessels (abnormal or normal). Examples of these labeled images are presented in Fig. 2.

Since the images were generated by different equipment in both datasets (high- and low-quality), variations in size, aspect ratio, color, focus, and quality are present. Consequently, all the images had to be anonymized and standardized in a size of 299×299 pixels (in RGB scale), where the ocular fundus was centered and cropped.

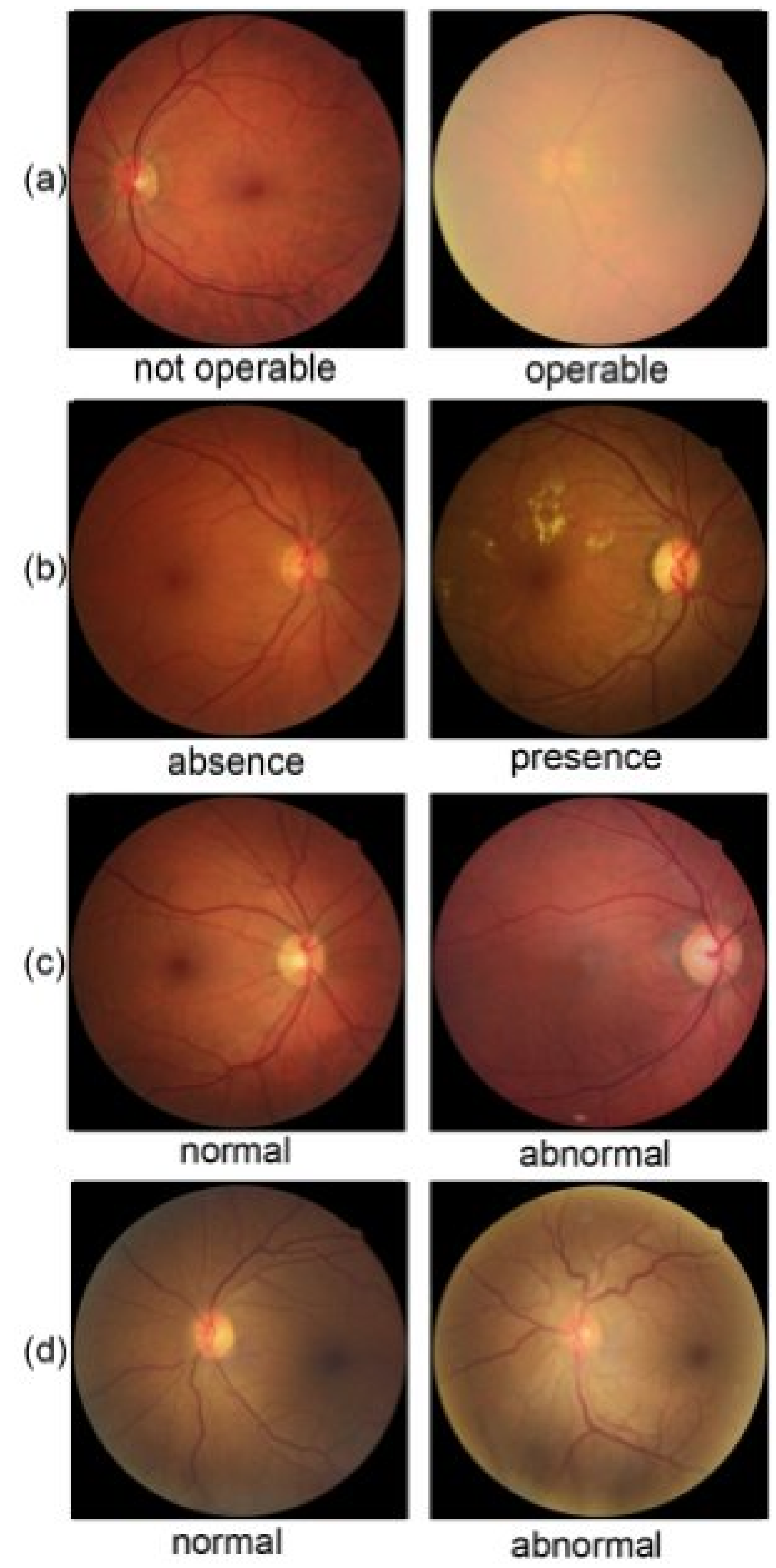

FIGURE 2. Examples of fundus images for: (a) operable and not operable cataract; (b) presence and absence of referable DR; (c) abnormal and normal excavation; and (d) abnormal and normal blood vessels.

In both datasets there are mis-captured images, where a large portion of the image is practically black or white (as shown in Fig. 3).

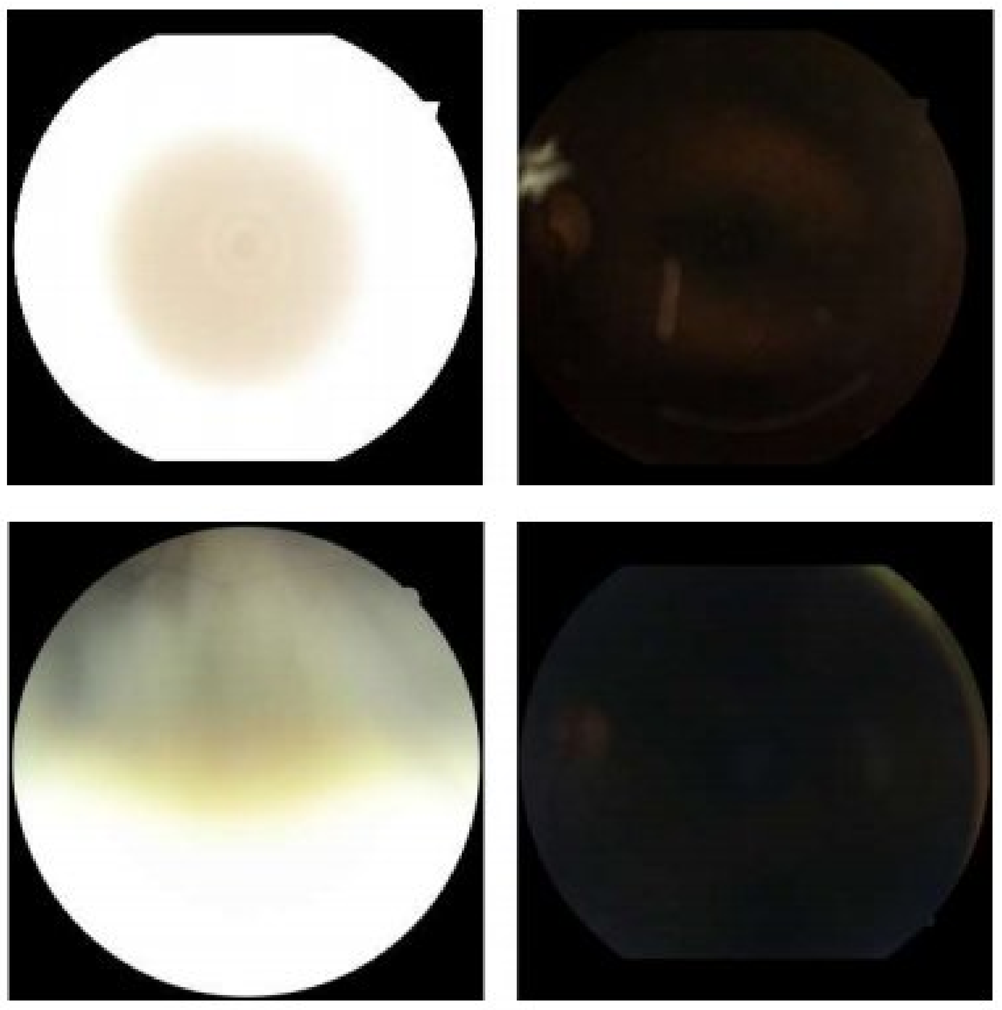

FIGURE 3. Example of mis-captured images.

Thus, a filter was created to remove them, which converts the RGB image into grayscale and calculates the average value of the pixels (avg_{pixels}). In this sense, two thresholds were empirically defined ($thr_{black} = 15$ and $thr_{white} = 145$) to eliminate the images when the following rule is true:

$$\textit{if}\ (avg_{pixels} < thr_{black})\ \textit{or}\ (avg_{pixels} > thr_{white}).$$

Less than 500 images were discarded from the high- and low-quality datasets. Thus, the high-quality dataset consists of 68,171 images (being 19,360 of cataract; 5,833 of DR; 3,866 of abnormal excavation, 19,488 of abnormal blood vessels; and 19,624 of normal condition). Otherwise, the low-quality dataset was formed by 7,850 images (being 1,474 of cataract; 986 of DR; 763 of abnormal excavation; 1,046 of abnormal blood vessels; and 3,581 of normal condition).

B. ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS

The CNN VGG16 model is basically composed of convolutional, pooling and dense layers. It aims to convert the image into representations of depth [38], filtering the pixel information for a smaller mapping. As can be seen in Fig. 4, this process was initialized by the convolutional layers (which scan a matrix of pixels), followed by a max pooling layer. The weights of convolutional and pooling layers were kept in accordance with those obtained for ImageNet, i.e., these layers were pre-trained with the premise of granting greater adaptability to the network [39]. Next, the dense layers were modified (in relation to the CNN VGG16 base model) by using fully connected layers and a global average pooling (GAP) layer. The purpose of using the GAP layer is to avoid overfitting as it averages all feature maps as opposed to a flattened layer which transforms a normal layer into a one-dimensional layer keeping all the original values, thus using a layer GAP makes the use of dropout optional [40]. Not to mention that, for future work, the GAP layer supports the generation of a Class Activation Map, a very important feature to corroborate the regions of interest of an image.

Different from the convolutional and pooling layers, the dense layers were trained in the domain referred to in this work, i.e., the high-quality fundus images. At last, a softmax function was used to generate the output. It is important to mention that, to prevent the propagation of negative values between the network layers, the rectified linear unit (ReLU) activation function was used to the convolutional and fully connected layers.

The Tensorflow framework with the Keras library was adopted, considering the use of stochastic gradient descent, a learning rate of 0.01, a weight decay of 1^{-6} and momentum of 0.9.

Since we have employed a CNN for each specific domain (cataract, referable DR, excavation and blood vessels), the output layer was binary. The CNNs were trained by considering a batch size of 32. The entire training has 30 epochs, being the steps per epoch defined as the number of samples divided by the batch size. For each domain, 80% of the total number of images were used to train the CNNs, while 10% was used to validate and the remaining 10% to test the models.

C. TRANSFER LEARNING

Transfer learning is used to improve learning from one domain by transferring information from a related domain. As exemplified in [41], consider two people who want to learn to play piano. One person has no previous experience playing music, and the other person has extensive musical knowledge of playing the guitar. The person with a musical background will be able to learn piano more efficiently, transferring previously learned musical knowledge to the task of learning to play the piano.

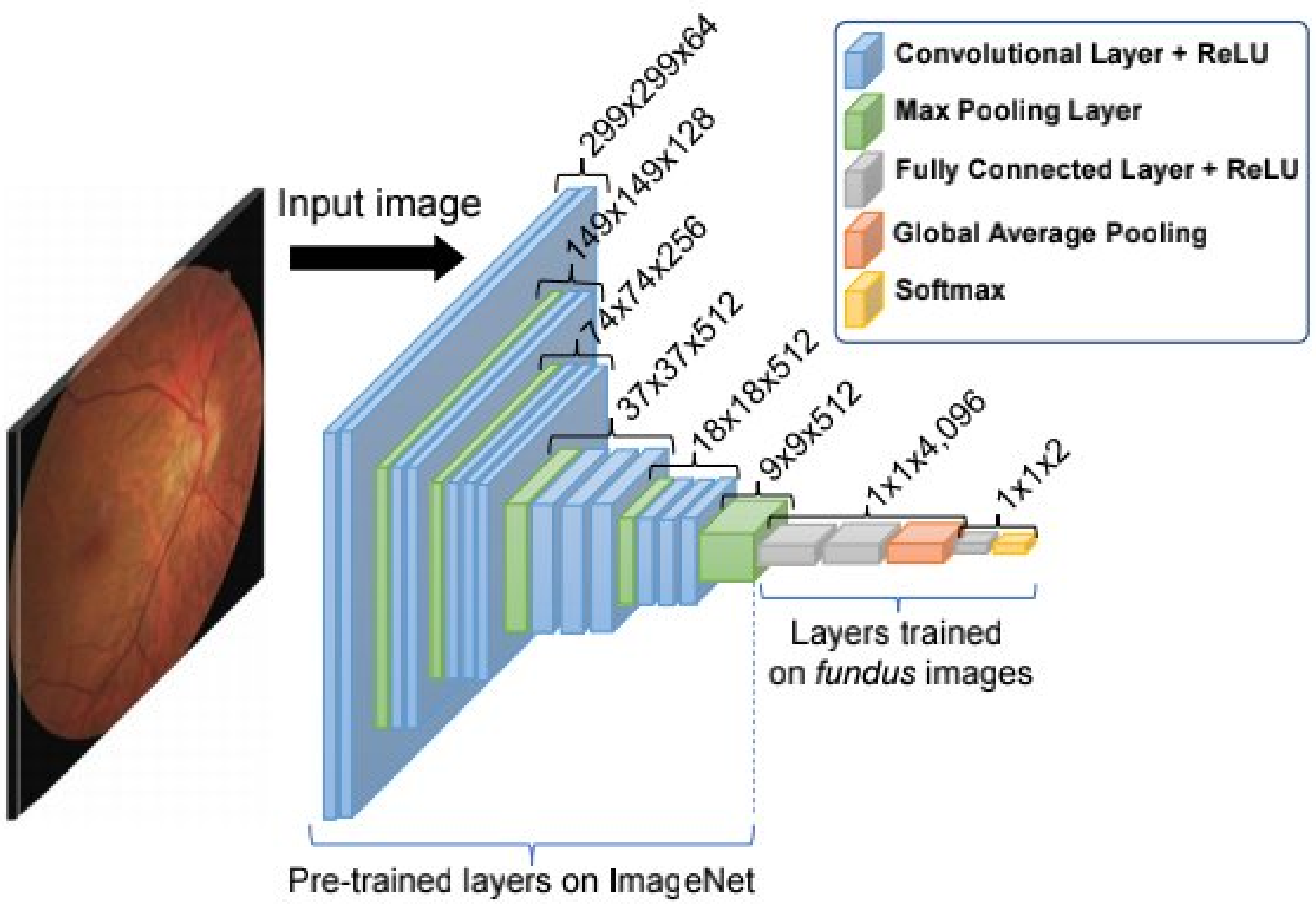

FIGURE 4. CNN VGG16 architecture with dense layers trained on fundus images.

According to [42] there are three categories of transfer learning, that is, transductive, inductive, and unsupervised. To summarize, transductive learning represents the scenario when all the knowledge comes only from the source domain. inductive learning refers to a situation where the label information of the target domain is available. Whether the information is unknown for both, the source and target domain, the transfer learning is categorized as unsupervised.

There are numerous applications of transfer learning in the context of machine learning, however, in the elaboration of CNNs, this process has gained notoriety for increasing the efficiency in the acquisition of knowledge of a certain domain of images, through the learning of a generic domain and not so correlated [43]. An example of this process is the use of weights that were trained in the ImageNet database, a database of generic images such as balls, cars, and similar, in a specific domain such as lung disease detection [44].

This paper starts from the premise that low-quality images present unique aspects concerning image proportion, colors, and sharpness. Hence, it is conceived that there is a transfer learning between one variety of image to another and not a permutation of images that compose the same set. An example of the differences amid the two groups of images can be seen in Fig. 5.

By using the transfer learning procedure on CNNs, it is possible to load the weights from the previous domain into a network and unlock only the last trainable layers (arbitrarily chosen) or unlock all layers for retraining in the new domain. Among all the tests done to acquire the best transfer learning process, the one that unfreezes all the trainable layers, including those pre-trained on ImageNet and previously frozen,

FIGURE 5. Differences between fundus images: (a) high-quality; and (b) low-quality.

from the base CNN architecture (high-quality model), and processes the new dataset (low-quality) with a lower learning rate, precisely 0.0001, was the one that generated the best results.

D. PERFORMANCE EVALUATION METRICS

To evaluate the performance of each CNN, metrics regularly seen in similar works were chosen. In this sense, it was considered the calculation of sensitivity (sens), specificity (spec), and accuracy (acc). These metrics are respectively presented in the sequence:

$$sens = \frac{TP}{TP + FN}, \tag{1}$$

$$spec = \frac{TN}{TN + FP}, \tag{2}$$

$$acc = \frac{(TP + TN)}{(TP + TN + FP + FN)}. \tag{3}$$

Next, the results were analyzed and discussed to demonstrate the robustness and effectiveness of the proposed deep transfer learning strategy against the state-of-the-art papers.

IV. RESULTS AND DICUSSIONS

This section divides the results into: (A) proprietary dataset, which implies that a comparison is not applicable; and (B) public datasets, in which the proposed approach was compared with similar works. In both scenarios (public or proprietary datasets), the proposed work has been trained each eye condition separately, each model used its own images for the altered retina (of each condition), but the normal images are shared with all models from the dataset.

A. PROPRIETARY DATASET

Focusing on the proprietary dataset, the 4 models representing each eye condition addressed were trained in a high-quality set, after this, they were retrained in a low-quality dataset denoting the transfer learning method. To compare with this approach, another 4 models were trained only in the low-quality dataset. Elucidating the importance of the proposed approach, the results obtained with the model with transferred knowledge were shown in Table 1.

Given the fact that this research proposes an approach capable of assisting decision-making by ophthalmologists,

TABLE 1. Comparative results on proprietary dataset. Both strategies were trained in low-quality datasets, but one of them has knowledge transferred from a high-quality dataset.

CNN Model	Strategy	Spec.	Sens.	Acc.
Cataract	w/o high-quality TL	0.814	0.886	0.853
	with high-quality TL	0.829	0.901	0.874
DR	w/o high-quality TL	0.720	0.855	0.754
	with high-quality TL	0.886	0.932	0.908
Excavation	w/o high-quality TL	0.810	0.902	0.866
	with high-quality TL	0.809	0.911	0.875
Blood Vessels	w/o high-quality TL	0.682	0.807	0.739
	with high-quality TL	0.741	0.813	0.791
Average	w/o high-quality TL	0.757	0.863	0.803
	with high-quality TL	**0.816**	**0.889**	**0.862**

it is possible to state that transferring the knowledge from one kind of dataset to another (high- to low-quality) shows a considerable improvement in the models' performance.

Although the proposed approach has achieved average results that exceed 81.6% of the evaluation metrics, it is difficult to compare it to the state-of-the-art due to the use of private data. The same issue occurs when analyzing the literature since some of the papers did not employ public datasets, or perform a new grading method which is not shared [23], [45]–[48]. Additionally, there is no consensus regarding the evaluation metrics. For these reasons, in the sequence, the proposed approach was tested on public datasets of cataract, DR and glaucoma. Thus, the results could be adequately compared with those of other works.

B. PUBLIC DATASETS

To verify the assertiveness of the model on the public data, this work used some datasets commonly approached in literature. The domains mentioned in the comparison were DR, cataract, and glaucoma. Additionally, as DR is the most researched classification disease in eye fundus image, this work proposed two comparisons with the commonly used datasets in this study field, Messidor-2 [49] and EyePACS [50]. Considering the cataract domain, the ODIR [51] dataset was selected, and REFUGE [52] for glaucoma.

It is worth noting that the use of public datasets, in essence, is of paramount importance to define benchmarks for predictive models. However, it can be noticed that many of these researches curate data and, therefore, change the class label of some instances. In this sense, it is difficult to establish an adequate benchmark, since such changes are just mentioned (not discussed) in the papers.

1) Diabethic Rethinopaty

As mentioned above, Messidor-2 and EyePACS were considered to evaluate the DR domain. The Messidor-2 data has 1748 images with 4 classes, from 0 (no DR) to 3 (severe DR). Whereas, EyePACS consists of 9963 divided into: No DR, Mild, Moderate, Severe or Proliferative DR. This approach divided the data into referable DR and no referable DR for both datasets.

A meta-analysis of Deep Learning Systems for DR was made by [53] with the intention of listing the types of approaches with their respective performances, and based on those types, Table 2 shows the comparison between this research and works with the same scope, i.e., using the two aforementioned datasets, containing all stages of the pathology including the mild ones to create a binary classifier. It is important to emphasize that the studies condensed in this comparison use the same data labeling system, without a specialized curation of other professionals, as the curation process is subjective and difficult to standardize.

TABLE 2. Comparison with the state-of-the-art papers on DR public datasets.

Reference	Dataset	AUC ROC	Sens.	Spec.
[25]	EyePACS	0.951	0.906	0.847
	Messidor-2	0.853	0.818	0.712
[28]	EyePACS	–	0.832	0.890
	Messidor-2	–	–	–
[54]	EyePACS	0.764	0.911	0.500
	Messidor-2	0.912	0.940	0.500
Proposed approach	EyePACS	0.951	0.913	0.847
	Messidor-2	0.953	0.915	0.849

2) Cataract

As an illustration of the positive result in cataract ambit, the dataset called ODIR of the Kaggle platform was chosen, whose state-of-the-art contained in 21 shows an accuracy of 98.62%, for tests without the data augmentation process. The dataset consists of 1,400 images, and the training, validation and test set configuration contained in this research is the same as in [55]. Even knowing that accuracy is not the most relevant metric in medical analyses, the average result of 99.28% obtained in this present work corroborates the adherence of the proposed model to aid cataract classification, as compared in Table 3.

TABLE 3. Comparison with the state-of-the-art papers on cataract public dataset.

Reference	Dataset	Acc.
[21]	ODIR	0.986
[55]	ODIR	0.975
[56]	ODIR	1.00
Proposed approach	ODIR	0.993

3) Glaucoma

In the context of glaucoma, it is common to find works on segmentation of the optic nerve, however, in addition to the ground-truth values of the exact segmentation of excavations, the REFUGE dataset [52] also has the classifications of healthy or pathological retinas, in this case, with glaucoma. The dataset consists of 1,200 annotated retinographies, of which 121 correspond to eyes with glaucoma. The retinographies in this dataset are centered on the macula and have a size of $1,634 \times 1,634$ or $2,124 \times 2,056$ pixels. The default dataset split consists of 400 images for testing and 800 for training and validation. Even so, the networks proposed in this present research stood out, as demonstrated in Table 4.

TABLE 4. Comparison with the state-of-the-art papers on glaucoma public dataset.

Reference	Dataset	AUC ROC
[32]	REFUGE	0.976
[52]	REFUGE	0.989
Proposed approach	REFUGE	0.996

4) Public Datasets Discussions

Even with the generalist proposal, there is a great capacity of the proposed network to adapt, since the results obtained in the aforementioned public domains generated good results. This process endorses the possibility of canonical models that, added to supplementary tools for filtering, collecting, and improving data, serve with mastery to support decision-making in multiple domains with scalar performance for all domains.

The acquired results can be interpreted as a possibility to calculate the results of multiple networks in sequence in a pipeline format for the identification of a retinography, thus generating a ranking of probabilities of pathologies that support the analysis and inference of the professional in the area of ophthalmology.

It is clear that cured data improve the accuracy of the models since their performance is mainly a result of the quality and labeling of the data, however, it should be noted that the brightness filtering process (subsection III-A) can be used in a real environment as a support tool in the image collection, validating the submitted images that fit the acceptable thresholds.

V. CONCLUSIONS

It was found that the problem of scarcity of data in a basic health unit, which, occasionally, can only have low-quality images, was solved with transfer learning. Compared with related studies, this is the first time that a machine learning model has been proposed to classify four eye conditions, contributing to better guiding decisions in a realistic scenario of basic health units in emerging and underdeveloped countries and, consequently, preventing visual impairment or blindness.

This research presented a generalist approach in the generation of the models, it was able to produce considerably accurate results. The premise of this work was to support the inference of the health professionals of the Instituto da Visão de São Paulo with the same information that they already work in their proprietary dataset. Thereupon, with the results obtained, it is possible to conclude that the approach has a viable practical adhesion.

Undoubtedly that by condensing several domains into a single solution, with more common metrics of the medical environment, it contributes to the acceleration of the use of support systems for decision-making in real scenarios.

As the dataset offered was not designed to be used as an input in Deep Learning algorithms, it is possible to audit the data in order to improve the natural classification of images or to increase the domain of observed conditions. Furthermore, with the amount of data already collected and the constant collection of new ones, it is possible to create and share a robust dataset, which will allow more efficient training for the idealized models.

ACKNOWLEDGMENT

The authors would like to thank the Federal University of São Carlos for the facilities provided.

REFERENCES

[1] WHO, "Blindness and vision impairment," 2021.

[2] ——, "World report on vision," 2019.

[3] R. Bourne, J. D. Steinmetz, S. Flaxman, P. S. Briant, H. R. Taylor, S. Resnikoff, R. J. Casson, A. Abdoli, E. Abu-Gharbieh, A. Afshin *et al.*, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study," *The Lancet Global Health*, vol. 9, no. 2, pp. e130–e143, 2021.

[4] J. M. Furtado, A. Berezovsky, N. N. Ferraz, S. Muñoz, A. G. Fernandes, S. S. Watanabe, C. C. Cunha, G. C. Vasconcelos, P. Y. Sacai, M. Cypel *et al.*, "Prevalence and causes of visual impairment and blindness in adults aged 45 years and older from parintins: the brazilian amazon region eye survey," *Ophthalmic Epidemiology*, vol. 26, no. 5, pp. 345–354, 2019.

[5] A. Baker, *Crossing the quality chasm: a new health system for the 21st century*. British Medical Journal Publishing Group, 2001, vol. 323, no. 7322.

[6] E. A. McGlynn, S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr, "The quality of health care delivered to adults in the united states," *New England Journal of Medicine*, vol. 348, no. 26, pp. 2635–2645, 2003.

[7] I. De la Torre-Díez, B. Martínez-Pérez, M. López-Coronado, J. R. Díaz, and M. M. López, "Decision support systems and applications in ophthalmology: literature and commercial review focused on mobile apps," *Journal of Medical Systems*, vol. 39, no. 1, pp. 1–10, 2015.

[8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[9] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep transfer learning for modality classification of medical images," *Information*, vol. 8, no. 3, p. 91, 2017.

[10] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Dream: diabetic retinopathy analysis using machine learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 5, pp. 1717–1728, 2013.

[11] R. Priya and P. Aruna, "Diagnosis of diabetic retinopathy using machine learning techniques," *ICTACT Journal on Soft Computing*, vol. 3, no. 4, pp. 563–575, 2013.

[12] S. Somasundaram and P. Alli, "A machine learning ensemble classifier for early prediction of diabetic retinopathy," *Journal of Medical Systems*, vol. 41, no. 12, pp. 1–12, 2017.

[13] A. Ali, S. Qadri, W. Khan Mashwani, W. Kumam, P. Kumam, S. Naeem, A. Goktas, F. Jamal, C. Chesneau, S. Anam *et al.*, "Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image," *Entropy*, vol. 22, no. 5, p. 567, 2020.

[14] T. Vijayan, M. Sangeetha, A. Kumaravel, and B. Karthik, "Gabor filter and machine learning based diabetic retinopathy analysis and detection," *Microprocessors and Microsystems*, p. 103353, 2020.

[15] M. Alam, D. Le, J. I. Lim, R. V. Chan, and X. Yao, "Supervised machine learning based multi-task artificial intelligence classification of retinopathies," *Journal of Clinical Medicine*, vol. 8, no. 6, p. 872, 2019.

[16] L. Math and R. Fatima, "Adaptive machine learning classification for diabetic retinopathy," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5173–5186, 2021.

[17] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2693–2701, 2015.

[18] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Computers in Industry*, vol. 69, pp. 72–80, 2015.

[19] J.-J. Yang, J. Li, R. Shen, Y. Zeng, J. He, J. Bi, Y. Li, Q. Zhang, L. Peng, and Q. Wang, "Exploiting ensemble learning for automatic cataract detection and grading," *Computer methods and programs in biomedicine*, vol. 124, pp. 45–57, 2016.

[20] T. Pratap and P. Kokil, "Computer-aided diagnosis of cataract using deep transfer learning," *Biomedical Signal Processing and Control*, vol. 53, p. 101533, 2019.

[21] M. S. Junayed, M. B. Islam, A. Sadeghzadeh, and S. Rahman, "Cataractnet: An automated cataract detection system using deep learning for fundus images," *IEEE Access*, vol. 9, pp. 128 799–128 808, 2021.

[22] K. Ananthapadmanaban and G. Parthiban, "Prediction of chances-diabetic retinopathy using data mining classification techniques," *Indian Journal of Science and Technology*, vol. 7, no. 10, pp. 1498–1503, 2014.

[23] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[24] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *Jama*, vol. 318, no. 22, pp. 2211–2223, 2017.

[25] M. Voets, K. Møllersen, and L. A. Bongo, "Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *PloS One*, vol. 14, no. 6, p. e0217541, 2019.

[26] S. Keel, J. Wu, P. Y. Lee, J. Scheetz, and M. He, "Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma," *JAMA ophthalmology*, vol. 137, no. 3, pp. 288–292, 2019.

[27] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150 530–150 539, 2019.

[28] L. Xie, S. Yang, D. Squirrell, and E. Vaghefi, "Towards implementation of ai in new zealand national diabetic screening program: Cloud-based, robust, and bespoke," *Plos One*, vol. 15, no. 4, p. e0225015, 2020.

[29] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, "Glaucoma detection based on deep convolutional neural network," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 715–718.

[30] M. N. Bajwa, M. I. Malik, S. A. Siddiqui, A. Dengel, F. Shafait, W. Neumeier, and S. Ahmed, "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–16, 2019.

[31] R. Hemelings, B. Elen, J. Barbosa-Breda, S. Lemmens, M. Meire, S. Pourjavan, E. Vandewalle, S. Van de Veire, M. B. Blaschko, P. De Boever, and I. Stalmans, "Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning," *Acta Ophthalmologica*, vol. 98, no. 1, pp. e94–e100, 2020.

[32] Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images," *Applied Soft Computing*, vol. 116, p. 108347, 2022.

[33] S. Zhang, R. Zheng, Y. Luo, X. Wang, J. Mao, C. J. Roberts, and M. Sun, "Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network," *IEEE Access*, vol. 7, pp. 57 561–57 573, 2019.

[34] X. Xu, W. Ding, M. D. Abràmoff, and R. Cao, "An improved arteriovenous classification method for the early diagnostics of various diseases in retinal image," *Computer methods and programs in biomedicine*, vol. 141, pp. 3–9, 2017.

[35] E. Deepika and S. Maheswari, "Earlier glaucoma detection using blood vessel segmentation and classification," *Proceedings of the 2nd International Conference on Inventive Systems and Control (ICISC)*, pp. 484–490, 2018.

[36] D. Maji and A. A. Sekh, "Automatic grading of retinal blood vessel in deep retinal image diagnosis," *Journal of Medical Systems*, vol. 44, no. 10, pp. 1–14, 2020.

[37] B. Remeseiro, A. M. Mendonça, and A. Campilho, "Automatic classification of retinal blood vessels based on multilevel thresholding and graph propagation," *The Visual Computer*, vol. 37, no. 6, pp. 1247–1261, 2021.
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representation*, 2015, pp. 1–14.
[39] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
[40] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013. [Online]. Available: https://arxiv.org/abs/1312.4400
[41] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.
[42] M. Hussain, J. J. Bird, and D. R. Faria, "A study on cnn transfer learning for image classification," in *UK Workshop on computational Intelligence*. Springer, 2018, pp. 191–202.
[43] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.
[44] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," *Advances in neural information processing systems*, vol. 32, 2019.
[45] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks," *IEEE Access*, vol. 7, pp. 3360–3370, 2018.
[46] J. Ran, K. Niu, Z. He, H. Zhang, and H. Song, "Cataract detection and grading based on combination of deep convolutional neural network and random forests," in *2018 international conference on network infrastructure and digital content (IC-NIDC)*. IEEE, 2018, pp. 155–159.
[47] F. Li, Z. Wang, G. Qu, D. Song, Y. Yuan, Y. Xu, K. Gao, G. Luo, Z. Xiao, D. S. Lam *et al.*, "Automatic differentiation of glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network," *BMC medical imaging*, vol. 18, no. 1, pp. 1–7, 2018.
[48] S. AlBadawi and M. Fraz, "Arterioles and venules classification in retinal images using fully convolutional deep neural network," in *International conference image analysis and recognition*. Springer, 2018, pp. 659–668.
[49] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.
[50] Kaggle, "Diabetic retinopathy detection," Tech. Rep., 2015, https://www.kaggle.com/c/diabetic-retinopathy-detection.
[51] ——, "Ocular disease intelligent recognition (odir)," Tech. Rep., 2020, https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.
[52] J. I. Orlando, H. Fu, J. B. Breda, K. van Keer, D. R. Bathula, A. Diaz-Pinto, R. Fang, P.-A. Heng, J. Kim, J. Lee *et al.*, "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical image analysis*, vol. 59, p. 101570, 2020.
[53] J.-H. Wu, T. A. Liu, W.-T. Hsu, J. H.-C. Ho, C.-C. Lee *et al.*, "Performance and limitation of machine learning algorithms for diabetic retinopathy screening: Meta-analysis," *Journal of Medical Internet Research*, vol. 23, no. 7, p. e23863, 2021.
[54] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. T. Salles, "Diabetic retinopathy detection using red lesion localization and convolutional neural networks," *Computers in biology and medicine*, vol. 116, p. 103537, 2020.
[55] M. S. M. Khan, M. Ahmed, R. Z. Rasel, and M. M. Khan, "Cataract detection using convolutional neural network with vgg-19 model," in *IEEE World AI IoT Congress (AIIoT)*, 2021, pp. 0209–0212.
[56] H. H. Ali, A. Y. Al-Sultan, and E. H. Al-Saadi, "Cataract disease detection used deep convolution neural network," in *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, 2022, pp. 102–108.

GABRIEL D. A. ARANHA was born in Lins, Brazil in 1992. He did his B.Sc. in Information Systems at the Centro Universitário de Lins - UNILINS, Lins, Brazil in 2013, and his M.Sc. degree in Computer Science at the Federal University of São Carlos (UFSCar), São Carlos, Brazil in 2016. Currently, he is working towards a Ph.D. in Computer Science at UFSCar. His research interests include feature engineering and machine learning (mainly deep learning and transfer learning) as well as its applications as assistive technologies.

RICARDO A. S. FERNANDES (M'10) was born in Barretos, Brazil in 1984. He did his B.Sc in Electrical Engineering at the Educational Foundation of Barretos, Barretos, Brazil in 2006, and his M.Sc. and Ph.D. degrees in Electrical Engineering at the University of São Paulo, São Carlos, Brazil in 2009 and 2011, respectively. In 2015 and 2017, he was a Visiting Professor at the Polytechnic Institute of Porto. Currently, he is an Associate Professor at the Federal University of São Carlos, São Carlos Brazil. His research interests include signal processing, feature engineering, and machine learning.