J. Vis. Commun. Image R. 64 (2019) 102597

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

An hybrid feature space from texture information and transfer learning for glaucoma classification

Maíla Claro [a,*], Rodrigo Veras [a], André Santana [a], Flávio Araújo [c], Romuere Silva [c], João Almeida [b], Daniel Leite [a]

[a] Universidade Federal do Piauí, Teresina, PI, Brazil
[b] Universidade Federal do Maranhão, São Luís, MA, Brazil
[c] Universidade Federal do Piauí, Picos, PI, Brazil

ARTICLE INFO

Article history:
Received 2 May 2018
Revised 5 May 2019
Accepted 31 July 2019
Available online 16 August 2019

Keywords:
Glaucoma detection
Feature selection
Pre-trained CNNs
Transfer learning

ABSTRACT

Glaucoma is a progressive eye disease due to the increase in intraocular pressure. Accurate early detection may prevent vision loss. Most algorithms in the literature are not feasible for use in screening programs since they are not able to handle a wide diversity of images. We conducted an extensive study to determine the best set of features for image representation. Our feature extraction methodology included the following descriptors: LBP, GLCM, HOG, Tamura, GLRLM, morphology, and seven CNN architectures, that results in 30.682 features. Then, we used the gain ratio to order the features by importance and select the best set for glaucoma classification. Our tests were performed using 1675 images of DRISHTI, RIM-ONE, HRF, JSIEC, and ACRIMA databases. We concluded that a combination of the GLCM and pre-trained CNN's has the potential to be used in a computer aid system for glaucoma detection. Our approach achieved an accuracy of 93.61%.

1. Introduction

The diagnosis of diseases using computational approaches has become the focus of several types of research. These studies aim to support specialists in identifying pathologies, and possibly to prevent their development once early diagnosis has been achieved.

In ophthalmology, a large number of computer-aided diagnosis (CAD) systems have been developed to assist the detection of various types of ocular diseases, including glaucoma [1–5]. These systems have the potential to provide an alternative solution to mass screening programs, which need to examine a large number of fundus images efficiently and robustly.

The initial stages of glaucoma tend to be asymptomatic, and its clinical manifestations appear only in the more advanced stages of the disease. The early diagnosis of this pathology is therefore essential, as there are treatments to prevent its progression and consequently to prevent blindness. According to the World Health Organisation [6], there are around 60 million people with glaucoma worldwide, with 2.4 million more cases each year. One estimation suggests that by 2020, about 80 million people will have glaucoma.

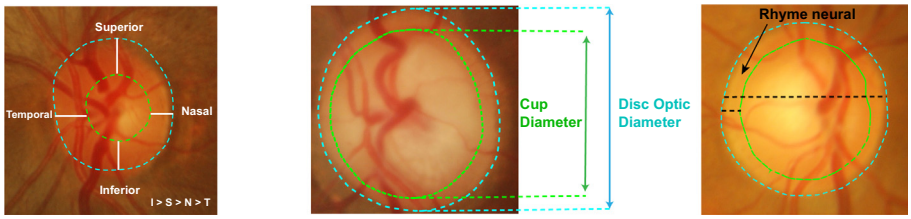
Specialists analyze retinal images to diagnose glaucoma. This determination is possible because the amount of fiber loss from the optic nerve has a direct effect on the neural rhyme configuration. As the nerve fibers die, the cup becomes wider than the optic disc (OD), which change the ISNT pattern and increases the values of the cup-to-disc ratio (CDR) and Disc Damage Likelihood Scale (DDLS) [7], as shown in Fig. 1. The highlighted areas represent the OD region (represented by the entire area within the white marking), and the cup region (the whole area within the green marking).

Although ISNT, CDR and DDLS analysis are the most widely used clinical method, it is difficult to replicate in automatic systems. According to Hallem et al. [8], the edge between the OD and optic cup is not usually visible in retinal images with normal contrast. There is therefore a need to determine certain parameters that can distinguish between these regions. The extraction of the optic cup may also require segmentation of the vascular structure and its in-painting with the neighbourhood for accurate dimension measurement.

In this scenario, deep learning is an emerging technology with several applications in Ophthalmology. The deep learning methods have demonstrated utility in the assessment of various disease processes including cataracts, glaucoma, age-related macular degeneration, and diabetic retinopathy. They have been applied

* Corresponding author.
E-mail address: claromaila@gmail.com (M. Claro).

https://doi.org/10.1016/j.jvcir.2019.102597
1047-3203/



(a) ISNT (b) CDR (c) DDLS

Fig. 1. Examples of findings used to detect glaucoma in retinal images: (a) the rhyme neural thickness is usually higher in the inferior region, followed by the upper, nasal and temporal regions (inferior > upper > nasal > temporal); (b) The CDR is a relation that compares the cup diameter with the OD diameter; (c) DDLS calculates the ratio between the shortest rhyme neural thickness and OD diameter.

to different diagnostic modalities including digital images and optical coherence tomography [9].

The literature contains several algorithms for the processing and analysis of retinal images for the automatic diagnosis of glaucoma. However, researchers typically evaluate their proposals using a restricted set of images that are often private, and use strategies that are dependent on the characteristics of the databases in which they are tested.

The goal of this paper is to perform the automatic detection of glaucoma from an extensive diversity of retinal images created by combining several different databases. Although we propose a complete system, the main focus of this paper is on describing the image. In other words, we aim to define the best set of features that can be used to differentiate between healthy and pathological images.

To develop the proposed method, we carried out numerous experiments using texture and shape descriptors, and pre-trained convolutional neural networks (CNNs) [10] as input for random forest (RF) [11], support vector machine (SVM) [12] and multilayer perceptron (MLP) [13] algorithms. Based on these experiments, we achieved higher performance metrics by combining feature vectors, and this paper therefore proposes a set of features for the automatic detection of glaucoma.

The present work is organised as follows: Section 2 discusses related works on the identification of glaucoma; Section 3 describes the image databases, the descriptors tested, the attribute selection step, the classifiers used and the evaluation metrics; Section 4 presents the obtained results; Section 5 contains a discussion of these results; and Section 6 presents our conclusions and next steps.

2. Related work

We carried out a literature review to identify the main techniques used in the diagnosis of glaucoma based on images.

The work of Acharya et al. [1] presented a combination of texture descriptors and higher order spectra (HOS) in 60 digital retinal images for feature extraction. The images were collected from Kasturba Medical College in Manipal, India; 30 of these were normal and 30 were glaucomatous. Of the classifiers evaluated in this work, RF [11] achieved the best result, obtaining an accuracy of the order of 91%.

Simonthomas et al. [2] proposed a method for glaucoma detection using a Haralick texture descriptor and the k-nearest neighbours (KNN) classifier. The set of images used were the same as those used by Acharya et al. [1]. The authors applied the grey level co-occurrence matrix (GLCM) [14] to the image, and Haralick's 13 features were combined to achieve better performance, yielding an accuracy rate of 98%.

Srinivasan et al. [3] applied two texture descriptor techniques, the GLCM and local binary pattern (LBP) [15] to a database containing 200 retinal images. The authors used OD segmentation to extract these characteristics and an SVM to classify the images. The system proposed in this work showed promising results, with a sensitivity of 100%, a specificity of 99% and an accuracy of 99%.

In Kotyk et al. [4], the authors proposed a semi-automated system to detect both the OD and the cup, and these regions were then used to calculate the CDR. For this, they used a thresholding process for the red channel of the RGB colour model in the retinal images of the RIM-ONE-1 base. The images were further processed using GLCM [14], and the obtained features were used in the MLP classifier to determine the accuracy of the system. The accuracy of the proposed method was of the order of 86.43%.

Salam et al. [16] combined structural descriptors (CDRs) with non-structural descriptors (texture and intensity) to give better accuracy in automated diagnosis. This algorithm used 100 images acquired from a local database. The images were preprocessed, and the OD region and its cup were extracted for use in the CDR calculation. The texture descriptors used were wavelets, multi-wavelets, LBP, GLCM and Haralick characteristics. Colour moments and an autocorrelogram were used in the description of intensity. The authors applied these descriptors to training and testing images for the SVM classifier, and the mean sensitivity, specificity and accuracy of the proposed system were 100%, 87% and 91%, respectively.

Claro et al. [17] applied the GLCM and the entropy to extract texture information from retinal images for glaucoma detection. They used the red, green, blue (RGB) channels, hue, saturation, intensity (HSI) and the luminosity, chromaticity U and chromaticity V (LUV). The database used was RIM-ONE-1, and their method achieved an accuracy of 93.03% by applying the MLP classifier.

In [18] the authors observed that features cited in paragraphs above are strongly related to expert knowledge and they have restricted representation power. Thus, for a huge dataset it cannot show the discriminative power. To overcome this problem and enhance the classification performance, many authors applied deep learning techniques for glaucoma diagnosis.

In the literature, we found three different approaches that use deep learning to classify medical images [10]: the proposal of new CNN architectures, the use of pre-trained CNNs for feature extraction, and the fine-tuning of pre-trained CNNs.

In the papers [19–21] the authors proposed new CNN architectures for glaucoma diagnosis. In [19], Chen et al. proposed a CNN architecture with six layers: four convolutional and two fully-connected layers. The experiments were performed on the ORIGA [22] and SCES [23] datasets. The results showed area under ROC curve (AUC) at 0.831 and 0.887 in the two databases.

Ahn et al. [20] proposed a CNN with three convolutional layers and max pooling applied at each layer, following by two fully connected layers. The authors used a private dataset of 1,542 images (754 for training, 324 for validation and 464 for the test). They achieved accuracy and AUC of 87.9% and 0.94 on the test data.

Raghavendra et al. [21] presented a CNN with four convolutional layers, and batch normalization, one ReLU, and one Max-pooling applied at each layer, following by a fully connected and a Soft-max layers. The authors achieved accuracy of 98.13% using

1426 (589: normal and 837: glaucoma) fundus images of a private dataset.

Despite the promising results, the definition of new CNNs presents significant challenges because it is necessary a large computational power and a huge database to define the quantity, size, and configuration of layers. Therefore, the papers [5,24,25], stand out among those using pre-trained CNNs to the extraction of features. These features are known as deep features.

In Orlando et al. [5], the Vgg-S [26] and Overfeat [27] models were applied to images from the DRISHTI database [28] to generate the feature vectors. These authors also used preprocessing techniques to improve the image quality before the feature extraction step, such as histogram equalization, vessel painting and a cut in the region of the optic nerve head. The AUC of 0.78 was achieved with the logistic regression (LR) classifier [29].

Al-Bander et al. [24] used the AlexNet architecture [30] for feature extraction and the SVM to classify 455 (70% for training and 30% for test) images of the RIM-ONE-2 database. The authors reported accuracy of 88.2%, a specificity of 90.8% and a sensitivity of 85%. Similarly, Lima et al. [25] used the ResNet network [31] for feature extraction and the Logistic Regression to classify the images of the RIM-ONE-2 database and reported an AUC of 0.957. The test in 159 images of RIM-ONE-3 database, demonstrated that the InceptionResNet [32] was the best architecture to the extraction of features, with an AUC of 0.86.

The use of deep features has been presented with a viable alternative to the use of characteristics that depend on the knowledge of specialists. However, as the CNNs used have been trained with general purpose images; they are not able to extract specific features of the problem. In this way, some authors [33–39] chose to refine the weights of pre-trained networks with specific images.

The methodologies in [34,33] were divided into two stages, namely: detection of the region of interest (ROI); and image classification. In both approaches, the detection of the ROI in the first stage is used as the input for the second-stage. The authors trained one CNN for each stage. Cerentini et al. [34] used two GoogLeNet [32] in 772 images of four public databases and achieved an average accuracy of 87.6%. Jun et al. [33] used two DenseNet [40] in 1022 fundus images and achieved an accuracy of 96.46%.

In [35–37] the authors evaluated some pre-trainded CNNs and fine tuned the CNN ResNet [31]. In [35] the authors achieved an AUC of 0.965 using a training dataset containing 3132 images (1,364 with glaucoma and 1,768 without glaucoma) and a testing dataset containing 110 images (60 with glaucoma and 50 without glaucoma). Christopher et al. [36] related an AUC of 0.91 in a large dataset of 14,822 images. Liu et al. [37] achieved 86.7% of sensitivity and specificity. They used 4,364 images of different databases to fine-tuning the CNN and tested it in 30 images from HRF database.

Gómez-Valverde [38] achieved an AUC of 0.94 evaluating four CNNs and fine-tuning the VGG 19. They used three different datasets with 2313 images. Diaz-Pinto [39] evaluated five pre-trained CNNs in five public databases. The fine-tuned Xception [41] achieved an average AUC of 0.9605. Otherwise, when the authors assessed the databases separately, the AUC varied from 0.76 to 0.85 showing the dependence of the training data.

To define the best results, the presented papers evaluated different CNNs. Many of these are repeated, for example GooGleNet [32] was evaluated in [34,38], VGG 19 was evaluated in [38,39], ResNet50 [31] was evaluated in [37,36,38,39]. In this way, we conclude that the refinement of the network is extremely dependent on the set of images used (quantity and diversity of images). Thus, although promising, the use of this strategy does not guarantee good hit rates for images from other databases.

3. Materials and methods

The goal of this work is enable the diagnosis of glaucoma based on retinal images. We performed experiments using texture and shape descriptors and pre-trained CNNs on the OD region. We conducted an evaluation of the potential separate description of each of these descriptors, and concluded that a combination of different characteristics would lead to better classification rates for glaucoma. We therefore formed a new image representation by using the set of features that produced the best results.

Segmentation of the OD region is used to analyze features to find the differences between a pathological and non-pathological retina. The main features analyzed here are the thickness of the vessels, the cup region and the homogeneity or heterogeneity in terms of colour between the areas of the OD and the cup, among other characteristics.

3.1. Proposed method

The proposed method follows the flowchart shown in Fig. 2. This methodology starts with a segmented OD image following the extraction of the characteristics of texture and shape using CNNs. From an analysis of the results, we observed that a concatenation of the GLCM features with the last fully connected layer in the CNN-CaffeNet [42], CNN-Vgg-f, CNNVgg-m, CNN-Vgg-s [26] models produced the best results. We applied the gain ratio algorithm to rank the features from the most significant to the least significant, and made an feature selection. Finally the RF classifier was applied to determine whether the image was healthy or glaucomatous.

3.2. Feature extraction

The aim of the feature extraction step is to describe images in terms of the extracted attributes. These characteristics are used to recognise certain pattern, and specific properties are employed to make the decision process easier during classification. According to Salam et al. [16], the identification of glaucomatous images can be efficiently performed using texture and shape descriptors. Thus, in this study, the obtained results were evaluated using the local binary pattern (LBP) [15],GLCM, histogram of oriented gradients (HOG) [43], Tamura [44], grey level run length matrix (GLRLM) [45–47], and shape descriptors. We also included CNNs in our analysis.

We extracted the texture feature using five different colour channels, converting the images to the hue, saturation, value (HSV) colour models and RGB patterns. The V colour band of the HSV model was discarded, since in the retinal images the colour band is almost the same as the red colour channel [48]. In the following paragraphs, we present an overview of the descriptors used in the experiments.

Local Binary Pattern (LBP): In the LBP texture descriptor [15], the pixel labelling of a greyscale image is performed using thresholding. Each pixel is compared to the centre of a 3x3 neighbourhood window, which constitutes the threshold value. When all windows are computed, the attribute vector (with 256 attributes) is assembled by merging the obtained values. Thus, the histogram of the cell is calculated using values of zero and one.

Gray Level Co-occurrence Matrix (GLCM): We can define several characteristics based on the GLCM [14], but the most important ones used were: contrast, correlation, energy, homogeneity and entropy.

Histogram of Oriented Gradients (HOG): The HOG [43] calculates the histogram of the gradient orientations in the image. It

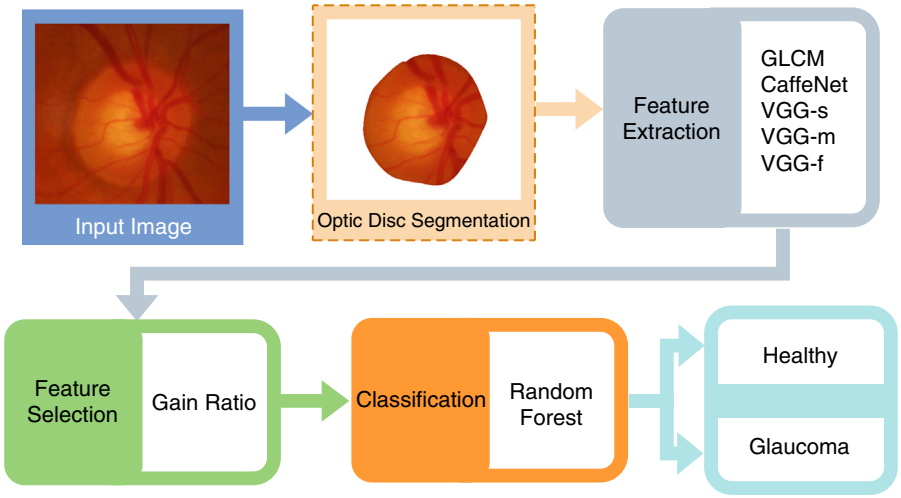


Fig. 2. Flowchart for the proposed method of automatic diagnosis of glaucoma.

converts the image to greyscale and normalises it based on illumination. The gradients are computed and the image is divided into small spatial regions called cells.

A local 1D histogram of the orientations of the pixels is calculated for each cell. After the histograms are computed, they are normalised based on an accumulation of local histograms into slightly larger spatial regions called blocks. The accumulated histograms are then used to normalise all the cells in a given block. After normalisation, a detection window is computed for the generated histograms that consists of the output of the HOG descriptor with 81 attributes.

Tamura: [44], this texture descriptor calculates the three characteristics of coarseness, contrast and directionality for the image.

Gray Level Run Length Matrix (GLRLM): The GLRLM (also known as the grey level primitive length method) is a texture descriptor that uses auxiliary matrices and checks the repetition of the pixels (primitive). A GLRLM matrix is created for each cut, and contains the number of grey level primitives in the θ (0°, 45°, 90° or 135°) and at a distance d. Thus, this descriptor returns 44 features. The sum of all GLRLM matrices of each cut is calculated, and based on the number of replicates found, the matrix is divided and normalised.

According to Galloway [45], five statistical measures can be extracted from the GLRLM matrix: short-run emphasis, long-run emphasis, grey level distribution, run length distribution and run percentage. Two further statistical measures were determined by Chu et al. [46]: the low and high grey level run emphasis.

Following the idea of statistical measures, four other feature extraction functions were described by Dasarathy and Holder [47]. These are the short-run low grey level emphasis primitives, short-run high grey level emphasis primitives, long run low grey level emphasis, and long run high grey-level emphasis.

Morphology: The most important characteristics are area, perimeter, convex area, equivalent diameter and extension. Although several other morphological features were tested and analysed, we observed after selection with the gain ratio algorithm [49] that the above characteristics were the most effective for application to retinal images.

Convolutional Neural Networks: CNNs have now been established as the leading state-of-the-art technique in computer vision systems. They have been successfully used in various different applications [50,51] such as object recognition [52] and image segmentation for medical diagnosis [53].

In recent works, researchers have presented two different ways of using the power of CNNs. The first is the usual method in which training is performed with a large image database. The second involves transfer learning using pre-trained networks [35]. In our work, we used pre-trained CNNs in a large natural image database. In this way, the neural network can assimilate generic features to facilitate its application in small databases. One way to perform the transfer learning is to use the last fully connected layer of the network as a feature vector. This layer typically contains 4,096 attributes. This technique can be used for several types of tasks, for example in the extraction of features from images of faces, objects, and diseases. In this work, we tested seven recent architectures available from the MatConvNet website to build our approach: AlexNet [30], CaffeNet [42], Vgg-f, Vgg-m, Vgg-s [26], Vgg-16 and Vgg-19 [54].

Table 1 gives a summary of the 30,682 features evaluated in this paper.

3.3. Feature selection and classification

After creating the characteristics vector, we performed a feature selection. This aimed to eliminate unnecessary features and consequently to simplify the prediction model to improve the performance of the algorithms, reduce the computational cost of these models and provide a better understanding of the obtained results.

The algorithm used for attribute selection was the gain ratio, which is a metric of information gain. This algorithm is a type of filter that tends to overestimate the quality of attributes with many values [49]. In this context, 10 tests were performed using 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% and 90% of the characteristic vector size, allowing us to exclude the least significant attributes.

The software used in this classification was the Waikato Environment for Knowledge Analysis (WEKA) [55]. The classifiers tested in the experiments in this work were the SVM [12], MLP [13] and random forest [11]. The parameters used for each classier were the standard ones in WEKA and the validation method used was k-fold cross validation (with k = 10).

3.4. Evaluation metrics

To analyse the classification results, we computed the confusion matrix. From these matrix values we calculate the precision (P), AUC and accuracy (A) [56].

We also computed the Kappa index (K), which is recommended as an appropriate measure of exactitude as it can fully represent the confusion matrix; it takes all elements of the matrix into

Table 1
Summary of the descriptors used in this paper.

Descriptor	Features	N. of Features
LBP [15]	Texture histogram	1280
GLCM [14]	Contrast, Correlation, Energy, Homogeneity (0°, 45°, 90° e 135°) and Entropy	85
HOG [43]	Histogram of the gradients orientation	405
Tamura [44]	Contrast, Directionality and Coarseness	15
GLRLM [45–47]	Short-Run Emphasis, Long-Run Emphasis, Gray Level distribuition, Run Lenght distribuition, Run Percentage, Low Gray Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, and Long Run High Gray Level Emphasis (0°, 45°, 90° e 135°)	220
Morphology	Area, Perimeter, Convex Area, Diamenter and Extension	5
CNNs [30,42,26,54].	Transfer Learning	7 × 4096
Total	Texture, Morphology and Transfer Learning	30,682

account, rather than just those on the main diagonal, which occurs when calculating the global accuracy of the classification [57]. This metric can be calculated based on Eq. 1.

$$K = \frac{(observed - expected)}{1 - expected}. \quad (1)$$

In this case, "observed" indicates the total of images correctly classified, in other words, the sum of the main diagonal divided by the number of elements; and "expected" indicates the values calculated based on the totals of each row and each column of the confusion matrix. The quality of the results can be evaluated by the Kappa index value, as observed in Table 2, and as defined by Landis and Koch [58].

3.5. Image databases

One of the challenges in the automatic detection of glaucoma is the ability to diagnose the disease in databases with distinct characteristics, such as images captured by various professionals and with different cameras, patients from different ethnic groups (for example, hospitals in India (DRISTH), Spain (RIM-ONE), Germany (HRF), and China (JSIEC)), different resolutions and diagnoses provided by different specialists. However, several authors have tested their systems with only one database [1–3,5,4,16,17,33]. In the present work, we separate the databases into development and performance sets.

The development set was made up of the DRISHTI [28] database and three versions of RIM-ONE [59]. We used this set for parameter tuning.

The DRISHTI database is composed of 101 samples collected at the *Aravind Eye Hospital*, India. All images were taken centred on the OD with a field of view (FOV) of 30° and dimensions of 2896 × 1944 pixels in the PNG uncompressed image format. The ground truth was collected from data experts with 3, 5, 9 and 20 years of clinical experience.

The images in the RIM-ONE 1, 2 and 3 databases were recorded in three different regions of Spain. RIM-ONE-1 consists of 158 retinal images, of which 118 are healthy, 12 contain primary glaucoma, 14 have moderate glaucoma and 14 advanced glaucoma. A Nidek AFC-210 camera was used to capture these images. The database was created with the collaboration of four ophthalmologists and an optometrist.

The RIM-ONE-2 database contains 455 retinal images (255 healthy and 200 glaucomatous). All images have the ground truth for the OD segmentation and their size varies from 345 × 386 to 828 × 818 pixels.

The RIM-ONE-3 database contains 159 retinal images (85 healthy and 74 glaucomatous). All images were taken with a Kowa WX 3D non-stereo camera. The images were centred in the OD with a FOV of 34°, and the stereoscopic images were captured in the same shot, giving a resolution of 2144 × 1424 pixels. The diagnosis for each image in the database was made by two specialists.

Table 2
Level of classification hit, based on the Kappa index.

Kappa Index (K)	Quality
$K \leqslant 20$	Poor
$20 < K \leqslant 40$	Moderate
$40 < K \leqslant 60$	Good
$60 < K \leqslant 80$	Very Good
$K > 80$	Excellent

Table 3
Number of images in the development set.

Database	Healthy	Non-pathological	Total
DRISHTI	31	70	101
RIM-ONE-1	118	40	158
RIM-ONE-2	255	200	455
RIM-ONE-3	85	74	159
Total	489	384	873

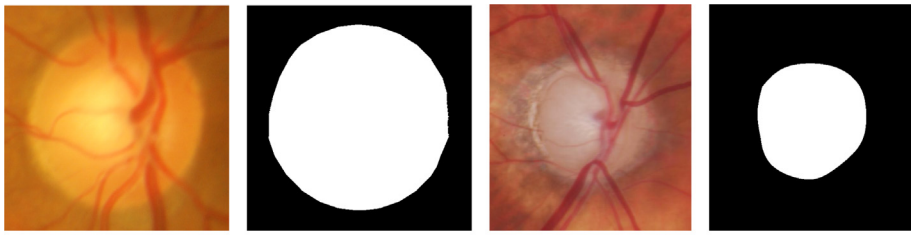
The ground truth of the OD segmentation of all images in the development set was provided by the database. Table 3 presents the number of images in the development set, and Fig. 3 shows samples of these images and the respective ground truth.

To demonstrate that the proposed feature vector is capable to be used in a diversity of images, it was applied to the images in the performance set formed from the HRF [60], JSIEC[1] and ACRIMA [39] databases.

The HRF database [60] was composed of 45 fundus images, which were divided into three groups of 15 images. The first one represented glaucoma, while the second group showed healthy eyes and the last one, diabetic retinopathy. For the performance evaluation, only the groups showing glaucoma and normal images were selected. The images were captured in the Department of Ophthalmology, Friedrich-Alexander University, Erlangen-Nuremberg (Germany). All images have dimensions of 3504 × 2336 pixels.

The JSIEC database contains 1087 high-quality fundus images divided into 37 categories. The images were captured at the Joint Shantou International Eye Centre in China and have varying dimensions. Of the 37 categories, we used 2. The first one has 54 images categorised as normal, and the second one is composed by 13 possible glaucoma images.

[1] Available in https://www.kaggle.com/linchundan.



(a) (b) (c) (d)

Fig. 3. Samples of: (a) image of the DRISHTI database, (b) ground truth of the DRISHTI database, (c) image of the RIM-ONE-1 database and (d) ground truth of the RIM-ONE-1 database.

Table 4
Number of images in the performance set.

Database	Healthy	Glaucomatous	Total
HRF	15	15	30
JSIEC	54	13	67
ACRIMA	309	396	705
Total	378	424	802

The ACRIMA database [39] is composed of 705 fundus images, including 396 glaucomatous and 309 normal images. The project is founded by the Ministerio de Economía y Competitividad of Spain. They were captured using the Topcon TRC retinal with a field of view of 35°.

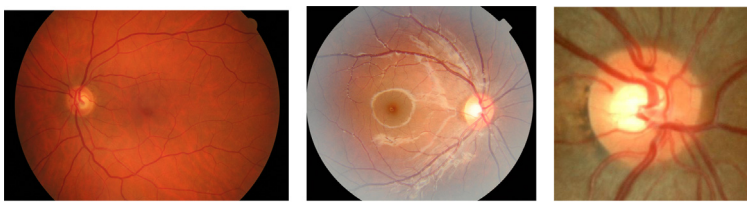
The ground truth for the OD segmentation in the performance test images was not available. Therefore, an ophthalmologist from our team marked the OD boundary. Table 4 presents the number of images in the performance set, and Fig. 4 shows samples of these images.

4. Results

The parameters used for classification were the standard ones for each classifier in WEKA [55], and the validation method used was k-fold cross-validation (with k = 10). Figs. 5 and 6 presents the Accuracy and Kappa results obtained for each set of features and each classifier. For these tests, we used the 873 images in the development set.

Fig. 5 shows that in all three classifiers tested, the GLCM and CNNs obtained a higher accuracy than the other descriptors, achieving an accuracy higher than 80%. The VGG-19 descriptor generated the best result for the SVM classifier, with an accuracy of 82.70% and a Kappa value rated as very good. GLCM generated the best results for the MLP and RF classifiers, with accuracies of 84.99% and 83.16%.

Comparing the results presented in Fig. 6, we can see that the random forest algorithm achieved most results with a Kappa of above 60%. We can infer that random forest performed better than



(a) (b) (c)

Fig. 4. Samples of: (a) an image from the HRF database; (b) image from the JSIEC database; and (c) image from the ACRIMA database.

	Accuracy (%)		
Descriptor	SVM	MLP	RF
LBP	56.01%	74.45%	74.45%
GLCM	80.18%	84.99%	83.16%
HOG	78.00%	72.73%	75.02%
Tamura	69.64%	70.33%	70.90%
GLRLM	63.68%	74.34%	75.71%
Morphology	52.11%	63.00%	66.32%
CNN-AlexNet	82.47%	83.50%	80.29%
CNN-CaffeNet	80.87%	81.55%	81.55%
CNN-Vgg-f	80.75%	79.95%	82.13%
CNN-Vgg-m	82.13%	83.27%	81.78%
CNN-Vgg-s	80.87%	81.67%	82.60%
CNN-Vgg-16	78.23%	80.06%	80.87%
CNN-Vgg-19	82.70%	81.44%	80.98%

Best

Worst

Fig. 5. Accuracy obtained for each set of features and for each classifier in the development set of images.

	Kappa (%)		
Descriptor	SVM	MLP	Random Forest
LBP	0.00%	48.12%	47.91%
GLCM	60.57%	69.69%	65.70%
HOG	55.42%	44.90%	49.24%
Tamura	37.54%	39.98%	40.19%
GLRLM	25.67%	47.70%	50.44%
Morphology	5.78%	22.60%	31.27%
CNN-AlexNet	64.30%	66.49%	59.77%
CNN-CaffeNet	60.91%	62.40%	62.18%
CNN-Vgg-f	60.75%	59.59%	63.33%
CNN-Vgg-m	63.68%	66.04%	62.97%
CNN-Vgg-s	61.06%	62.72%	64.63%
CNN-Vgg-16	55.78%	59.51%	60.77%
CNN-Vgg-19	64.89%	62.30%	61.11%

Best

Worst

Fig. 6. Kappa obtained for each set of features and for each classifier in the development set of images.

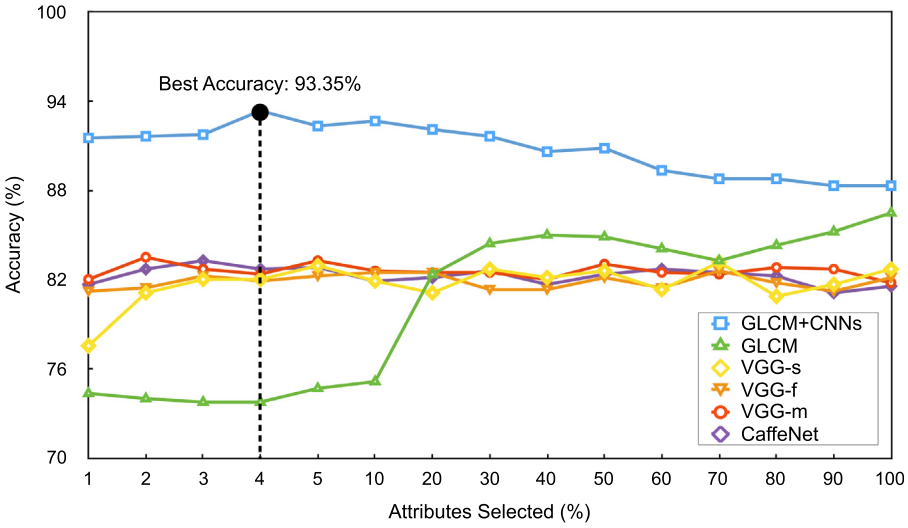


Fig. 7. Attribute selection results using random forest and GLCM, CNN-CaffeNet, CNN-VGGFNet, CNN-VGGMNet, CNN-VGGSNet and GLCM + CNNs.

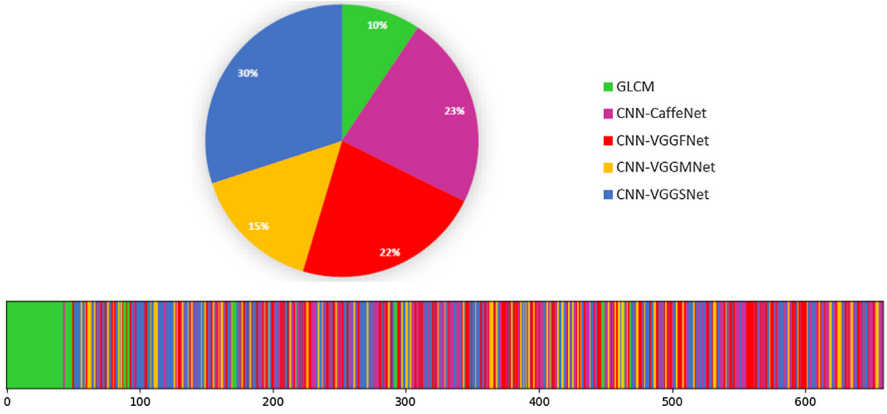


Fig. 8. Composition of the proposed descriptor.

the other classifiers. To set up our feature vector, we decided to use the five descriptors that achieved the best result with the random forest classier, and therefore concatenated the descriptors of GLCM, CNN-CaffeNet, CNN-Vgg-f, CNN-Vgg-m, and CNN-Vgg-s.

We performed tests to identify the best subset of features. Thus, we used feature vectors with 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100% of the most significant attributes, giving a total of 90 tests (15 feature vectors of different lengths

multiplied by six descriptors). Fig. 7 shows all the results obtained using the random forest approach, and Table 5 shows the numbers of attributes that gave the best results.

The GLCM and CNN descriptors obtained accuracy rates of between 73.76% and 86.48%. These descriptors achieved a Kappa index that was rated as very good. However, by concatenating them, the result was significantly improved, reaching an accuracy of 93.35% with only 4% of the attributes, i.e. 659 attributes from a possible 16,469. The Kappa index was 86.48%, which is rated as excellent.

Fig. 8 shows the features percentage chosen from each descriptor as well as their distributions along the positions of the vector sorted by the gain ratio algorithm in GLCM + CNNs feature vector. It is notorious that the GLCM descriptor, which has only 85 attributes, obtained a significant rate on the essential characteristics subset.

The results presented in the last line of Table 5 were obtained using a heuristic search of the best parameters. To demonstrate that the proposed method works well in images of various databases, we applied it to the HRF, JSIEC and ACRIMA databases with the parameters defined previously. Table 6 presents the results of this evaluation. We applied the same validation method used for the development set (*k*-fold cross validation).

We can see from Table 6 that the method achieved accuracy, precision and AUC rates above than 90%. The value for the Kappa index is also rated as excellent. We can therefore conclude that the proposed feature set can be used for a wide range of images and databases.

Optimal segmentation is still a challenge in the development of CAD systems. In order to analyse the performance of the proposed method even when there are segmentation errors, we performed tests in two scenarios, under-segmentation and over-segmentation, in a similar analysis to that carried out in [61]. The low contrast between the OD and the background, with a similar colour level and low variations in intensity, can cause these types of segmentation errors.

Fig. 9 illustrates how these segmentation errors were simulated. In Fig. 9(a), the ground truth is indicated by the white contour. To simulate the under-segmentation error, we decreased the ground truth area by 10%, and the result can be seen in Fig. 9(b), indicated by the red contour. In the over-segmentation error, the ground truth area was increased by 10% and the result can be seen in Fig. 9(c), indicated by the green contour.

Tables 7 and 8 present the results obtained with segmentation errors (under- and over-segmentation) using the proposed GLCM + CNN approach.

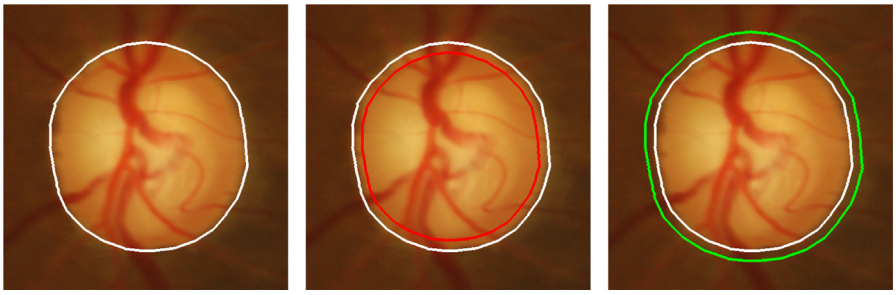
Table 5
Number of features and best results obtained in the feature selection step.

Descriptor	Data(%)	N. of Attributes	A(%)	K(%)	P(%)	AUC(%)
GLCM	100	85	86.48	72.52	86.50	95.00
CNN-CaffeNet	3	123	83.27	65.79	83.30	90.40
CNN-VGGFNet	70	2867	82.59	64.37	82.60	89.60
CNN-VGGMNet	2	82	83.50	66.38	83.50	89.90
CNN-VGGSNet	70	2867	82.59	65.67	82.60	89.80
GLCM + CNNs	4	659	**93.35**	**86.48**	**93.40**	**98.00**

These values in bold mean that they are the best results obtained when compared to the other results in the table.

Table 6
Classification results of the proposed method in images of performance set.

Database	A(%)	K(%)	P(%)	AUC(%)
HRF	90.00	80.00	90.20	95.30
JSIEC	95.52	85.25	95.50	98.10
ACRIMA	95.31	90.45	95.40	99.10



(a) Ground truth. (b) Under-segmentation. (c) Over-segmentation.

Fig. 9. Examples of under- and over-segmentation.

Table 7
Classification results for the proposed method for the images in the development set, with segmentation errors.

Segmentation Error	A(%)	K(%)	P(%)	AUC(%)
Under	82.58	64.32	82.60	90.10
Over	87.62	74.57	87.60	94.30

Table 8
Classification results for the proposed method for the images in the performance set, with segmentation errors.

Segmentation Error	A(%)	K(%)	P(%)	AUC(%)
Under	83.50	56.11	83.10	93.40
Over	84.54	56.32	85.90	92.60

It is possible to observe a decrease in the rates compared to the previous results. However, the method obtained a very good Kappa index for the development set and a good Kappa index for the performance set. This result shows that even when there are segmentation errors, the proposed method still obtains promising results.

5. Discussion

The results obtained by the proposed method are comparable to the State of Art. In some cases, researchers use private glaucoma labeling databases and, for that reason, it is difficult to reproduce the results presented in their work. Table 9 gives a comparison between our method and others of the State of the Art that use image databases different from those used by us.

Among the works listed in Table 9 only ours and that of Chen et al. [19] used more than one image database. The number of images used by Chen et al. [19] is larger than ours. However, the SCES database is unbalanced, of the 1676 images only 46 are considered unhealthy. Such characteristic of the database can interfere with the generalization of the proposed method.

In Ahn et al. [20], of 1542, 464 images were used in the test, While our set was composed of 802 images from three different bases. Raghavendra et al. [21] achieved the best accuracy rate of the methods listed. However, the validation method was different from other studies. The authors chose the hold-out method and randomly split the database in 70% of images for training and 30% for testing.

Christopher et al. [36] and Shibata et al. [35] used the largest sets of images. However, due to the need for a large number of images for CNN fine-tuning, the tests were performed in a few images. Christopher et al. [36] used only 10% of the 14822 images for testing while Shibata et al. [35] used only 3.69% (120 of 3252 images).

We compared our results with those obtained from the methods of Kotyk et al. [4] and Claro et al. [17]. Table 10 presents the results from applying these algorithms to the 873 images in the development set, and Table 11 presents the results for the 802 images in the performance set.

Our approach obtained better performance than the other methods. For the development set, we achieved an accuracy of 93.35% and a Kappa value of 86.48%, which is considered excellent, while the other two papers obtained a Kappa value rated as very good. We observed a similar situation for the performance set: we achieved an accuracy of 92.78% and a Kappa of 81.85%, rated excellent, while the other two papers reached an Kappa index rated good. In both tests, our method achieved the lowest values for standard deviation, demonstrating the description capacity of the proposed feature set.

We performed a Z-test [62] to statistically compare and evaluate the results presented in Tables 10 and 11 with a significance level of 5% in order to assess whether the results were significantly different from each other. The results show that our approach achieved markedly higher performance than the methods of Kotyk et al. [4] and Claro et al. [17].

As the DRISHTI, RIM-ONE-1, RIM-ONE-2, RIM-ONE-3, HRF and ACRIMA are public databases, a comparison with other state-of-the-art algorithms that used this databases is also possible. Tables 12–17 present these comparisons. The results presented in these Tables were taken from the original papers.

From the Tables 12–17, we compared our method with others that use pre-trained CNNs for feature extraction and others that fine-tuned CNNs. In most situations, our method was superior in performance. Only the method of Cerentine et al. [34] achieved a better result in the RIM-ONE-1 base and a similar result in the HRF database.

Particularly noteworthy is a comparison with the methods of Orlando et al. [5], Al-Bander et al. [24] and Lima et al. [25]. These methods chose to use only CNNs as descriptors. In our proposed descriptor, we have included texture information (GLCM). The results lead us to conclude that the texture characteristics are fundamental in the characterization of the existence or otherwise of glaucoma in retinal images.

Table 9
Comparison results with related works which used different image databases.

Methods	Dataset	N. of images	Performance
Chen et al. [19]	ORIGA, SCES	2326	AUC ORIGA: 0.831 AUC SCES: 0.887
Ahn et al. [20]	Private	1542	Accuracy: 87.9% AUC: 0.94
Raghavendra et al. [21]	Private	1426	Accuracy: 98.13%
Shibata et al. [35]	Private	3132	AUC: 96.5%
Christopher et al. [36]	Private	14822	AUC: 91%
Proposed method	RIM-ONE-1, RIM-ONE-2, RIM-ONE-3, DRISHTI	873	Accuracy: 93.35% AUC: 0.98
Proposed method	HRF, JSIEC, ACRIMA	802	Accuracy: 93.61% AUC: 0.975

Table 10
Comparison of the proposed method with those of Kotyk et al. [4] and Claro et al. [17] for the development set of images.

Methods	A(%)	K(%)	P(%)	AUC(%)
Kotyk et al. [4]	86.48 ± 1.18	72.43 ± 0.01	86.50 ± 0.00	95.20 ± 0.00
Claro et al. [17]	87.00 ± 0.52	73.65 ± 0.01	87.05 ± 0.01	95.50 ± 0.00
Proposed	**93.35** ± 0.49	**86.48** ± 0.01	**93.40** ± 0.01	**98.00** ± 0.00

Table 11
Comparison of the proposed method with those of Kotyk et al. [4] and Claro et al. [17] for the performance set of images.

Methods	A(%)	K(%)	P(%)	AUC(%)
Kotyk et al. [4]	80.41 ± 0.84	49.63 ± 0.03	79.70 ± 0.01	83.80 ± 0.02
Claro et al. [17]	82.47 ± 1.29	52.82 ± 0.04	82.10 ± 0.02	76,60 ± 0.03
Proposed	**92.78** ± 1.29	**81.85** ± 0.04	**92.80** ± 0.01	**96.60** ± 0.00

Table 12
Comparison of the proposed method with those of the state of the art in the DRISHTI image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Orlando et al. [5]	–	–	–	76.20
Diaz-Pinto et al. [39]	75.25	–	–	80.41
Proposed	**86.13** ± 2.4	**64.13** ± 0.07	**86.20** ± 0.03	**88.30** ± 0.03

Table 13
Comparison of the proposed method with those of the state of the art in the RIM-ONE-1 image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Cerentini et al. [34]	94.20	–	–	–
Proposed	**93.03** ± 0.29	**80.11** ± 0.01	**93.30** ± 0.03	**98.30** ± 0.00

Table 14
Comparison of the proposed method with those of the state of the art in the RIM-ONE-2 image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Al-Bander et al. [24]	88.2	–	–	–
Cerentini et al. [34]	86.20	–	–	–
Lima et al. [25]	90.1	–	–	95.7
Diaz-Pinto et al. [39]	71.21	–	–	85.75
Proposed	**99.56** ± 0.10	**99.11** ± 0.00	**99.60** ± 9.42	**100.00** ± 0.00

Table 15
Comparison of the proposed method with those of the state of the art in the RIM-ONE-3 image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Cerentini et al. [34]	86.40	–	–	–
Lima et al. [25]	79.90	–	–	86.00
Proposed	**91.19** ± 1.02	**82.24** ± 0.02	**91.30** ± 0.00	**96.70** ± 0.00

Table 16
Comparison of the proposed method with those of the state of the art in the HRF image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Cerentini et al. [34]	90.00	–	–	–
Liu et al. [37]	86.7	–	–	89.00
Diaz-Pinto et al. [39]	80.00	–	–	83.54
Proposed	**90.00** ± 1.57	**80.00** ± 0.03	**90.20** ± 0.01	**95.30** ± 0.01

Table 17
Comparison of the proposed method with those of the state of the art in the ACRIMA image database.

Methods	A(%)	K(%)	P(%)	AUC(%)
Diaz-Pinto et al. [39]	70.21	–	–	76.78
Proposed	**95.31** ± 0.20	90.45 ± 0.00	95.40 ± 0.00	**99.10** ± 9.42

6. Conclusion

The literature shows that many researchers have directed their efforts to the field of CAD systems, in which the automatic diagnosis of glaucoma is a subarea. In this study, we have presented an automated approach to the diagnosis of glaucoma classification using texture and CNN descriptors.

Based on the literature, it seems that authors typically use private databases (or a single database) for the evaluation of their results. However, this situation does not represent a real environment for application of this type of system. We therefore evaluated our approach using six different image databases, each with specific characteristics.

Feature extraction and classification are essential steps in the automatic diagnosis of different computational systems. In this

paper, we evaluated the descriptors of texture, shape and CNNs. Our best result was achieved by concatenating GLCM with CNNs, obtaining an excellent Kappa index, and an accuracy of 93.35% for the development set and 92.78% for the performance set. We also verified that the proposed method works well even when there are imperfections in the OD segmentation.

The results presented here are promising, but they can be improved. In future work on image representation, we intend to investigate the feature extraction of the cup region in addition to the OD region. In these two areas, it is possible to calculate several metrics that can aid in the diagnosis of glaucoma, such as the cup-to-disc ratio. Furthermore, in the present work, we do not pre-process the image for the removal of blood vessels, and we intend to investigate the effect of this operation on the final classification result. Another critical point is the classification step; since our primary objective was a description of the image, the evaluated classifiers were not tuned. We believe that our results could be improved by fine parameter tuning.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgements

The authors would like to thank the Foundation for Research Support of the State of Piauí (FAPEPI) for sponsoring our research.

References

[1] U.R. Acharya, S. Dua, X. Du, C.K. Chua, et al., Automated diagnosis of glaucoma using texture and higher order spectra features, IEEE Trans. Informat. Technol. Biomed. 15 (3) (2011) 449–455.
[2] S. Simonthomas, N. Thulasi, P. Asharaf, Automated diagnosis of glaucoma using haralick texture features, in: IEEE International Conference on Information Communication and Embedded Systems, 2014, pp. 1–6.
[3] C. Srinivasan, S. Dubey, T. Ganeshbabu, Complex texture features for glaucomatous image classification system using fundus images, Int. J. Eng. Res. Sci. 2 (13) (2016) 106–113.
[4] T. Kotyk, S. Chakraborty, N. Dey, T. Gaber, A.E. Hassanien, V. Snasel, Semi-automated system for cup to disc measurement for diagnosing glaucoma using classification paradigm, in: Proceedings of the Second International Afro-European Conference for Industrial Advancement, Springer, 2016, pp. 653–663.
[5] J.I. Orlando, E. Prokofyeva, M. del Fresnob, M. Blaschko, Convolutional neural network transfer for automated glaucoma identification, in: 12th International Symposium on Medical Information Processing and Analysis, International Society for Optics and Photonics, 2017, pp. 101600U–101600U.
[6] H.A. Quigley, A.T. Broman, The number of people with glaucoma worldwide in 2010 and 2020, British J. Ophthalmol. 90 (3) (2006) 262–267.
[7] P.S. Mittapalli, G.B. Kande, Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma, Biomed. Signal Process. Control 24 (2016) 34–46.
[8] M.S. Haleem, L. Han, J.I. van Hemert, B. Li, Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review, Computer. Med. Imag. Graphics 37 (2013) 581–596.
[9] P.S. Grewal, F. Oloumi, U. Rubin, M.T. Tennant, Deep learning in ophthalmology: A review, Can. J. Ophthalmol. 53 (4) (2018) 309–313.
[10] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imag. 35 (5) (2016) 1285–1298.
[11] L. Breiman, Random forests, Machine Learn. 45 (1) (2001) 5–32.
[12] C. Cortes, V. Vapnik, Support-vector networks, Machine Learn. 20 (3) (1995) 273–297.
[13] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, Tech. Rep., California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
[14] R.M. Haralick, K. Shanmugam, et al., Textural features for image classification, IEEE Trans. Syst., Man, Cybernet. 3 (6) (1973) 610–621.
[15] T. Ojala, M. Pietikäinen, Unsupervised texture segmentation using feature distributions, in: International Conference on Image Analysis and Processing, Springer, 1997, pp. 311–318.
[16] A.A. Salam, T. Khalil, M.U. Akram, A. Jameel, I. Basit, Automated detection of glaucoma using structural and non structural features, SpringerPlus 5 (1) (2016) 1519.
[17] M. Claro, L. Santos, W. Silva, F. Araújo, N. Moura, A. Macedo, Automatic glaucoma detection based on optic disc segmentation and texture feature extraction, CLEI Electronic J. 19 (2) (2016) 5.
[18] Y. Hagiwara, J.E.W. Koh, J.H. Tan, S.V. Bhandary, A. Laude, E.J. Ciaccio, L. Tong, U.R. Acharya, Computer-aided diagnosis of glaucoma using fundus images: A review, Comput. Methods Programs Biomed. 165 (2018) 1–12.
[19] X. Chen, Y. Xu, D.W. Kee Wong, T.Y. Wong, J. Liu, Glaucoma detection based on deep convolutional neural network, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 715–718. https://doi.org/10.1109/EMBC.2015.7318462.
[20] J.M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K.B. Lee, U.S. Kim, A deep learning model for the detection of both advanced and early glaucoma using fundus photography, PloS One 13 (11) (2018) e0207982.
[21] U. Raghavendra, H. Fujita, S.V. Bhandary, A. Gudigar, J.H. Tan, U.R. Acharya, Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images, Inf. Sci. 441 (C) (2018) 41–49, https://doi.org/10.1016/j.ins.2018.01.051.
[22] Z. Zhang, F.S. Yin, J. Liu, W.K. Wong, N.M. Tan, B.H. Lee, J. Cheng, T.Y. Wong, Origa-light: An online retinal fundus image database for glaucoma analysis and research, in: Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE, 2010, pp. 3065–3068.
[23] C.C. Sng, L.-L. Foo, C.-Y. Cheng, J.C. Allen Jr, M. He, G. Krishnaswamy, M.E. Nongpiur, D.S. Friedman, T.Y. Wong, T. Aung, Determinants of anterior chamber depth: the singapore chinese eye study, Ophthalmology 119 (6) (2012) 1143–1150.
[24] B. Al-Bander, W. Al-Nuaimy, M.A. Al-Taee, Y. Zheng, Automated glaucoma diagnosis using deep learning approach, in: 2017 14th International Multi-Conference on Systems, Signals & Devices (SSD), IEEE, 2017, pp. 207–210.
[25] A.C. de Moura Lima, L.B. Maia, R.M.P. Pereira, G.B. Junior, J.D.S. de Almeida, A.C. de Paiva, Glaucoma diagnosis over eye fundus image through deep features, in: 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE, 2018, pp. 1–4.
[26] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, arXiv preprint arXiv:1405.3531, 2014.
[27] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, arXiv preprint arXiv:1312.6229, 2013.
[28] J. Sivaswamy, S. Krishnadas, G.D. Joshi, M. Jain, A.U.S. Tabish, Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation, in: IEEE International Symposium on Biomedical Imaging, 2014 IEEE 11th, 2014, pp. 53–56.
[29] T. Hastie, R. Tibshirani, J. Friedman, Unsupervised learning, in: The Elements of Statistical Learning, Springer, 2009, pp. 485–585.
[30] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Informat. Process. Syst. (2012) 1097–1105.
[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[33] T.J. Jun, D. Kim, H.M. Nguyen, D. Kim, Y. Eom, 2sranking-cnn: A 2-stage ranking-cnn for diagnosis of glaucoma from fundus images using cam-extracted ROI as an intermediate input, CoRR abs/1805.05727, 2018. arXiv: 1805.05727. URL http://arxiv.org/abs/1805.05727.
[34] A. Cerentini, D. Welfer, M.C. d'Ornellas, C.J.P. Haygert, G.N. Dottob, Automatic identification of glaucoma sing deep learning methods, MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics, vol. 245, IOS Press, 2017, p. 318.
[35] N. Shibata, M. Tanito, K. Mitsuhashi, Y. Fujino, M. Matsuura, H. Murata, R. Asaoka, Development of a deep residual learning algorithm to screen for glaucoma from fundus photography, Sci. Rep. 8 (1) (2018) 14665.
[36] M. Christopher, A. Belghith, C. Bowd, J.A. Proudfoot, M.H. Goldbaum, R.N. Weinreb, C.A. Girkin, J.M. Liebmann, L.M. Zangwill, Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs, Sci. Rep. 8 (1) (2018) 16685.
[37] S. Liu, S.L. Graham, A. Schulz, M. Kalloniatis, B. Zangerl, W. Cai, Y. Gao, B. Chua, H. Arvind, J. Grigg, et al., A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs, Ophthalmol. Glaucoma 1 (1) (2018) 15–22.
[38] J.J. Gómez-Valverde, A. Antón, G. Fatti, B. Liefers, A. Herranz, A. Santos, C.I. Sánchez, M.J. Ledesma-Carbayo, Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning, Biomed. Opt. Exp. 10 (2) (2019) 892–913.
[39] A. Diaz-Pinto, S. Morales, V. Naranjo, T. Köhler, J.M. Mossi, A. Navea, Cnns for automatic glaucoma assessment using fundus images: an extensive validation, Biomed. Eng. Online 18 (1) (2019) 29.
[40] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[41] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[42] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.
[43] B. Dalal, Navneet e Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, vol. 1, 2005, pp. 886–893.
[44] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Trans. Syst., Man, Cybernet. 8 (6) (1978) 460–473.
[45] M.M. Galloway, Texture analysis using gray level run lengths, Comput. Graphics Image Process. 4 (2) (1975) 172–179.
[46] A. Chu, C.M. Sehgal, J.F. Greenleaf, Use of gray value distribution of run lengths for texture analysis, Pattern Recogn. Lett. 11 (6) (1990) 415–419.
[47] E.B. Dasarathy, Belur V.e. Holder, Image characterizations based on joint gray level—run length distributions, Pattern Recogn. Lett. 12 (8) (1991) 497–502.
[48] J. Cheng, J. Liu, Y. Xu, F. Yin, D.W.K. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, T.Y. Wong, Superpixel classification based optic disc and optic cup segmentation for glaucoma screening, IEEE Trans. Med. Imaging 32 (6) (2013) 1019–1032.
[49] J.R. Quinlan, Induction of decision trees, Machine Learn. 1 (1) (1986) 81–106.
[50] Y.-D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, S.-H. Wang, Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, Multimedia Tools Appl. (2017), https://doi.org/10.1007/s11042-017-5243-3.
[51] Y.-D. Zhang, K. Muhammad, C. Tang, Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on gpu platform, Multimedia Tools Appl. (2018), https://doi.org/10.1007/s11042-018-5765-3.
[52] S.-H. Wang, J. Sun, P. Phillips, G. Zhao, Y.-D. Zhang, Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units, J. Real-Time Image Proc. (2017), https://doi.org/10.1007/s11554-017-0717-0.
[53] L.H. Vogado, R.M. Veras, F.H. Araujo, R.R. Silva, K.R. Aires, Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification, Eng. Appl. Artif. Intell. 72 (2018) 415–422, https://doi.org/10.1016/j.engappai.2018.04.024. https://www.sciencedirect.com/science/article/pii/S0952197618301039.
[54] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
[55] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, ACM SIGKDD Explorat. Newsletter 11 (1) (2009) 10–18.
[56] D.M. Powers, Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, school of informatics and engineering, flinders university, adelaide, australia, Tech. Rep., TR SIE-07-001, Journal of Machine Learning Technologies 2: 1 37-63, 2007.
[57] G.H. Rosenfield, K. Fitzpatrick-Lins, A coefficient of agreement as a measure of thematic classification accuracy, Photogramm. Eng. Remote Sens. 52 (2) (1986) 223–227.
[58] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.
[59] F. Fumero, S. Alayón, J. Sanchez, J. Sigut, M. Gonzalez-Hernandez, Rim-one: An open retinal image database for optic nerve evaluation, in: IEEE 24th International Symposium on Computer-Based Medical Systems, 2011, pp. 1–6.
[60] A. Budai, R. Bock, A. Maier, J. Hornegger, G. Michelson, Robust vessel segmentation in fundus images, Int. J. Biomed. Imag. (2013) 154860.
[61] J.A. de Sousa, A.C. de Paiva, J.D.S. de Almeida, A.C. Silva, G.B. Junior, M. Gattass, Texture based on geostatistic for glaucoma diagnosis from fundus eye image, Multimedia Tools Appl. (2017) 1–18.
[62] R.G. Congalton, K. Green, Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, 2nd ed., CRC Press, Boca Raton, 2008.