Article

Accurate Diagnosis of Diabetic Retinopathy and Glaucoma Using Retinal Fundus Images Based on Hybrid Features and Genetic Algorithm

Nasser Tamim [1,2,]*[iD], Mohamed Elshrkawey [1][iD] and Hamed Nassar [1][iD]

[1] Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt; melshrkawey@ci.suez.edu.eg (M.E.); nassar@ci.suez.edu.eg (H.N.)
[2] Deanship of Community Services and Continuing Education, Jazan University, Gazan 82142, Saudi Arabia
[*] Correspondence: ntamim@jazanu.edu.sa; Tel.: +20-1006237484

Abstract: Diabetic retinopathy (DR) and glaucoma can both be incurable if they are not detected early enough. Therefore, ophthalmologists worldwide are striving to detect them by personally screening retinal fundus images. However, this procedure is not only tedious, subjective, and labor-intensive, but also error-prone. Worse yet, it may not even be attainable in some countries where ophthalmologists are in short supply. A practical solution to this complicated problem is a computer-aided diagnosis (CAD) system—the objective of this work. We propose an accurate system to detect at once any of the two diseases from retinal fundus images. The accuracy stems from two factors. First, we calculate a large set of hybrid features belonging to three groups: first-order statistics (FOS), higher-order statistics (HOS), and histogram of oriented gradient (HOG). Then, these features are skillfully reduced using a genetic algorithm scheme that selects only the most relevant and significant of them. Finally, the selected features are fed to a classifier to detect one of three classes: DR, glaucoma, or normal. Four classifiers are tested for this job: decision tree (DT), naive Bayes (NB), k-nearest neighbor (kNN), and linear discriminant analysis (LDA). The experimental work, conducted on three publicly available datasets, two of them merged into one, shows impressive performance in terms of four standard classification metrics, each computed using k-fold crossvalidation for added credibility. The highest accuracy has been provided by DT—96.67% for DR, 100% for glaucoma, and 96.67% for normal.

Keywords: retinal fundus imaging; diabetic retinopathy; glaucoma; genetic algorithm; feature extraction and selection; multi-classification

1. Introduction

The eyes can be considered a mirror of the human body, allowing for non-invasive diagnosis of numerous illnesses [1]. In particular, the retina can be used as a useful indicator of many diseases. As a person grows older, several diseases leave significant indicative signs in the retina.

According to the fact sheet of the World Health Organization (WHO), blindness and vision impairment affect about 2.22 billion people around the world. One billion of them have a chance to prevent the decrease in vision if they are recognized and diagnosed at an earlier stage [2].

Diabetic retinopathy (DR) and glaucoma, which are the focus of the present work, are regarded as the most important causes of blindness, mainly due to the progress in life prospects and permissive lifestyles [3]. DR is a complication of diabetes that inevitably leads to the terminal state of vision loss. Glaucoma is a chronic ocular disease caused by increased fluid pressure in the optic nerve and ends up also harming peripheral vision [4]. Globally, it is estimated that at least 6.9 million people have glaucoma and 4.5 million have DR [5].

The International Diabetic Federation estimates that, worldwide, one in every ten people suffers from diabetes [6], and that those persons have a significant chance to suffer from DR and glaucoma impairments. Moreover, it is estimated that the number of people who suffer from glaucoma is expected to reach 111.8 million by 2040 [7].

DR and glaucoma are irreversible diseases, so detecting them at an early stage is critical to avoid further deterioration in the retina. On the other hand, some countries, especially in the developing world, have a severe shortage of oculist doctors [8], requiring computer-aided diagnosis, such as the one proposed in the present work.

Luckily, images of the retina can help in diagnosing several retinal diseases, including DR and glaucoma [9]. This process is usually carried out and interpreted manually, which is laborious and prone to error due to minute image details, as shown in Figure 1. Consequently, if this process is automated, it would help ophthalmologists with a supportive tool that makes both diagnosis and decision making faster, easier, less expensive, and more accurate. To this end, computer vision and imaging techniques, such as those in the present work, can be leveraged.
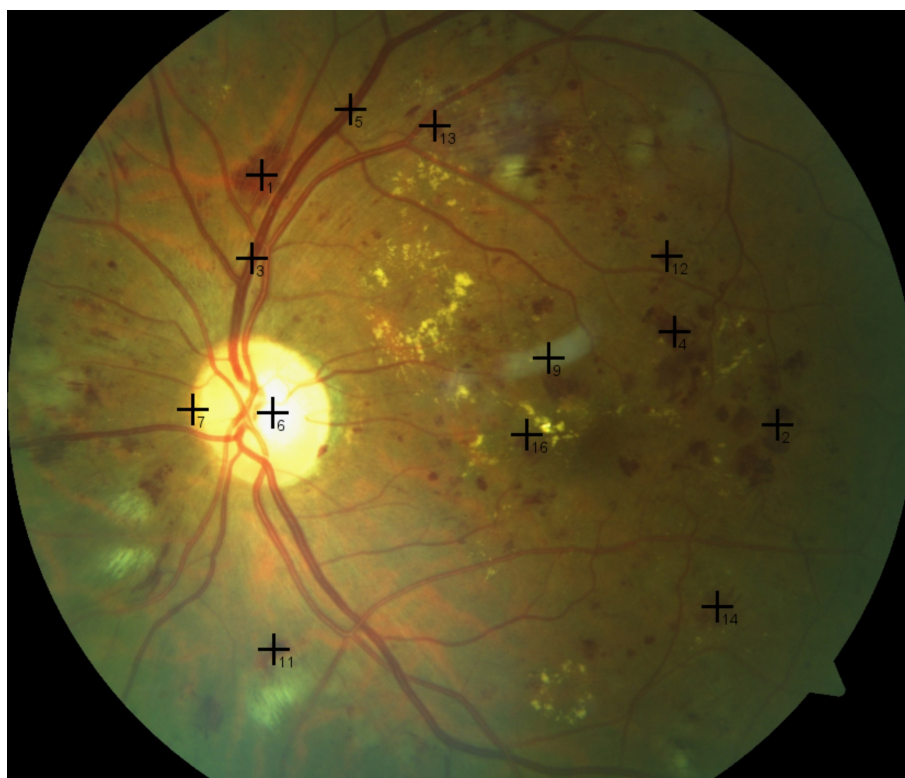

Figure 1. The subtle and tiny intensity variations in the fundus image are too difficult to identify by the naked eye, and this is where the proposed CAD system can help.

For ophthalmic fundus imaging analysis, well-known modalities, such as 2D fundus images and 3D optical coherence tomography (OCT), can be employed [10]. In this study, the 2D fundus image modality is adopted due to its low cost and the simplicity in computation compared with 3D OCT modality images [11]. To this end, we use three publicly available datasets of retinal fundus images. It should be noted, however, that disease diagnosis via pathological structures of fundus images poses numerous challenges, such as:

1. Finding relevant datasets for each disease is laborious;
2. Labeling (annotating) the dataset usually requires ophthalmology experts, and can be problematic due to such considerations as privacy, safety, or ethics. The only alternative is to use publicly available images, as is done in the present work;

3. Subtle and tiny variations between intensity values of retinal objects, as shown in Figure 1, can cause errors;
4. The curvature of the retina, in addition to photographic capturing conditions, can often cause illumination spots in retinal fundus images, negatively affecting their quality.
5. Retinal diseases share common characteristics in the fundus images, giving rise to confusion.

This article tackles the above challenges, culminating in an accurate retinal diagnostic and prognostic system that discriminates among DR, glaucoma, and normal (free of diseases) cases using a retinal fundus image. To this end, the RGB image is first transformed into its elementary components: red (R), green (G), and blue (B) channels. Then, three groups of distinct features are composed for the image. The first-order statistics (FOS) group contains 4 features per channel, giving rise to 12 features for the 3 channels. The higher-order statistics (HOS) group contains 14 features per channel per displacement, giving rise to 112 features for the 3 channels and 8 displacements. The histogram of oriented gradient (HOG) group contains 81 features. The three groups thus collectively contain 429 features that can discern between the three classes we consider. However, to make the processing more efficient, a genetic algorithm (GA)-based scheme is leveraged to select the most relevant and informative of these features, successfully reducing them to 105.

To evaluate the feature selection process and the proposed system's performance in general, we use four machine learning (ML) classifiers: decision tree (DT), naive Bayes (NB), k-nearest neighbor (kNN), and linear discriminant analysis (LDA). Based on our experimentation with other classifiers, such as support vector machine (SVM), random forest (RF), and Adaboost, and based on our previous work [12] with artificial neural networks (ANNs) in the same image processing context, the four chosen classifiers provide higher accuracy, easier setup, fewer training images, and less training and/or testing time. It is possible, however, that a classifier that we did not experiment with performs better using the proposed features than the chosen four.

The article is organized as follows. Section 2 surveys previous work on retinal CAD systems, especially feature extraction and image classification. Section 3 describes the retinal fundus datasets used in our experiments and the methodology adopted in feature extraction and selection. Section 4 provides the results of the experimental work and discusses their implications and the gained insights, with the discussions of those results presented in Section 5. Finally, concluding remarks are given in Section 6.

2. Related Work

A sizable amount of research exists for disease diagnosis through retinal fundus images. The research can be categorized in six directions. The first direction is the binary computer-aided diagnosis (CAD) system that is based on retinal blood vessel segmentation [13]. The second direction is binary CAD systems dedicated to identifying age-related macular degeneration (AMD) disease [9]. The third direction is binary retinal CAD systems mainly devoted to segmenting the optic disc (OD) region [3]. The fourth direction includes multi-class CAD systems devoted to grading the progress levels of only DR [14], or only glaucoma [15]. The fifth direction includes binary CAD systems used to classify more than one retinal disease together, such as DR and/or glaucoma and/or AMD. These diseases are identified, however, one at a time, as either existing or not using binary classification [16]. The sixth direction includes multi-class CAD systems applied to classify several retinal diseases together and evaluates them as a multi-classification problem [17]. This direction is the most relevant to the present work, and is therefore reviewed in detail next, with Table 1 summarizing its highlights.

Table 1. Selected methods for retinal CAD systems through retinal fundus images.

Author	Images	Number of Images in Use	Methods	Results
[18]	KMC	DR: 170 Normal: 170	Feature Trac transform	Overall Acc = 0.9941 Overall Se = 0.9941 Overall Sp = 0.9941
	MESSIDOR	(DR: 170) (Normal: 170)		Overall Se = 1.0 Overall Sp = 1.0 Overall Acc = 1.0
[19]	ARIA	AMD: 60 Normal: 101	Random transform, EMD, HOS nonlinear, entropy, LSDA, MRMR	Overall Acc = 0.8509
	KMC	Normal: 270 AMD: 270		Overall Acc = 0.9167
	STARE	AMD: 47 Normal: 36		Overall Acc = 1.0
[20]	KMC	Normal: 400 Abnormal: 400 (115 AMD, 170 DR, and 115 Glaucoma)	Bi-dimensional decomposition and empirical mode integrated index	Overall Acc = 0.8863 Overall Se = 0.8625 Overall Sp = 0.9100
[4]	KMC	Normal: 244 Glaucoma: 244	Features Variation mode decomposition entropies	Overall Acc = 0.9480
[16]	KMC	Normal: 404 Abnormal: 1082 (AMD: 381, DR: 195, and Glaucoma: 506)	Continuous wavelet of oriented transform and entropy's features	Overall Acc = 0.9248 Overall Se = 0.8937 Overall Sp = 0.9558
[21]	KMC	Normal: 404 Abnormal: 1400 (AMD: 529, DR: 365, and Glaucoma: 506)	Pyramid histogram of oriented and SURF features Canonical correlation analysis Particle swarm optimization	Overall Acc = 96.21 Overall Se = 0.9050 Overall Sp = 0.9742
[22]	KMC	Normal: 404 AMD: 529 DR: 356 Glaucoma: 506	PHOW-GMM and Fisher vector Canonical correlation analysis Particle swarm optimization	Overall Acc = 0.9679 Overall Se = 0.9673 Overall Sp = 0.9696

An automatic lesion detection system is proposed by [23] for early detection of DR-related lesions. A bag-of-visual-words (BoVW) algorithm is employed to discriminate among different DR lesions using a maximum margin classifier. Despite the robustness of the system, its only purpose is to identify the different types of DR signs rather than grading the DR severity.

Ganesan et al. [18] suggest a binary CAD system for DR detection by experimenting on two datasets of images. The RGB fundus image is transformed into a gray-level image. Out of a total of 840 extracted features, only 670 meaningful features are used. This system leverages a support vector machine (SVM), probabilistic neural network (PNN), and PNN improved by a genetic algorithm (PPN-GA). In spite of the effectiveness of the system, its binary operation is a limitation. Moreover, the system is designed to find whether the pathology is DR positive, without determining the severity grade of the disease.

Mookiah et al. [19] present a CAD system based on nonlinear feature extraction to distinguish between normal and AMD images. The feature vector is encoded and provided to a SVM classifier. The system is tested on three datasets, one private and two public,

the public being automated retinal image analysis (ARIA) and structured analysis of the retina (STARE). The system exhibits an accuracy of 85.09%, 91.67%, and 100% for the ARIA, private, and STARE datasets, respectively. Aside from the fact that the system is dedicated to only AMD diagnosis, it is only valid when the retinal image resolution is uniform.

Qaisar et al. suggest an automatic retinal system to recognize the five severity levels of DR: normal, mild NDPR, moderate NPDR, severe NPDR, and PDR [24]. This system was developed using deep visual features (DVFs) and was tested on fundus images from three public datasets and one private.

All the systems mentioned above are applied to diagnose only a single disease in the image dataset. Next, we will review work that is directed at multi-disease diagnosis DR, AMD, and glaucoma.

A system dedicated to identifying DR, AMD, and glaucoma is presented by [20]. The system is based on decomposing the fundus image into $2D$ intrinsic mode functions (IMFs) to detect the pixels' morphological variations. The system leverages an SVM classifier to differentiate between pathological symptoms pertaining to DR, AMD, and glaucoma and normal images. The system performance achieved 88.63%, 86.25%, and 91.00% for accuracy, sensitivity, and specificity, respectively. Although this system considers three diseases, the final decision output is a binary classification (normal or abnormal).

Koh et al. [16] present a two-class CAD system dedicated to identifying DR, AMD, glaucoma from fundus images. In total, 404 normal and 1082 abnormal fundus images were experimented on. The system achieved 92.48%, 89.37%, and 95.58% for overall accuracy, overall sensitivity, and overall specificity, respectively, using 15 features. Although the system uses four classes of images (AMD, DR, glaucoma, and normal), the system only predicts binary classification (normal or abnormal) for each disease. Since there are 26.7% more abnormal fundus images than normal ones, this imbalance makes the system's final decision more biased towards the abnormal class.

In a subsequent work, Koh et al. [22] present another CAD system to identify three retinal diseases, AMD, DR, and glaucoma. The bag-of-visual-words approach and Gaussian mixture model (GMM) were carried out on the training dataset to build the vocabulary. Random forest (RF) was employed as a classification model. However, aside from the fact that the system was tested on a private dataset of images, a requirement for reproducibility, the final results are somewhat disappointing. The reason why the results are not impressive could be because the system depends on the extracted vocabulary from image patterns, not from the image pixels themselves. In addition, the final results are provided on an aggregate basis, i.e., for all three classes of cases, DR, glaucoma and normal.

ANN in general and deep learning in particular have also been used heavily for ophthalmic diagnosis using retinal fundus images. An overview of the applications of deep learning of this topic is presented in [25], where the authors describe image datasets that can be used for deep learning purposes, and applications of deep learning for segmentation of the optic disc, optic cup, and blood vessels as well as detection of lesions.

A retinal CAD system based on deep residual NN classifiers using small labeled images is presented in [17]. The system employs three tasks related to eye disease diagnosis. The first is identifying five broad categories, DR, AMD, glaucoma, melanoma, and normal images. The second is predicting one of the 320 fine-grained disease sub-categories (grades). The third is creating a textual explanation for the diagnosis. Experimenting on a dataset of 7212 labeled and 35,854 unlabeled images from 3502 patients, the system achieved an overall accuracy of 83%, 75%, and 48% for the three tasks, respectively, which are low compared to the results of the present work.

A medical image analysis based on deep mining for screening DR is proposed by Quellec et al. [26], aiming to detect the four different DR lesion types in fundus images. Specifically, a deep convolutional NN with 26 different layers acts to automatically detect pathological features.

A combined convolutional NN and recurrent NN were developed for heightened glaucoma detection [27]. The CNN/RNN combined model attained an average *F* score

of 96.2% in identifying glaucoma. This system suffers, however, from low accuracy if the number of training images is not large enough.

A retinal CAD system based on a deep learning technique is presented by Zhao et al. [28]. A transfer deep learning architecture depends on a residual NN, including feature attention and channel re-calibration to extract features from the retinal fundus image. RGB Kaggle fundus images were used after pre-processing steps to mitigate noise effects of the background. This system resulted in 87.60% for the area under the curve (AUC) of the receiver operating characteristic (ROC) and 59.94% for overall accuracy. This system is highly robust, but the computational cost is relatively high.

In [29], a deep learning system is proposed to study its efficacy in diagnosing and grading glaucomatous optic neuropathy (GON) through retinal fundus images. The network comprises twenty-two layers, of which eleven are inception-v3 architecture modules. Out of 70,000, 48,116 fundus images were selected to be annotated by ophthalmology doctors (twenty-seven doctors) for labeling as GON images or not GON images. A minibatch gradient descent of a size of thirty-two was used for training using an Adam optimizer, with an initial learning rate of 0.002. The final results achieved for the system were 0.929, 0.956, and 0.920 for accuracy, sensitivity, and specificity [29].

A multi-layer perceptron (MLP) NN and hand-crafted features are presented by Tamim et al. [12], using retinal fundus images to diagnose DR by searching pixel by pixel in the retinal image for biomarkers of DR lesions. The system uses 24 hand-crafted features and a 3-layer NN. A post-processing technique depending on mathematical morphological operators is used to optimize the blood vessel segmentation procedure. A selected vector of features is proposed in this study. In comparison, three publicly available datasets are used. The experimental results, visually and quantitatively, indicate the robustness of the suggested methods. The proposed method gave 0.960, 0.754, and 0.984 for the DRIVE dataset, 0.963, 0.780, and 0.982 for the STARE dataset, and 0.957, 0.758, and 0.984 for the CHASSE_DB1 dataset, for accuracy, sensitivity, and specificity, respectively. Despite the robustness of this method, it is dedicated to only one unique ocular disease, *DR*.

The use of NNs and deep learning is still the focus of much research in the area of retinal image analysis. For example, a technique based on a gray wolf optimized NN is presented by Jerith and Kumar [30] for early recognition of glaucoma using retinal images. They first converted the original color images into gray-level, then the artifacts due to noise were suppressed using an adaptive median filter. The extracted features comprised gray-level co-occurrence matrix (GLCM) features, speedup robust features, histogram of oriented gradient (HOG) features, and global features that were encoded for gray wolf optimization.

In [31], Maqsood et al. studied hemorrhage detection based on a 3D CNN deep learning framework and feature fusion for evaluating retinal abnormality in diabetic patients. A pre-trained modified CNN model was employed to extract features to form a feature vector. The feature vector was fused by convolutional sparse image decomposition, and later the best features were selected by multi-logic regression controlled entropy variance techniques. They achieved 0.9771 for the average accuracy.

Li et al. [29] proposed a deep learning system to detect glaucomatous optic neuropathy from retinal fundus images. Their NN comprised 22 layers, with 11 inception-v3 architecture modules. Out of 70,000 images, 48,116 fundus images were selected for annotation by ophthalmologists as GON or not GON. The final results achieved for the system were 0.929, 0.956, and 0.920 for accuracy, sensitivity, and specificity, respectively.

In [32], Ramasamy et al. studied detection of DR using a fusion of textural and ridgelet features of retinal images and a sequential minimal optimization classifier. The ocular features were extracted and fused based on co-occurrence, run-length features, and ridgelet transform coefficients. The method was tested using sequential minimal optimization through retinal ocular images. The sensitivity, specificity, and accuracy obtained were 0.9887, 0.9524, and 0.9705 for the DIAREKDB1 dataset, and 0.909, 0.91, and 0.910 for the Kaggle dataset, respectively.

The problem with deep learning, however, is that it acts mostly as a black box, with plenty of stacked layers, providing little knowledge about local features at the image level. Furthermore, its training requires a huge number of images, which may not be readily available. For example, the work by Quellec et al. [26] depends on 90,000 images from Kaggle and 110,000 images from e-ophtha, in addition to 89 DiaretDB1 datasets. Another problem with deep learning is that it is remarkably time-consuming, primarily during training but also during testing. All these drawbacks are mitigated in the approach followed in the present work which depends mainly on hand-crafted features and conventional ML models.

3. Materials and Methods

Despite the huge body of work in the field of eye CAD systems, there is a gap in handling and classifying more than one retinal disease at once (direction six above), in particular distinguishing between cases of DR, glaucoma, and normal conditions in fundus images. This could be due to the scarcity of publicly available fundus images, especially ones with both glaucoma and DR. It could also be due to the fine and subtle differences between the intensity values of different pathological patterns (pixels) in the images, as shown in Figure 1. The present work is intended to bridge this gap.

It should be noted that both normal and DR fundus images are macula centered, so FOS and HOS features are applied for characterizing the pathological pattern of DR. On the other hand, glaucoma fundus images are OD centered [22]. Thus, HOG features are used to identify the pathological pattern for glaucoma.

To obtain an accurate retinal CAD system, this study employs three groups of hybrid features and a modified genetic algorithm (GA)-based scheme, resulting in the following contributions:

1. A fully automated system that is easy to use.
2. The classification model is developed using a balanced number of retinal fundus images in each class, ensuring robustness.
3. The center pixel of the OD in Dataset_2 is identified using three expert ophthalmologists, expediting the cropping of the OD for further processing and reaching the quality of the High-Resolution Fundus (HRF) dataset.
4. The model utilizes all the three channels (red, green, and blue) of images, capturing all potential information in the fundus image.
5. Three different groups of feature extractors are used to carefully obtain an optimal number of 429 features, ensuring both accuracy and efficiency.
6. A modified genetic algorithm is used as a feature selector to select the most relevant features among all the 429 extracted features.
7. Classifying more than two cases, DR, glaucoma and normal conditions, from the fundus image opens the door to generalization where any number of cases can be identified.
8. In-depth analysis plus a fair comparison with the nearest state-of-the-art systems.

The proposed system is based on texture features extracted from the region of interest (ROI) of the retinal fundus images. FOS features and HOS features are extracted from all three RGB channels to characterize DR disease, while HOG features are extracted from the optical disc (OD) region in gray-level images. The system encompasses four main stages: color transformation, feature extraction, feature selection, and classification, as shown in Figure 2.

The RGB color model is commonly used in CAD systems because it is based on the three primary colors embedded in most computerized input and output devices, unlike the HSV model which is closer to how humans perceive color. As such, it lends itself readily to digital image processing in general and digital retinal image research in particular. Even the few attempts that initially use HSV, such as the attempt by Zhou et al. [33], to provide some lamination and contrast enhancement for poor-quality images eventually revert back to RGB for further processing.

In the beginning, the RGB image is processed in two steps. The first step is to transform the RGB image into its essential R, G, and B channels. The second step is to identify and crop the OD region using the given center pixel location and the OD's diameter. The cropped OD region is converted into gray-scale.

Next, three different feature sets, FOS, HOS, and HOG, are then used for constructing three different sets of features that can distinguish between the pathological and normal patterns for each disease. The first set, FOS, consists of four features: mean, variance, skewness, and kurtosis while HOS features are 14 textural features: angular second moment (ASM), contrasts, correlation, sum of squares (variance), inverse difference moment (homogeneity), sum average, sum entropy, sum variance, entropy, difference variance, difference entropy, information measure of correlation 1, information measure of correlation 2, and auto-correlation.

The third set, HOG, comprises 81 features. Both FOS and HOS extractors are used to gather the statistical texture information for the entire retinal R, G, and B image, whereas HOG is employed to detect intensity changes in the gray-scale cropped OD region image. At the end of feature extraction, a feature vector with 429 distinct features is constructed.

In the next stage, a feature selection technique based on modified genetic algorithms is applied to reduce the feature space, which positively affects the final system performance.

In the last stage, four classifiers, decision tree (DT), k-nearest neighbor (kNN), naive Bayes (NB), and linear discriminant analysis (LDA) are used to make the diagnosis decision.
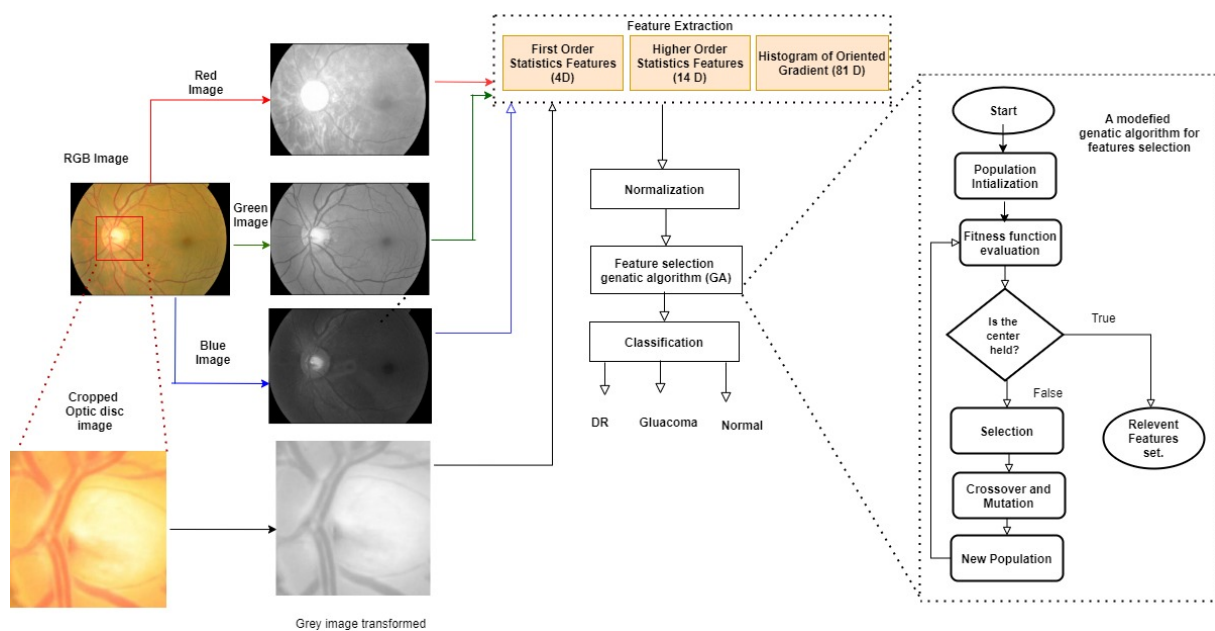

Figure 2. Block diagram of the proposed retinal computer-aided diagnosis system.

3.1. Data Acquisition

The first dataset in our work, High-Resolution Fundus (HRF) [34], comprises 45 RGB images of three different classes, DR, glaucoma, and normal conditions, with 15 images per class. The HRF dataset is provided with ma annotation by three expert ophthalmologists, specifying the OD center pixel location and the diameter of the OD. The second dataset, Dataset_2, consists of 288 RGB fundus images assembled from datasets. First, 192 images are from the Kaggle DR dataset [35] (96 images for DR class and 96 for the normal class). The remaining 96 images for glaucoma are collected from the BinRashed Eyes Glaucoma dataset [36]. Thus, the classes of Dataset_2 are balanced in that each class contains 96 images. Since Dataset_2 was acquired from two different sources, unlike HRF, it needs to be manually annotated to locate the OD center pixels. For this task, we made use of the expertise of three skilled ophthalmologists from King Fahd Central Hospital (KFCH), Saudi Arabia, who ably specified the center pixel and the diameter of the OD. The two

datasets used in the present work thus have three classes, DR, glaucoma, and normal (free of disease). The DR and normal images are macula centered, while the glaucoma images are OD centered. Table 2 provides general information related to the two image datasets.

Table 2. Number of images in each dataset.

Class	HRF dataset	Dataset_2
DR	15	96
Glaucoma	15	96
Normal	15	96
Total	**45**	**288**

Figure 3 shows the different pathological patterns for DR and glaucoma compared to the normal cases.
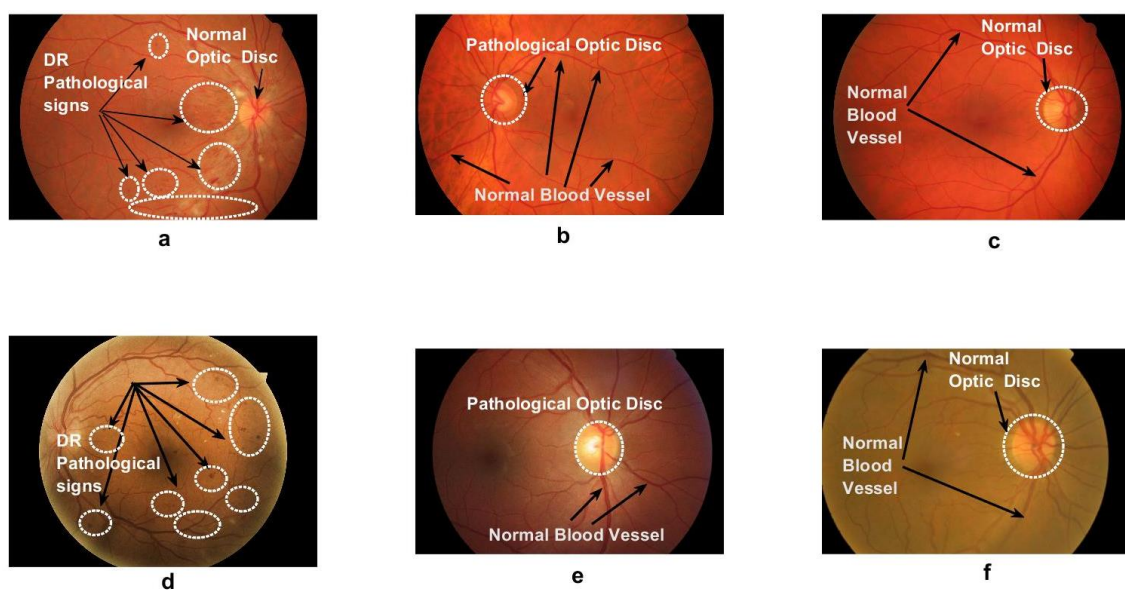

Figure 3. Sample images from the two datasets: (**a**) DR image from HRF dataset; (**b**) glaucoma image from HRF dataset; (**c**) normal (healthy) image form HRF dataset; (**d**) DR image from Dataset_2; (**e**) glaucoma image from Dataset_2; (**f**) normal image from Dataset_2.

3.2. Feature Selection

Feature selection reduces the amount of linearly dependent features to extract useful independent information for a given problem. Unlike prior work on diagnosing diseases through retinal fundus images, where the green channel of the RGB image is commonly used, in the present work, the three channels, red (R), green (G), and blue (B) are used to capture all the information available in the image. In addition, we make use of the gray-level version of the image because the textural features are mainly based on the spatial repetition of any two neighbor pixels in that version.

With the above in mind, three different feature sets are used to represent the retinal fundus image: FOS, HOS, and HOG. This is in sharp contrast to the less suitable method where only the green channel is used.

3.2.1. FOS Features

FOS texture features are the most straightforward and most commonly used to represent the image textural characteristics [37]. They are intensity based and mainly computed from a histogram of gray-level values of the image. They include mean, variance, skewness, and kurtosis which specify some of the image's textural characteristics and are obtained from the distribution of intensity values and individual pixel values in an image, as follows.

Let an image I, of size $H \times W$ pixels, where H and W are the height and width, respectively, of I in pixels. Then,

1. The mean, μ, gives information concerning the central tendency of the pixel intensities in the image. If (h, w), $h = 0, 1, \ldots, H - 1$, $w = 0, 1, \ldots, W - 1$ is a given pixel and $I(h, w)$ is the intensity of that pixel, then the mean of I is given by

$$\mu = \frac{\sum_{h=0}^{H-1} \sum_{w=0}^{W-1} I(h, w)}{H \times W}. \tag{1}$$

2. The variance σ^2, where σ is the standard deviation, determines how the intensities of pixels are scattered around the mean intensity μ and is given by

$$\sigma^2 = \frac{\sum_{h=0}^{H-1} \sum_{w=0}^{W-1} (I(h, w) - \mu)^2}{H \times W}. \tag{2}$$

3. The skewness μ^3 gives knowledge regarding the symmetry of the gray-level values around the mean. A given distribution of data is symmetric if it looks the same to the left and right side of the mean, as in the normal distribution; otherwise, it is skewed left or right, and the skewness is given by

$$\mu^3 = \frac{\sum_{h=0}^{H-1} \sum_{w=0}^{W-1} (I(h, w) - \mu)^3}{H \times W \times \mu}. \tag{3}$$

4. The kurtosis, μ^4, concerns the peakedness of the pixel intensity distribution in the image's histogram. It is a measure of the extreme values in the tail of the distribution, with a large kurtosis indicating that the tail peakedness exceeds that of the normal distribution. Data distributions with low kurtosis exhibit tail data that are generally less extreme than the normal distribution tails. The kurtosis is given by

$$\mu^4 = \frac{\sum_{h=0}^{H-1} \sum_{w=0}^{W-1} (I(h, w) - \mu)^4}{H \times W \times \mu} - 3. \tag{4}$$

Though simple to compute, FOS features ignore the spatial relationship between pixels and their surrounding neighbors. So, this group is not sufficient to quantify or discriminate between changes in retinal images [38].

3.2.2. HOS Features

The second set of features, HOS, contains textural features based on the gray-level co-occurrence matrix (GLCM) obtained from a 2D gray image [39]. The GLCM helps extract relevant textural features from an image by tracking the spatial distribution of pixels inside the region of interest (ROI) of the gray image. It is prevalent in medical image analysis and pattern recognition [40].

Consider a retinal 2D gray image, $\widetilde{I}$, with L gray levels and of size $H \times W$ pixels. We could characterize the textural information in terms of the frequent occurrence of the intensity of a reference pixel value, i, to its adjacent neighbor pixel value, j, with a predefined distance d and orientation θ. To quantify this image, the concept behind Haralick feature extraction [39] is to map it from the range $[\alpha, \beta]$, as a bit depth, into the range $[1, N_g]$, which holds the solicited number N_g of the gray level. The map quantization function, ψ, is defined as

$$\psi : [\alpha, \beta]^{H \times W} \longrightarrow [1, N_g]^{H \times W}. \tag{5}$$

Then, the quantized image, $\boldsymbol{I}$, is given by

$$\boldsymbol{I} = \psi(\widetilde{I}). \tag{6}$$

The non-normalized GLCM matrix is denoted by **P**, and its elements are obtained from the quantized image, *I*, through counting the number of times every pair of adjacent gray-level pixels occurs as a neighbor in the image *I*, or in an arbitrary area thereof. The neighborhood relation between pixels in terms of displacement between any two adjacent pixels can be defined by a displacement vector, $\nu = (d_x, d_y)$, in the *x* and *y* directions, with $d_x, d_y \in \mathbb{Z}$ representing the displacement in the *x* and *y* directions in terms of pixels with distance *d*. In the present work, $d = 1$. Each entry, $p(i,j)$, in **P** is given by

$$p(i,j) = \sum_{h=1}^{H} \sum_{w=1}^{W} \begin{cases} 1, & \text{if } i = I(h,w) \text{ and } j = I(h+d_x, w+d_y) \\ 0, & \text{otherwise} \end{cases},$$

The entry $p(i,j)$ tallies how many times gray values of the two nearby pixels occur in *I*, with $i, j \in [1, N_g]$.

According to the pioneering work of Haralick, eight displacement vectors can be used on 2D fundus retinal images to establish the direction within two adjacent pixels, as shown in Figure 4.
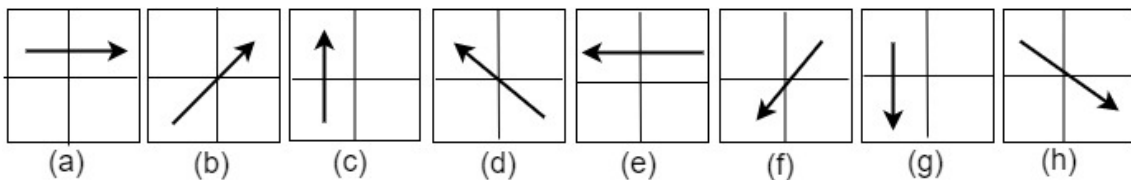
Figure 4. The eight displacement vectors that describe all possible directions of adjacency pixels. They are used to calculate the HOS features. (**a**) $\nu_{0°} = (1,0)$; (**b**) $\nu_{45°} = (1,1)$; (**c**) $\nu_{90°} = (0,1)$; (**d**) $\nu_{135°} = (-1,1)$; (**e**) $\nu_{180°} = (-1,0)$; (**f**) $\nu_{225°} = (-1,-1)$; (**g**) $\nu_{270°} = (0,-1)$; (**h**) $\nu_{325°} = (1,-1)$.

The eight displacement vectors described above are used in this work to conduct all possible relationships for the adjacent pixels in the eight directions in *I*. Therefore, in this work, there are eight GLCMs for each *I*. Once **P** is constructed based on a particular direction vector, then it is normalized, yielding the normalized GLCM:

$$\tilde{\mathbf{P}} = \begin{pmatrix} \tilde{p}(1,1) & \tilde{p}(1,2) & \cdots & \tilde{p}(1,N_g) \\ \tilde{p}(2,1) & \tilde{p}(2,2) & \cdots & \tilde{p}(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}(N_g,1) & \tilde{p}(N_g,2) & \cdots & \tilde{p}(N_g,N_g) \end{pmatrix},$$

where

$$\tilde{p}(i,j) = \frac{p(i,j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j)}$$

is the probability mass function for the gray level of adjacent pixels in *I*. So, the normalized matrix, $\tilde{\mathbf{P}}$, contains information about the retinal image pixels that have similar gray-level values, which is used later to compute the textural features for each of the eight displacement vectors per gray image.

The 14 textural features of an image are obtained from the $\tilde{\mathbf{P}}$ matrix, noting that N_g is the number of gray levels in the quantized image [41].

1. The angular second moment (ASM), f_1, feature is a measure of the consistency of textural information in an image. A consistent image is characterized by very few dominant gray level transitions between its pixels, and its related $\tilde{\mathbf{P}}$ matrix contains fewer entries of large magnitude. The more homogeneity in the image, the larger the value of ASM, with the range of this value being $[0,1]$. The f_1 feature is given by

$$f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (\tilde{p}(i,j))^2. \tag{7}$$

2. The contrast feature, f_2, is a measure of the status of the local variations in the intensity values between pixels of an image. The higher the local variation value, the larger the contrast. Contrast also represents a measure of intensity variation between a reference pixel and its neighbors in the image. Low contrast reflects low-intensity differences in $\tilde{\mathbf{P}}$ and vice versa. The value of the contrast feature, f_2, is given by

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left(\sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} \tilde{p}(i,j) \right). \tag{8}$$

There is no local contrast for a cell by itself in the $\tilde{\mathbf{P}}$ matrix, so the value of $|i-j|$ plays the role of a weight. There is no contrast if $|i-j| = 0$. The contract keeps increasing as $|i-j|$ becomes larger.

3. The correlation feature, f_3, indicates that pairs of adjacent pixels are correlated (positively, neutrally, or negatively) in the retinal image. It measures the amount of linear dependence of the gray-level values in the image.

    Let $\tilde{p}_x(i) = \sum_{j=1}^{N_g} \tilde{p}(i,j)$ be the ith element in the marginal probability matrix obtained by summing the columns of $\tilde{\mathbf{P}}$ and $\tilde{p}_y(j) = \sum_{i=1}^{N_g} \tilde{p}(i,j)$ be the jth element in the marginal probability matrix obtained by the summation of the rows of $\tilde{\mathbf{P}}$. Additionally, let $\mu_x = \sum_{i=1}^{N_g} i\tilde{p}_x(i)$ and $\mu_y = \sum_{j=1}^{N_g} j\tilde{p}_y(j)$ be the means of $\tilde{p}_x$ and $\tilde{p}_y$. Further, let

$$\sigma_x = \sqrt{\sum_{i=1}^{N_g} (i - \mu_x)^2 \tilde{p}_x(i)},$$

$$\sigma_y = \sqrt{\sum_{j=1}^{N_g} (j - \mu_y)^2 \tilde{p}_y(j)},$$

be the standard divisions of $\tilde{p}_x$ and $\tilde{p}_y$. Then the correlation feature, f_3, is given by

$$f_3 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{ij \ \tilde{p}(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}. \tag{9}$$

4. The sum of squares, or variance, feature, f_4, is a measure of the dispersion of the values around the mean and describes the higher weights that differ from the mean value. Let μ be the mean of the $\tilde{\mathbf{P}}$ matrix, then the variance features, f_4, are given by

$$f_4 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 \ \tilde{p}(i,j). \tag{10}$$

5. The inverse difference moment feature, f_5, shows how the distribution is close to the elements of the diagonal of $\tilde{\mathbf{P}}$. Conceptually, as the inverse difference moment decreases, the contrast increases. This feature is given by

$$f_5 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1 + (i-j)^2} \ \tilde{p}(i,j). \tag{11}$$

6. The sum average, f_6, feature is the higher weight to the higher index of the marginal $\tilde{\mathbf{P}}$ matrix. If $\tilde{p}_{x+y}(i)$ is the probability of the $\tilde{\mathbf{P}}$ matrix, then the f_6 feature is given by

$$f_6 = \sum_{i=2}^{2N_g} i\tilde{p}_{x+y}(i), \tag{12}$$

where

$$\tilde{p}_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{N_g} \sum_{j=1}^{N_g} \tilde{p}(i,j), \quad k = 2, 3, \ldots, 2N_g.$$

7. The sum variance feature, f_7, is a measure of the higher weights that differ from the sum entropy value (the f_8 feature below) of the marginal $\tilde{\mathbf{P}}$ matrix, and is given by

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2\, \tilde{p}_{x+y}(i). \tag{13}$$

8. The sum entropy feature, f_8, is calculated as

$$f_8 = -\sum_{i=2}^{2N_g} \tilde{p}_{x+y}(i) \log \tilde{p}_{x+y}(i). \tag{14}$$

Since $\log(0)$ is undefined, whenever a 0 probability is encountered during the computations, it should be replaced by an arbitrarily small positive value.

9. The entropy feature, f_9, is a concept of information theory, estimating the randomness of pixel intensities in the image. As such, its value is zero for a constant image. It is given by

$$f_9 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_9} \tilde{p}(i,j) \log \tilde{p}(i,j). \tag{15}$$

10. The difference variance feature, f_{10}, is a measure of the weights that differ from the difference entropy value (the f_{11} feature below) of the marginal $\tilde{\mathbf{P}}$ matrix, and is given by

$$f_{10} = \sum_{i=2}^{2N_g} (i - f_{11})^2\, \tilde{p}_{x-y}(i). \tag{16}$$

where

$$\tilde{p}_{x-y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^{N_g} \sum_{j=1}^{N_g} \tilde{p}(i,j), \quad k = 0, 1, \ldots, N_g - 1.$$

11. The difference entropy feature, f_{11}, is the weight of the higher difference of the index entropy and is given by

$$f_{11} = -\sum_{i=0}^{N_g-1} \tilde{p}_{x-y}(i) \log \tilde{p}_{x-y}(i). \tag{17}$$

12. The information measure of the correlation 1 feature, f_{12}, is considered as an entropy measure and is given by

$$f_{12} = \frac{f_9 - \delta}{\max(\alpha, \beta)}, \tag{18}$$

where α, β, and δ are the entropies of $\tilde{p}_x$, $\tilde{p}_y$, and $\tilde{p}_x\tilde{p}_y$, respectively, and are calculated as follows.

$$\alpha = -\sum_{i=1}^{N_g} \tilde{p}_x(i) \log \tilde{p}_x(i),$$

$$\beta = -\sum_{j=1}^{N_g} \tilde{p}_y(j) \log \tilde{p}_y(j),$$

$$\delta = -\sum_{i=1}^{N_g} \sum_{i=j}^{N_g} \tilde{p}(i,j) \log\big(\tilde{p}_x(i)\tilde{p}_y(j)\big).$$

13. The information measure of the correlation 2 feature, f_{13}, is given as follows. Let

$$\eta = -\sum_{i=1}^{N_g} \sum_{i=j}^{N_g} \tilde{p}_x(i)\tilde{p}_y(j) \log\big(\tilde{p}_x(i)\tilde{p}_y(j)\big)$$

be the entropy of $\tilde{p}_x(i)\tilde{p}_y(j)$. Then, f_{13} is given by

$$f_{13} = \sqrt{1 - \exp(-2(\eta - f_9))}. \tag{19}$$

14. The auto-correlation feature, f_{14}, is given by [41]

$$f_{14} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij\, \tilde{p}(i,j). \tag{20}$$

We content ourselves with only the first four FOS moments as they are usually enough to summarize the underlying data fairly accurately [42]. Higher moments add little value, and can complicate the computations. The compelling evidence of this claim is seen clearly in Table 3 where, out of the five FOS features selected by the GA, four belonged to the first two moments and only one to the fourth moment.

Table 3. A listing of the 105 features selected by GA from the original 429 features (See Section 4 for feature codes).

FOS2R	FOS2G	FOS1B	FOS2B	FOS4B					
HOS3Ra	HOS4Ra	HOS5Ga	HOS6Ga	HOS11Ga	HOS13Ga	HOS6Ba	HOS8Ba	HOS11Ba	HOS6Rb
HOS11Rb	HOS13Rb	HOS14Rb	HOS5Gb	HOS7Gb	HOS8Gb	HOS10Gb	HOS14Ga	HOS4Ba	HOS9Ba
HOS11Ba	HOS12Ba	HOS3Rc	HOS7Rc	HOS9Rc	HOS1Gc	HOS6Gc	HOS7Gc	HOS9Gc	HOS10Gc
HOS11Gc	HOS12Gc	HOS1Bc	HOS2Bc	HOS10Bc	HOS5Bc	HOS1Rd	HOS4Rd	HOS5Rd	HOS10Rd
HOS1Gd	HOS3Gd	HOS5Gd	HOS13Gd	HOS14Gd	HOS2Bd	HOS7Bd	HOS8Re	HOS9Re	HOS10Re
HOS5Ge	HOS6Ge	HOS7Ge	HOS8Ge	HOS14Ge	HOS1Be	HOS2Be	HOS6Be	HOS7Be	HOS11Be
HOS1Rf	HOS8Rf	HOS14Rf	HOS1Gf	HOS1Bf	HOS3Bf	HOS5Bf	HOS1Bf	HOS12Bf	HOS4Rg
HOS3Bg	HOS7Bg	HOS8Bg	HOS9Bg	HOS10Bg	HOS11Bg	HOS12Bg	HOS14Bg	HOS1Rh	HOS5Rh
HOS9Gh	HOS13Gh	HOS1Bf	HOS5Bf	HOS8Bf					
HOG111	HOG112	HOG134	HOG135	HOG139	HOG146	HOG147	HOG149	HOG154	HOG155
HOG161	HOG163	HOG165	HOG181	HOG182					

3.2.3. HOG Features

The third group of features is extracted using a histogram of orientation gradient (HOG) as a feature descriptor [43]. The basic idea of the HOG descriptor is that features within an image can be investigated through intensity derivative distribution by observing the occurrence of gradient orientation in a local image pattern. In this work, and as in [44], the RGB retinal fundus images are cropped using the central pixel coordinate and the

radius size of the OD for every image to localize the region of the optic disc, OD, a region of size 200×200, from the RGB retinal image, and the cropped image is pictured in Figure 5.
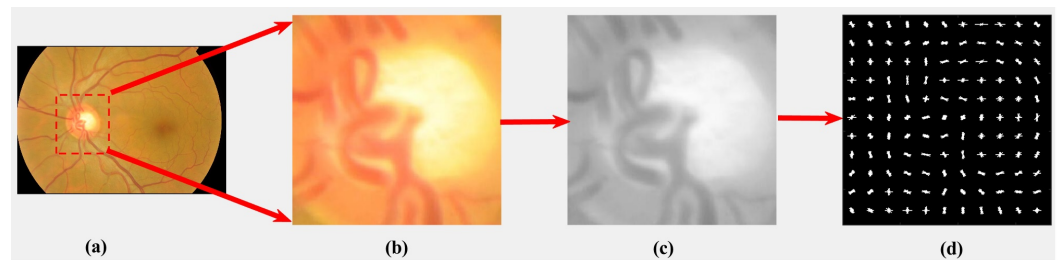

Figure 5. Stages of the HOG features. (**a**) RGB image; (**b**) cropped RGB image; (**c**) cropped gray-level image; (**d**) HOG feature image.

The cropped retinal fundus image is divided into smaller patterns called cells; a histogram of gradient directions for each cell is compiled separately for its entry pixels, then all histograms are concatenated for the descriptor. In more detail, the cropped RGB retinal fundus image is transformed to a gray-scale image; then, the cell size is indented using a window of size 3×3. Experimentally, filter kernels $[-1, 0, 1]$ and $[-1, 0, 1]^T$ are used to calculate the magnitude of the gradient orientation horizontally and vertically for each pixel within cells in the image.

Let $I(x, y)$ represents the intensity of the gray value of a pixel at a given location in image I, then the horizontal gradient, G_x, vertical gradient, G_y, and the total magnitude of gradient M, for that pixel are obtained as follows.

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y),$$
$$G_y(x, y) = I(x, y + 1) - I(x, y - 1),$$
$$M(G_x(x, y), G_y(x, y)) = \left((G_x(x, y))^2 + (G_y(x, y))^2 \right)^{0.5}.$$

On the other hand, the angle θ, representing the orientation, is given by

$$\theta = \tan^{-1} \left(\frac{G_y(x, y)}{G_x(x, y)} \right).$$

The magnitude and orientation values for each pixel in the image are used to construct the histogram. A weighted vote for each pixel in the cell is computed for the orientation, Herein, 9 histogram bins are used where the histogram channel is ultimately expanded over degrees from $0°$ to $180°$. Spatially, cells are grouped as connected patches or blocks. Afterwards, normalization using the L2 norm is carried out. To illustrate, let v be a non-normalized vector for all histograms in a given block, and let $||v||_i$ be the i-norm, for $i = 1, 2$, and ϵ be some small constant. Then the normalized vector, $\hat{v}$, is given by

$$\hat{v} = \frac{v}{(||v||_2^2 + \epsilon^2)^{0.5}}. \tag{21}$$

Finally, vectors for all normalized blocks are concatenated together, forming a HOG as a feature descriptor. It should be mentioned that the L2 norm is appropriated for each block to be more invariant to contrast, shadow, and illumination in the image. Finally, the generated feature vector with 3×3 cells and a histogram with 9 bins gives 81 features for each OD cropped image.

After the feature extraction process, a vector containing 429 features is formulated, and all feature vectors are stored in a feature matrix for further processing. Table 4 shows the different number of extracted features and their types.

Table 4. Details of the original 429 features, before reduction by GA.

Type of Feature	Image Channel			Gray-Scale	Total
	R	G	B		
FOS Features	4	4	4	-	12
HOS features	112	112	112	-	336
HOG features	-	-	-	81	81
Total	116	116	116	81	429

3.3. Min–Max Scaling for Normalization

Because features are extracted from a variety of sources, they span wide ranges. On the other hand, the performance of machine learning algorithms is sensitive to un-scaled features, which makes re-scaling them necessary. In this article, a min-max normalization is successfully used to re-scale the features, limiting their value to a common range. The limited range is useful because it makes the feature values end up with smaller standard deviations, suppressing the effect of outliers in the different features. Min–max normalization entails a linear transformation of the original features.

3.4. Feature Selection by Genetic Algorithm

Feature selection is defined as picking the most reliable and discriminative features that reduce the high dimension of feature space to a minimum. In this work, at the posterior stage of feature extraction, the curse of the dimensionality problem may appear, mostly when the number of features is immense. In our case, 429 features are extracted, which may affect not only the cost of computation in terms of training time for the model but also the final accuracy results for the classifiers. Moreover, over-fitting may be found as well due to the subset of redundant and irrelevant features [45]. So, it is critical to get rid of that subset, if it exists. To this end, we use a genetic algorithm (GA) scheme which proves to be successful, as indicated by the experimental results.

GAs consider that each individual (chromosome) in the population is described by a vector of features (genes) drawn from the entire feature set. The vector is randomly composed such that it contains a random combination of features. The fitness function is then evaluated to find the most relevant features among those that exist. The end result is used to produce successive generations. Two individuals are chosen as parents, with the criteria based on each of their fitness function values. While there are various selection systems, the roulette wheel is used in this study.

A chromosome with high fitness value is considered a candidate for the next generation. In this work, initially, a chromosome of 429 genes is created for each dataset, where the HRF comprises 45 chromosomes, with another 288 chromosomes for Dataset_ 2. Specifically, each chromosome is represented by a string of 1s and 0s, with each bit pointing to a gene. For example, a chromosome with 10 genes could be 1100101011. This means that genes 1, 2, 5, 7, 9, and 10 are included and genes 3, 4, 6, and 8 are excluded. A weighted random selection is employed in the initial population where the probability of a candidate chromosome being selected is based on its accuracy function response. So, chromosomes with high probability will have a greater chance of selection. Two parent chromosomes are selected and mated. In this process, two new chromosomes (children) inherit half of each parent's chromosome characteristics in a crossover operation. The new chromosome may be further processed using mutation where the state of some genes may be changed from 1 to 0 or the opposite. The GA effectively reduces the feature space where, out of 429 features, we end up with 105 features which are the most relevant and informative.

3.5. Classification

Once the GA selects the best discriminative features, they are used to estimate the parameters of the classification model. Re-sampling statistical techniques are regularly used to avoid variability and uncertainty in evaluating the model performance. That is

done by estimating the parameters of the model multiple times from feature samples, using a technique such as the *k*-fold crossvalidation, where the best discriminative feature set is partitioned into *k* groups. Each group can be used as a test set while the other groups are used for training.

In this article, we use the *k*-fold crossvalidation technique for both datasets. For the HRF dataset, a leave-one-out crossvalidation is applied, i.e., $k = 1$. That is good for small datasets, such as the HRF dataset, which has 45 images only. The HRF feature set is split such that, every time, 44 images are used for training and the remaining images are used for testing.

On the other hand, there are 288 images in Dataset_2. Thus, a 9-fold crossvalidation is the better choice. The features of Dataset_2 are split into 9 equal parts, 8 used for training and 1 for testing. These values lead to accurate estimation with low bias and modest variance while every fold has the same sample number of images, namely 32. The process of random partitioning is iterated 45 times for the HRF dataset and 9 times for Dataset_2, with testing sets determined at run time. The results of this exercise are reported in the tables below.

Determining the most applicable supervised ML algorithm is an overwhelming task. The most suitable option to find the proper algorithm that can achieve a satisfying performance is experimentation. After extensive analyses and trade-offs between the speed, accuracy, and complexity of several classifiers, we found that four supervised ML algorithms achieved a good performance in our problem: *k*-nearest neighbor (*k*NN) with Euclidean distance measure, naive Bayes with Gaussian distribution (NB-G), decision tree (DT), and linear discriminant analysis (LDA).

4. Results

We have tested the proposed system by experimenting extensively on two datasets, originating from three publicly available datasets. Our experiments were run on a platform including Intel(R) Core(TM) i7-9750H CPU @ 2.60 GHz, 2592 MHz, 6 Core(s), 12 logical processor(s), 16 GB memory using MATLAB 9.2, R2017b.

The experimental results, presented below, show that the combination of the three sets of features and feature selection process provides a discriminative vector of 105 features, which was later fed into each of the four mentioned classifiers for evaluation. The final results obtained from our experiments are excellent.

The first activity we performed was to use the GA to select the most relevant of the 429 originally calculated features. This activity ended up selecting the 105 features shown in Table 3, for which we use the following coding for the feature names:

- FOS*x*A is a feature of the FOS group (Section 3.2.1), with $x = 1, 2, 3, 4$ being the feature number within the group, and A $\in$ {R, G, B} being the color channel used. For example, FOS2G is feature no. 2 in the FOS feature family (i.e., variance) calculated for the green channel of the image.
- HOS*x*A*y* is a feature of the HOS group (Section 3.2.2), with $x = 1, 2, \ldots, 14$, being the feature number within the group, A $\in$ {R, G, B} being the color channel used, and $y \in$ {a, b, $\ldots$, h} being the displacement used. For example, HOS9Rd is feature no. 9 in the HOS feature family (i.e., entropy) calculated for the red channel of the image with the displacement number d in Figure 4.
- HOG*xyz* is a feature of the HOG group (Section 3.2.3), with *xy* referring to entry (x, y) in the 3×3 matrix and *z* being the bin number in the corresponding histogram. For example, HOG125 is the feature referred to by the entry in the 1st row and 2nd column of the 3×3 matrix and the 5th bin number in the histogram.

The four classifiers considered in the present work, DT, kNN, NB, and LDA, were tested separately to classify the HRF dataset and Dataset_2. For each classification experiment, the following 3×3 confusion matrix was constructed:

$$
\textbf{Actual} \quad
\begin{bmatrix}
Class & \textbf{1} & \textbf{2} & \textbf{3} \\
\textbf{1} & A_{11} & A_{12} & A_{13} \\
\textbf{2} & A_{21} & A_{22} & A_{23} \\
\textbf{3} & A_{31} & A_{32} & A_{33}
\end{bmatrix}
$$

where $1 = $ DR, $2 = $ glaucoma and $3 = $ normal, and A_{ij} is the count of cases of class i that are classified as class j. This matrix is used to obtain five metrics: precision (Pr), sensitivity (Se), specificity (Sp), F1 score ($F1$), accuracy (Acc), and AUC, as follows.

$$Pr = \frac{TP}{TP+FP}, \tag{22}$$

$$Se = \frac{TP}{TP+FN}, \tag{23}$$

$$Sp = \frac{TN}{TN+FP}, \tag{24}$$

$$F1 = \frac{2TP}{2TP+FP+FN}, \tag{25}$$

$$Acc = \frac{TP+TN}{TP+FP+TN+FN}. \tag{26}$$

$$AUC = 0.5\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right). \tag{27}$$

where

- FP, false positives, the number of images belonging to the class but incorrectly labeled as not.
- FN, false negatives, the number of images belonging to the class but incorrectly labeled as not.
- TP, true positive, the number of images belonging to the class and correctly labeled as such.
- TN, true negative, the number of images not belonging to the class and correctly labeled as such.

The area under the receiver operating characteristic curve (ROC-AUC) is a useful and broadly used measure to evaluate the performance of binary classifiers. Recently [46], it has been extended to multi-class classifiers, where for each class, the other classes are simply lumped together as the second class.

Furthermore, we obtained the *Overall Se*, *Overall Sp*, and *Overall Acc* parameters, which are calculated from the aggregation of all three classes for each dataset as follows.

$$Overall\ Se = \frac{\sum_{i=1}^{3} TP_i}{\sum_{i=1}^{3}(TP_i+FN_i)} \tag{28}$$

$$Overall\ Sp = \frac{\sum_{i=1}^{3} TN_i}{\sum_{i=1}^{3}(TN_i+FP_i)} \tag{29}$$

$$Overall\ Acc = \frac{\sum_{i=1}^{3}(TP_i+TN_i)}{\sum_{i=1}^{3}(TP_i+FP_i+TN_i+FN_i)} \tag{30}$$

where $i = 1, 2, 3$ is the class label with $1 = $ DR, $2 = $ glaucoma and $3 = $ normal.

Unlike most previous studies, we evaluate the classifiers of the proposed system in two ways. First, the performance metrics (22)–(27) are used to measure the classification performance for every classifier for each class (DR, glaucoma, and normal). Second, the performance metrics (28)–(30) are used to measure the overall classification performance for all classes.

Tables 5 and 6 show the results regarding precision, sensitivity, specificity, F1 score, and accuracy metrics for the two datasets. In all tables, the proposed system results are shown in detail for each class and for each dataset. From the tables, it is obvious that the

proposed system functions remarkably well. Tables 5 and 6 represent the classification performance of the four classifiers when tested on the HRF features and Dataset_2 features, respectively. Each row represents results for only one class.

Table 5. Classification performance per classifier for every class in the HRF dataset.

Classifier	Class	*Pr*	*Se*	*Sp*	*F1*	*Acc*	*AUC*
	DR	0.7000	0.7700	0.8334	0.7368	0.8148	0.8055
DT	Glaucoma	0.8000	0.7273	0.8750	0.7619	0.8148	0.8011
	Normal	0.7000	0.7000	0.8334	0.7000	0.7857	0.7666
	DR	0.5000	0.5555	0.7619	0.5263	0.7000	0.6587
*k*NN	Glaucoma	0.8000	0.6667	0.8667	0.7272	0.7778	0.7666
	Normal	0.8887	0.8889	0.8667	0.8421	0.8752	0.8777
	DR	0.4000	0.5000	0.7142	0.4444	0.6551	0.6071
NB	Glaucoma	0.8001	0.5714	0.8467	0.6667	0.7037	0.7087
	Normal	0.7000	0.8750	0.8000	0.7778	0.8260	0.8375
	DR	0.8000	0.7273	0.8947	0.7619	0.8334	0.8110
LDA	Glaucoma	1	1	1	1	1	1
	Normal	0.7000	0.7776	0.8571	0.7368	0.8334	0.8178

Table 6. Classification performance per classifier for every class in Dataset_2.

Cassifier	Class	*Pr*	*Se*	*Sp*	*F1*	*Acc*	*AUC*
	DR	1	0.9090	1	0.9523	0.9667	0.9545
DT	Glaucoma	1	1	1	1	1	1
	Normal	0.9000	1	0.9524	0.9473	0.9667	0.9761
	DR	0.7000	0.8750	0.8636	0.7778	0.8667	0.8693
kNN	Glaucoma	1	0.9000	1	0.9524	0.9629	0.9545
	Normal	0.9001	0.8181	0.9445	0.8571	0.8965	0.8813
	DR	0.7000	1	0.8695	0.8235	0.9000	0.9347
NB	Glaucoma	1	0.9090	1	0.9523	0.9642	0.9545
	Normal	1	0.8333	1	0.9090	0.9310	0.9166
	DR	0.8000	1	0.9000	0.8888	0.9285	0.9500
LDA	Glaucoma	0.9000	0.8182	0.9445	0.8571	0.8965	0.8813
	Normal	0.9000	0.8182	0.9444	0.8571	0.8965	0.8813

Table 7 represents the classification performance for the two datasets as overall sensitivity (*Overall Se*), overall specificity (*Overall Sp*), and overall accuracy (*Overall Acc*). Clearly, the proposed system performs extremely well.

Table 7. Overall classification performance for the two datasets.

Dataset	Classifier	*Overall Se*	*Overall Sp*	*Overall Acc*
	DT	0.7400	0.0740	0.7400
HRF dataset	*k*NN	0.6600	0.6600	0.6600
	NB	0.6400	0.6400	0.6400
	LDA	0.8300	0.8300	0.8300
	DT	0.9667	0.9667	0.9667
Dataset_2	*k*NN	0.8670	0.8670	0.8670
	NB	0.9000	0.9000	0.9000
	LDA	0.9360	0.9360	0.9360

Tables 8 and 9 show the best and worst results for each class in the HRF dataset and Dataset_2.

Table 8. The best and the worst classification results for the HRF dataset.

Class		*Pr*	*Se*	*Sp*	*F1*	*Acc*	*AUC*
DR	*Best*(max)	0.8000	0.7700	0.8947	0.7619	0.8334	0.8110
	Worst(min)	0.4000	0.5000	0.7142	0.4444	0.6551	0.6071
Glaucoma	*Best*(max)	1.000	1.000	1.000	1.000	1.000	1
	Worst(min)	0.8000	0.5714	0.8467	0.6667	0.7037	0.7087
Normal	*Best*(max)	0.8887	0.8889	0.8421	0.8752	0.8752	0.8777
	Worst(min)	0.7000	0.7000	0.8000	0.7000	0.7857	0.7666

Table 9. The best and the worst classification results for Dataset_2.

Class		*Pr*	*Se*	*Sp*	*F1*	*Acc*	*AUC*
DR	*Best*(max)	1.000	1.000	1.000	0.9523	0.9667	0.9545
	Worst(min)	0.7000	0.8750	0.8636	0.7778	0.8667	0.8693
Glaucoma	*Best*(max)	1.000	1.000	1.000	1.000	1.000	1.000
	Worst(min)	0.9000	0.8082	0.9445	0.8571	0.8965	0.8965
Normal	*Best*(max)	1.000	1.000	1.000	1.000	1.000	0.9761
	Worst(min)	0.9000	0.8181	0.9444	0.8571	0.8965	0.8813

Table 10 shows a recently published system proposed by Chelaramani [17], which is based on automatic deep learning and performs five-class classification, in contrast to our system which is based on hand-crafted features, conventional ML algorithms, and three-class classification. The common factor between the two systems is the complete results provided for every individual class in terms of precision, sensitivity, specificity, *F1* score, and accuracy.

Table 10. Comparison between the results of the proposed model and the state-of-the-art method as an automatic deep feature extraction.

Method	Class	*Pr*	*Se*	*Sp*	*F1*	*Acc*
	Melanoma	0.8193	0.9701	0.9511	0.8883	0.9479
	Glaucoma	0.8388	0.9365	0.9619	0.8850	0.9510
[17]	AMD	0.8090	0.8317	0.9516	0.8202	0.9196
	DR	0.8829	0.7211	0.9711	0.7938	0.9047
	Normal	0.7788	0.7136	0.9296	0.7448	0.8894
	DR	1.000	0.9090	1.000	0.9523	0.9667
The proposed model via DT classifier	Glaucoma	1.000	1.000	1.000	1.000	1.000
	Normal	0.9000	1.000	0.9524	0.9473	0.9667

5. Discussion

Rather than classify the severity level of a certain disease, we have opted in this work to focus on distinguishing between three classes, namely DR, glaucoma, and normal. The reason why our system performed remarkably well is multi-faceted. First, the choice of the features (of the three groups) plays an important role. For example, we used entropy and ASM as features because the former is a measure of disorder, just as the latter is a measure of sameness and consistency in the image. The higher the entropy, the higher the disorder (pathological pattern-like DR) in the image, and vice versa. So, entropy is a good biomarker for the degree of DR in an image. On the other hand, ASM is a good biomarker for normal and DR images. In addition, energy is a good biomarker for the local uniformity or homogeneity that can differentiate between DR and glaucoma. Furthermore, contrast is a good biomarker for DR, particularly in the green channel. So the obtained GLCM in general provides good biomarkers for discerning between DR and normal images. Meanwhile, the HOG features are a good biomarker for glaucoma through cropping the OD of the pixels. During the feature extraction and selection procedure, we observed

that the combination of the three feature sets is wholly leveraged to discern within every pathological pattern in the retinal image.

The classification results are provided comprehensively in Section 4, not only as overall metrics for all three classes (DR, glaucoma, and normal) but also for each class individually. The results show that although the same set of features is used with all four classifiers considered in the present work, the classification performance of each classifier is to some extent dataset dependent. This could be due to the extraordinary complexity of the retinal image. In our situation, the DT classifier performed better with Dataset_2 than with the HRF dataset. For example, DT achieved, with Dataset_2, an accuracy of 95.45% for DR, 100% for glaucoma, and 97.61% for normal conditions. By contrast, the LDA classifier performed better with the HRF dataset than with Dataset_2. Section 4 provides all the metric values for all four classifiers and both datasets.

The classification performance is also to some extent class dependent. That is, no classifier is good at detecting all diseases. Again, this can be attributed to the immense complexity of the retinal image. For example, let us consider the HRF dataset. The best results for the DR class are achieved by the LDA classifier under all metrics, except for Se, in which DT is the winner. At the same time, the worst DR results are obtained by the NB classifier. Concerning glaucoma, the best results are shown by the LDA classifier and the worst by the NB classifier. For the normal cases, the best results are obtained by the kNN classifier and the worst results by DT.

Based on the results, displayed in Section 4, the proposed system is by and large highly successful in identifying the three possible classes: DR, glaucoma, and normal. Concerning Dataset_2, the best results for DR, glaucoma, and normal conditions are achieved by the DT classifier, while the worst results are obtained by kNN for the DR class for Pr, F1, and AUC and for the normal class for SE and F1 metrics, respectively. The worst results for the glaucoma class are obtained in terms of Sp, F1, and AUC using the LDA classifier, while the worst results for the normal class are obtained in terms of Sp and AUC using the LDA classifier. According to the Diabetes Society in the United Kingdom, it is recommended that the sensitivity measure of any CAD system should exceed 80% [24] to be permitted for use. Our system clearly surpasses this threshold by a wide margin.

Our system shows a remarkable performance with low cost compared to automatic feature selection using time-consuming deep learning, which transfers the input (image) pixels into multiple successive layers for the purpose of learning the features. However, the proposed system has a limitation in the event that the image has pathological manifestations of both DR and glaucoma at the same time. We did not face this event because no image in the used datasets had this issue. It is considered an open point for future study if a suitable image dataset is found.

6. Conclusions

In this article, we have proposed an accurate retinal CAD system. The accuracy of the proposed system has been confirmed by extensive experimental work that was conducted on publicly available real datasets. The system comprises three elements that collectively contribute to its success. The first element is choosing three different sets of features that skillfully identify and discriminate between the pathological features, despite their slight and subtle differences. The second element is designing a modified GA-based scheme to reduce the highly dimensional feature space of the retinal image by selecting only the most relevant and essential features. The third element is the sampling technique used for each dataset.

The experimental results show that the system succeeds both absolutely and compared with competitive systems. More importantly, they refute the idea that the only way to achieve high classification performance in a system is to resort to automatic feature extractor systems (deep learning). The proposed system outperforms systems based on deep learning at a fraction of the time cost and of the number of training images.

It is evident from the experimental results that the 105 features selected by the GA from the original 429 calculated features are highly successful in discriminating between the three classes we consider: DR, glaucoma, and normal. This degree of success, however, depended on both the classifier and the dataset. The features classified DR best in Dataset_2 using the DT classifier. However, the same features classified DR best in the HRF dataset using the LDA classifier.

Experimentally, it is noted that despite the relatively small number of images used, the results are promising, and this proves the good choice of the features and the effectiveness of the modified genetic algorithm as a feature selection algorithm. However, using more images is sure to improve system performance further. This conclusion is confirmed by obtaining better classification for Dataset_2 than the HRF dataset, given that the former dataset is larger than the latter.

The proposed system can serve as a supportive diagnostic tool that can easily be installed and used in healthcare units, small polyclinics, and remote villages and communities, especially in developing countries where ophthalmologists are in short supply. Even in the presence of ophthalmologists, it can help them make the right diagnosis decision quickly. In the future, this system can be configured as a personal retinal healthcare commodity that can be used directly by individuals before seeking the services of a medical doctor.

References

1. Li, J.; Hu, Q.; Imran, A.; Zhang, L.; Yang, J.J.; Wang, Q. Vessel recognition of retinal fundus images based on fully convolutional network. In Proceedings of the IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 July 2018; Volume 2, pp. 413–418. [CrossRef]
2. WHO. Blindness and Vision Impairment. 2020. Available online: http://www.who.int/health-topicsblindness-and-vision-loss#tab=tab_1 (accessed on 7 May 2020).
3. Sarhan, A.; Rokne, J.; Alhajj, R. Approaches for early detection of glaucoma using retinal images: A performance analysis. In *Data Management and Analysis. Studies in Big Data Analysis*; Alhajj, R., Moshirpour, M., Far, B., Eds.; Springer: Cham, Swtzerland, 2020; Volume 65. pp. 213–238. [CrossRef]
4. Maheshwari, S.; Pachori, R.B.; Acharya, U.R. Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images. *IEEE J. Biomed. Health Inform.* **2016**, *21*, v803–v813. [CrossRef] [PubMed]
5. Sundaram, R.; Ravichandran, K.S.; Jayaraman, P. Extraction of blood vessels in fundus images of retina through hybrid segmentation approach. *Mathematics* **2019**, *7*, 169. [CrossRef]
6. International Diabetes Foundation (IDF). Diabetes Eye Health: A Guide for Health Professionals. 2020. Available online: https://idf.org/our-activities/care-prevention/eye-health/eye-health-guide.html (accessed on 7 June 2020).

7. Tham, Y.C.; Li, X.; Wong, T.Y.; Quigley, H.A.; Aung, T.; Cheng, C.Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis. *Ophthalmology* **2014**, *121*, 2081–2090. [CrossRef] [PubMed]

8. Samuel, P.M.; Veeramalai, T. Multilevel and multiscale deep neural network for retinal blood vessel segmentation. *Symmetry* **2019**, *11*, 946. [CrossRef]

9. Dupont, G.; Kalinicheva, E.; Sublime, J.; Rossant, F.; Paques, M. Analyzing age-related macular degeneration progression in patients with geographic atrophy Using Joint autoencoders for unsupervised change detection. *J. Imaging* **2020**, *6*, 57. [CrossRef]

10. Kamran, S.A.; Saha, S.; Sabbir, A.S.; Tavakkoli, A. Optic-net: A novel convolutional neural network for diagnosis of retinal diseases from optical tomography images. In Proceedings of the 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 964–971.

11. Qureshi, I.; Ma, J.; Shaheed, K. A hybrid proposed fundus image enhancement framework for diabetic retinopathy. *Algorithms* **2019**, *12*, 14. [CrossRef]

12. Tamim, N.; Elshrkawey, M.; Abdel Azim, G.; Nassar, H. Retinal blood vessel segmentation using hybrid features and multi-layer perceptron neural networks. *Symmetry* **2020**, *12*, 894. [CrossRef]

13. Tchinda, B.S.; Tchiotsop, D.; Noubom, M.; Louis-Dorr, V.; Wolf, D. Retinal blood vessels segmentation using classical edge detection filters and the neural network. *Inform. Med. Unlocked* **2021**, *23*, 100521–100529. [CrossRef]

14. Riaz, H.; Park, J.; Choi, H.; Kim, H.; Kim, J. Deep and densely connected networks for classification of diabetic retinopathy. *Diagnostics* **2020**, *10*, 24. [CrossRef]

15. Hagiwara, Y.; Koh, J.E.W.; Tan, J.H.; Bhandary, S.V.; Laude, A.; Ciaccio, E.J.; Tong, L.; Acharya, U.R. Computer-aided diagnosis of glaucoma using fundus images: A review. *Comput. Methods Programs Biomed.* **2018**, *165*, 1–12. [CrossRef]

16. Koh, J.E.; Acharya, U.R.; Hagiwara, Y.; Raghavendra, U.; Tan, J.H.; Sree, S.V.; Bhandary, S.V.; Rao, A.K.; Sivaprasad, S.; Chua, K.C.; et al. Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies. *Comput. Biol. Med.* **2017**, *84*, 89–97. [CrossRef] [PubMed]

17. Chelaramani, S.; Gupta, M.; Agarwal, V.; Gupta, P.; Habash, R. Multi-task knowledge distillation for eye disease prediction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Virtual Conference, 5–9 January 2021; pp. 3983–3993.

18. Ganesan, K.; Martis, R.J.; Acharya, U.R.; Chua, C.K.; Min, L.C.; Ng, E.Y.K.; Laude, A. Computer-aided diabetic retinopathy detection using trace transforms on digital fundus images. *Med. Biol. Eng. Comput.* **2014**, *52*, 663–672. [CrossRef] [PubMed]

19. Mookiah, M.R.K.; Acharya, U.R.; Fujita, H.; Koh, J.E.; Tan, J.H.; Chua, C.K.; Bhandary, S.V.; Noronha, K.; Laude, A.; Tong, L. Automated detection of age-related macular degeneration using empirical mode decomposition. *Knowl.-Based Syst.* **2015**, *89*, 654–668. [CrossRef]

20. Bhandary, S.V.; Rao, K.A. Automated screening system for retinal health using bi-dimensional empirical mode decomposition and integrated index. *Comput. Biol. Med.* **2018**, *75*, 54–62.

21. Koh, J.E.; Ng, E.Y.; Bhandary, S.V.; Laude, A.; Acharya, U.R. Automated detection of retinal health using PHOG and SURF features extracted from fundus images. *Appl. Intell.* **2018**, *48*, 1379–1393. [CrossRef]

22. Koh, J.E.; Ng, E.Y.; Bhandary, S.V.; Hagiwara, Y.; Laude, A.; Acharya, U.R. Automated retinal health diagnosis using pyramid histogram of visual words and Fisher vector techniques. *Comput. Biol. Med.* **2018**, *92*, 204–209. [CrossRef]

23. Pires, R.; Jelinek, H.F.; Wainer, J.; Valle, E.; Rocha, A. Advancing bag-of-visual-words representations for lesion classification in retinal images. *PLoS ONE* **2014**, *9*, e96814. [CrossRef]

24. Abbas, Q.; Fondon, I.; Sarmiento, A.; Jiménez, S.; Alemany, P. Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. *Med. Biol. Eng. Comput.* **2017**, *55*, 1959–1974. [CrossRef]

25. Sengupta, S.; Singh, A.; Leopold, H.A.; Gulati, T.; Lakshminarayanan, V. Ophthalmic diagnosis using deep learning with fundus images—A critical review. *Artif. Intell. Med.* **2020**, *102*, 101758–101767. [CrossRef]

26. Quellec, G.; Charrière, K.; Boudi, Y.; Cochener, B.; Lamard, M. Deep image mining for diabetic retinopathy screening. *Med. Image Anal.* **2017**, *39*, 178–193. [CrossRef]

27. Gheisari, S.; Shariflou, S.; Phu, J.; Kennedy, P.J.; Agar, A.; Kalloniatis, M.; Golzan, S.M. A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Sci. Rep.* **2021**, *11*, 1945. 1554-4. [CrossRef]

28. Zhao, Z.; Chopra, K.; Zeng, Z.; Li, X. Sea-Net: Squeeze-and-excitation attention net for diabetic retinopathy grading. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2496–2500.

29. Li, Z.; He, Y.; Keel, S.; Meng, W.; Chang, R.T.; He, M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* **2018**, *125*, 1199–1206. [CrossRef]

30. Jerith, G.G.; Kumar, P.N. Recognition of Glaucoma by means of gray wolf optimized neural network. *Multimed. Tools Appl.* **2020**, *79*, 10341–10361. [CrossRef]

31. Maqsood, S.; Damaševičius, R.; Maskeliūnas, R. Hemorrhage detection based on 3D CNN deep learning framework and feature fusion for evaluating retinal abnormality in diabetic patients. *Sensors* **2021**, *21*, 3865. [CrossRef]

32. Ramasamy, L.K.; Padinjappurathu, S.G.; Kadry, S.; Damaševičius, R. Detection of diabetic retinopathy using a fusion of textural and ridgelet features of retinal images and sequential minimal optimization classifier. *PeerJ Comput. Sci.* **2021**, *7*, e456–e477. [CrossRef]

33. Zhou, M.; Jin, K.; Wang, S.; Ye, J.; Qian, D. Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Trans. Biomed. Eng.* **2017**, *65*, 521–527. [CrossRef] [PubMed]

34. Budai, A.; Bock, R.; Maier, A.; Hornegger, J.; Michelson, G. Robust vessel segmentation in fundus images. *Int. J. Biomed. Imaging* **2013**, *2013*, 154860. [CrossRef]

35. Kaggle. Diabetic Retinopathy Detection (Data). 2015. Available online: https://www.kaggle.com/c/diabetic-retinopathy-detection/data (accessed on 7 May 2020).

36. Almazroa, A.; Alodhayb, S.; Osman, E.; Ramadan, E.; Hummadi, M.; Dlaim, M.; Alkatee, M.; Raahemifar, K.; Lakshminarayanan, V. Retinal fundus images for glaucoma analysis: The RIGA dataset. In Proceedings of the Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications, Houston, TX, USA, 6 March 2018. [CrossRef]

37. Balasubramanian, T.; Krishnan, S.; Mohanakrishnan, M.; Rao, K.R.; Kumar, C.V.; Nirmala, K. HOG feature based SVM classification of glaucomatous fundus image with extraction of blood vessels. In Proceedings of the IEEE Annual India Conference (INDICON), Bangalore, India, 16–18 December 2016; pp. 1–4. [CrossRef]

38. Löfstedt, T.; Brynolfsson, P.; Asklund, T.; Nyholm, T.; Garpebring, A.A. Gray-level invariant Haralick texture features. *PLoS ONE* **2019**, *14*, e0212110. [CrossRef] [PubMed]

39. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, 610–621. [CrossRef]

40. Sahlol, A.T.; Abdeldaim, A.M.; Hassanien, A.E. Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. *Soft Comput.* **2019**, *23*, 6345–6360. [CrossRef]

41. Soh, L.K.; Tsatsoulis, C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 780–795. [CrossRef]

42. Hoaglin, D.; Mosteller, F.; Tukey, J. *Understanding Robust and Exploratory Data Analysis*; Wiley: Hoboken, NJ, USA, 2000.

43. Rosyidi, L.; Prasetyo, A.; Romadhon, M.S. Object tracking with raspberry Pi using Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM). In Proceedings of the 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 24–26 June 2020; pp. 1–6. [CrossRef]

44. Jeena, R.S.; Kumar, A.S.; Mahadevan, K. A novel method for stroke prediction from retinal images using HoG approach. In Proceedings of the International Symposium on Signal Processing and Intelligent Recognition Systems SIRS 2018: Advances in Signal Processing and Intelligent Recognition Systems, Bangalore, India, 19–22 September 2018; pp. 137–146.

45. Cekik, R.; Uysal, A.K. A novel filter feature selection method using rough set for short text data. *Expert Syst. Appl.* **2020**, 113691. [CrossRef]

46. Idrees, F.; Rajarajan, M.; Conti, M.; Rahulamathavan, Y.; Chen, T.P. PIndroid: A novel Android malware detection system using ensemble learning methods. *Comput. Secur.* **2017**, *68*, 36–46. [CrossRef]

Short Biography of Authors

Nasser Tamim is a lecturer at the Computer and Information section, Deanship of Community Services, and continuing education at Jazan University, KSA, since 2011. He obtained his M.sc in Artificial Intelligent and Expert Systems from the Computer Science Department Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt, in 2011. His current research interests are Pattern Recognition, Machine Learning (Deep Learning), Pattern recognition, and medical image processing.

M. Elsharkawey received his B.Sc. in Electrical engineering from the Military Technical College, Cairo, Egypt, in 1987. He received his M. Sc. in Computer Engineering from the Faculty of Engineering, Al Azhar University in June 2007, and received his Ph.D in Network Security from the Faculty of Computers & Informatics, Cairo University, in June Cairo in 2012. He is currently an associate professor in the Faculty of Computers & Informatics, Suez Canal University. Ismailia, Egypt. His current research interests are Computer networking & cloud computing, simulation, and Image Processing.

Hamed Nassar is full professor in the Computer Science Department, Suez Canal University, Egypt, since 2004. He obtained his BSc in Electrical Engineering, from Ain Shams University, Egypt, in 1979, and obtained his MSc in Electrical Engineering and PhD in Computer Engineering, from the New Jersey Institute of Technology, USA, in 1985 and 1989, respectively. Prof. Nassar has taught computer science and engineering courses in USA, Egypt, Saudi Arabia, and Lebanon. He has published in reputed international journals and conferences. His research interests include wireless communications, cloud computing, machine learning, and mathematical modelling of ICT systems mainly mainly queueing theory and stochastic geometry.