Med Biol Eng Comput (2017) 55:1959–1974
DOI 10.1007/s11517-017-1638-6

ORIGINAL ARTICLE

Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features

Qaisar Abbas[1] · **Irene Fondon**[2] · **Auxiliadora Sarmiento**[2] · **Soledad Jiménez**[3] · **Pedro Alemany**[3]

Received: 6 September 2016 / Accepted: 13 March 2017 / Published online: 28 March 2017

Abstract Diabetic retinopathy (DR) is leading cause of blindness among diabetic patients. Recognition of severity level is required by ophthalmologists to early detect and diagnose the DR. However, it is a challenging task for both medical experts and computer-aided diagnosis systems due to requiring extensive domain expert knowledge. In this article, a novel automatic recognition system for the five severity level of diabetic retinopathy (SLDR) is developed without performing any pre- and post-processing steps on retinal fundus images through learning of deep visual features (DVFs). These DVF features are extracted from each image by using color dense in scale-invariant and gradient location-orientation histogram techniques. To learn these DVF features, a semi-supervised multilayer deep-learning algorithm is utilized along with a new compressed layer and fine-tuning steps. This SLDR system was evaluated and compared with state-of-the-art techniques using the measures of sensitivity (SE), specificity (SP) and area under the receiving operating curves (AUC). On 750 fundus images (150 per category), the SE of 92.18%, SP of 94.50% and AUC of 0.924 values were obtained on average. These results demonstrate that the SLDR system is appropriate for early detection of DR and provide an effective treatment for prediction type of diabetes.

Keywords Diabetic retinopathy · Retinal fundus images · Computer-aided diagnosis · Deep learning · Deep visual feature · Color dense SIFT features · Gradient location-orientation histogram

1 Introduction

Diabetes is the fourth leading [41] cause of death. According to estimation, the frequency of diabetes [38] will increase to 7.7% by 2030, while there are currently 336 million people who suffer from this disease. The diabetic retinopathy (DR) is one of the complications of diabetes, which is an ultimate cause of blindness. The DR has two main types of diabetes. In patients with type I diabetes, 15 years after the initial diagnosis, 75–95% of patients show some signs of retinopathy. In type 2 patients, the prevalence is around 60% after 16 years of the disease. The DR affects up to 80% of the patients who suffered from diabetes for 10 years or more. However, with an early diagnosis and continuous monitoring of DR, the progression of eyesight loss can be reduced.

The diabetic retinopathy study (DRS) demonstrated that the effectiveness of scatter laser photocoagulation can help to reduce the risk of severe [10] visual loss in 90%. In the diagnosis of DR, clinical experts such as ophthalmologists are currently using non-mydriatic cameras to get images along with a computer-aided diagnosis (CAD) system to assess or grade a severity level of the patient's diabetes [29]. Those assessments are made based on several DR-related lesions. These DR-related lesions are unhealthy objects caused by the leakage of blood and fluid on the retina. These DR-related lesions are presented in retinal images such as microaneurysms (MAs), hemorrhages (HRs), hard exudates (HE) and cotton wool spots

✉ Qaisar Abbas
qaisarabbasphd@gmail.com; qaabbas@imamu.edu.sa

[1] College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

[2] Escuela Superior de Ingenieros, Universidad de Sevilla, Camino de los Descubrimientos, s/n, 41092 Seville, Spain

[3] Surgery Department and Glaucoma Unity, University Hospital Puerta del Mar, Cádiz, Spain

(CWS) [7]. It was very difficult for clinical experts and CAD systems to identify and count DR-related lesions because these have different appearances on images. To visually show an example, Fig. 1 contains these DR-related lesions.

Diabetic retinopathy (DR) is classified into two major classes such as non-proliferative DR (NPDR) and proliferative DR (PDR) [7]. The NPDR is further subdivided according to the level of damage in the retina into three different stages such as mild, moderate and severe. According to clinical experts, the color fundus images should be classified into five categories such as normal, mild NPDR, moderate NPDR, severe NPDR and PDR. In the case of a normal image, there is no observation of DR-related lesions and no leakage of blood vessels. However, mild NPDR is the stage in which at least one MA is presented with or without mild HRs, HEs and CWS. In the case of moderate NPDR, there must exist hemorrhages, microaneurysms, hard exudates and cotton wool spots less than in severe stage. Severe NPDR is characterized by one of these findings: numerous HEs and MAs in 4 quadrants, venous beading in at least 2 quadrants and/or intraretinal microvascular abnormalities (IRMA) in at least 1 quadrant or severe in less of 4 quadrants, beaded vessels in 1 quadrant and other moderate vessel abnormalities. Severe NPDR diagnosis is based on the rule of 4-2-1, that is: MAs or severe HRs in 4 quadrants, beaded vessels in at least 2 quadrants and other moderate or severe vessel abnormalities in at least 1 quadrant [22], whereas the PDR is highly characterized by a proliferation of optic disk or retinal neovascularization and/or preretinal vitreous, hemorrhages, vessels and large HRs [9]. An example of each class of DR is shown in Fig. 2.

Identification of these five stages is a tedious and difficult task for ophthalmologists due to the presence of a variety of DR-related lesions along with their complicated structure. Also, the human grading of DR is subjective and depends on the reader's experience. As a result, there is a dire need for automatic CAD system to make the grading of larger datasets as much fine as possible [17, 30] and to reduce the inter-reader variability. In past studies, the computer-aided diagnosis (CAD) systems were proposed to address this issue [2, 8, 11, 18, 21, 31, 34, 36]. In fact, the performance of automatic computerized systems on diagnosing DR task has been found similar to human experts. However, those CAD systems for assessment of severity level depend on DR-related lesions, which is highly depended on technical domain expert knowledge and pre- or post-processing steps. To extract the features from images, those methods focused

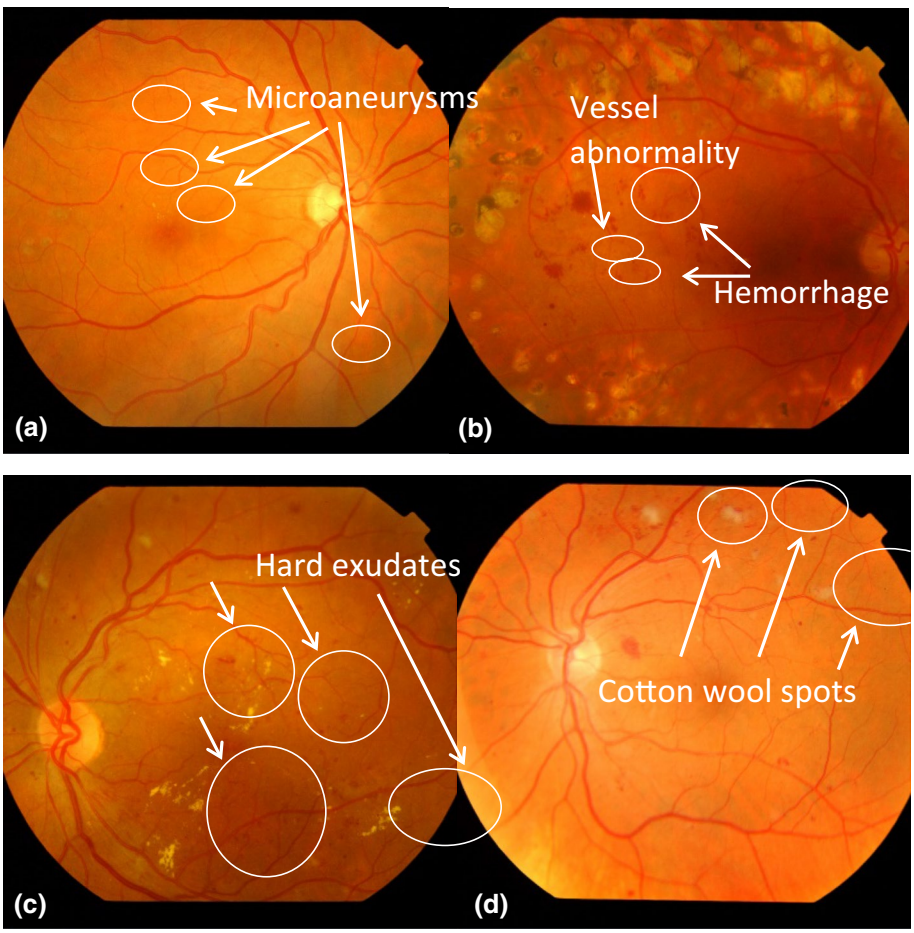

Fig. 1 Example of DR-related lesions of diabetic retinopathy where each fundus image represents **a** normal, **b** microaneurysms, **c** hemorrhages and vessel abnormality, **d** hard exudates and **e** cotton wool spots

on particular elements such as microaneurysms (MAs), hemorrhages (HRs), hard exudates (HE), cotton wool spots (CWS) and vascular network. Those systems were developed based on different publicly available datasets using various evaluation measures. The STARE [19], DRIVE [1], MESSIDOR [26] and private databases were mainly utilized to validate the obtained results. Some international competitions were also arranged to annotate these datasets. However, there were a few authors that proposed systems to detect different DR-related lesions [4, 25, 27, 28, 32, 33, 40, 42, 43] together with the limited severity-level category. The study of current state-of-the-art CAD systems is presented in Table 1. In this table, the CAD systems show current failures of different techniques remaining quite a number of problems without solution. Those CAD systems were briefly explained in the subsequent paragraphs.

In [32], DR-related lesions were detected using bag-of-visual-words (BOW) [5, 39], semi-soft coding and max pooling techniques. They did not evaluate diabetic retinopathy severity level and detected only the presence or absence of DR-related lesions. The author achieved a 94.2% of the area under the receiver operating characteristic curve (AUC), whereas in [28], a system was proposed to differentiate only between normal and diabetic retinopathy using a feature extraction step and reported 94.17% of AUC. They extracted features from retinal images using the discrete wavelet transform (DWT) and the stationary wavelet transform (SWT) coefficients along with a selection of the top-rank algorithm and AdaBoost methods. In [42], Welikala et al. used many 4D feature vectors to feed a support vector machine classifier (SVM) to recognize only PDR. They obtained a sensitivity (SE) of 91% and specificity (SP) of 96%. Prakash et al. in [33] tried to classify the fundus images into the five levels of DR by a dual-stage classification with the neural network (NN) and support vector machine (SVM). First, they performed an optic disk and blood vessels segmentation using modified region growing (MRG) method and morphological operations, respectively. Then, the retinal images were classified into normal and abnormal using NN. For this classification, mean, variance, entropy and area were extracted from the segmented optic disk, and also mean, variance, entropy, area, diameter and number of regions were extracted from the segmented blood vessels of retinal images. After that, the abnormal images were classified into the four severity stages by using SVM classifier. They achieved an accuracy of 80%. Although the method was computationally expensive. In [4], Akram et al. presented a system for grading the retina image as normal or as belonging to any one of the three stages of NPDR. The PDR stage was not considered in this work. The proposed system consisted of

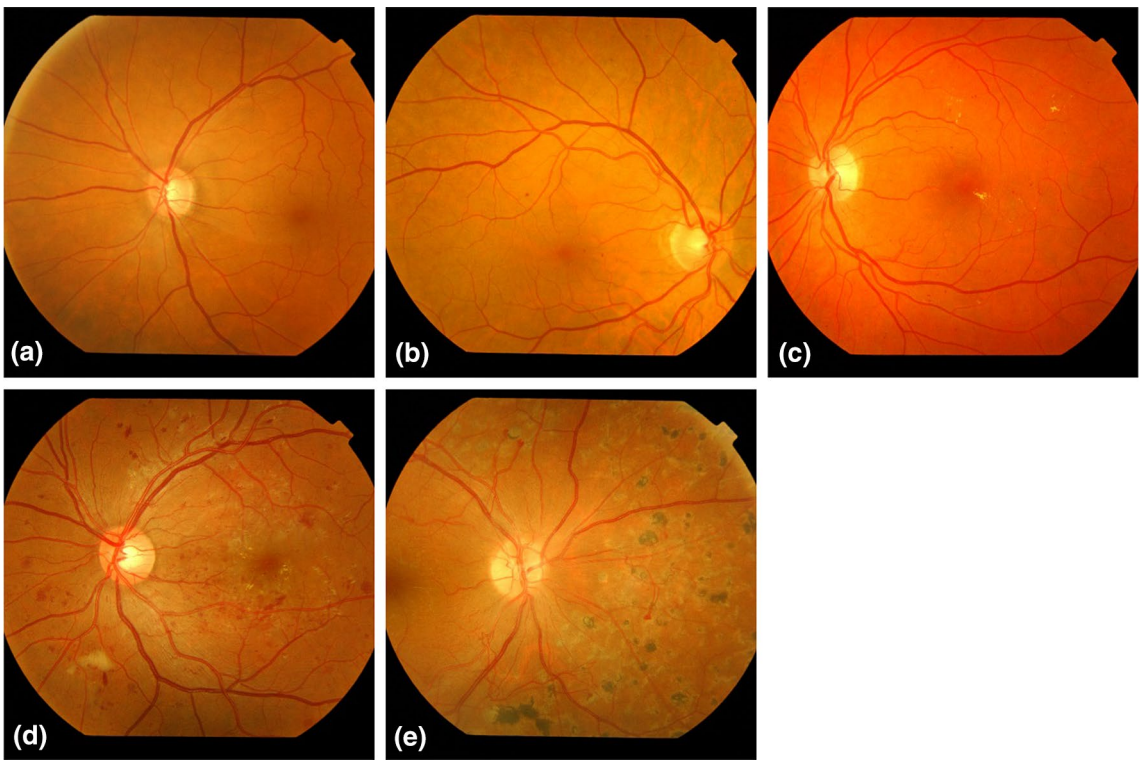

Fig. 2 Example of five severity levels and DR-related lesions of diabetic retinopathy where each fundus image represents **a** normal, **b** mild NPDR, **c** moderate NPDR, **d** severe NPDR, **e** proliferative PDR

Table 1 Current state-of-the-art computer-aided diagnosis (CAD) systems for recognition of limited category of severity level on retinograph images

Cited papers	Diabetic severity level	Methodology	Results
Pires et al. [32]	Detection of DR-related lesions	Bag-of-visual-words using speed-up robust features (SURF) along with semi-soft coding and max pooling	AUC: 94.2
Mookiah et al. [28]	Differentiation between normal and diabetic retinopathy	Features extracted using discrete wavelet transform (DWT) and stationary wavelet transform (SWT) coefficients. It classifies these features using top-rank and AdaBoost machine learning methods	ACC: 94.17
Welikala et al. [42]	Detection of PDR	Numerous 4D feature vectors fed into support vector machine (SVM) classifier	SN: 91, SP: 96
Prakash et al. [33]	Grading images into five severity levels	Pre-processing, segmentation, feature extraction and classification phases using support vector machine (SVM) and neural network (NN) algorithms	ACC: 80
Akram et al. [4]	Grading images into normal mild NPDR, moderate NPDR and severe NPDR	Detect dark and bright regions to recognize DR-related lesions. Classify these regions with a hybrid classifier comprising Gaussian mixture model and m-Medoids-based classifiers	AUC: 98.1
Wong Li Yun et al. [43]	Identified normal retina, moderate NPDR, severe NPDR and PDR	Four stages of diabetes are classified using neural network (NN) classifier	SN: 90, SP: 100
Mishra et al. [27]	Detect limited DR-related lesions	Blood vessel, optic disk and exudates detection followed by a KNN classifier	unknown
ManjulaSri et al. [25]	Detection of four DR levels such as normal, mild NPDR, moderate and severe NPDR and PDR	Pre-processing, vessel extraction, detection of perimeter and area of retinal blood vessel steps were developed	ACC: 85%
Verma et al. [40]	Detection of three DR levels such as normal, moderate NPDR and severe NPDR	Perimeter and area of retinal blood vessels, hemorrhages were detected and these features were classified with random forest (RF) algorithm	ACC. 87.5%
Ahmad Fadzil et al. [3]	This paper presented a system to obtain no DR but severe NPDR/PDR stages	The algorithm used measurement of perimeter and spread area of blood vessels along with pre-processing steps	ACC: 98%

AUC area under the receiver operating curve, *ACC* accuracy, *SN* sensitivity, *SP* specificity, diabetic-related lesions: *MAs* microaneurysms, *HRs* hemorrhages, *HE* hard exudates, *CWS* cotton wool spots, *PDR* proliferative diabetic retinopathy, *DME* diabetic macular edema, *NPDR* non-proliferative diabetic retinopathy

preprocessing, extraction of candidate lesions, feature set formulation and classification stages. They obtained an accuracy of 98.52%.

In [43], four groups were detected by Wong Li et al. through identifying normal retina, moderate NPDR, severe NPDR and PDR stages on 124 retinal photographs. They extracted six features on pre-processed images after a contrast enhancement process. These features were the perimeter and area of red, green and blue layers. Afterward, the features were fed into a three-layer feed-forward neural network (FF-NN) for classification. They showed a sensitivity of 90% and specificity of 100%. However, the authors presented a system to extract better features to improve the efficiency of the classification.

Similarly, in [27], the severity level of DR was identified by detecting a number of DR-related elements such as thickness of blood vessels, fovea region and blood clotting area. Based on these features, they determined normal, mild, moderate and severe level by using a KNN classifier. However, they did not present numerical results that could demonstrate the accuracy. In [25], the severity of DR was identified based on the perimeter and area of retinal blood vessels which were previously extracted from fundus images. The images were graded into four subsets, such as normal, mild NPDR, moderate and severe NPDR and PDR, yielding an accuracy of 85%. Nevertheless, they did not distinguish between moderate and severe PDR, and therefore, the algorithm was only based on retinal blood vessel characteristics. Moreover, in [40] retinal images were classified into normal, moderate and severe NPDR cases. The basis for the classification was the detection and quantification of blood vessels and hemorrhages. The classification decision was performed by the random forests technique based on the area and perimeter of the blood vessels and hemorrhages. The authors reported 90% accuracy for the normal case, while success rate of moderate and severe NPDR cases was 87.5%. Also in [3], a computerized DR grading system was developed for detection of four stages by using pre-processing, segmentation and feature extraction steps. The authors tested this algorithm on the FINDeRS database and achieved the sensitivity of 84%, specificity of 97% and accuracy of 95%.

From this view of the current development of automatic techniques for DR grading, we can highlight some facts. Some of the existing methods rely on the correct segmentation of certain DR-related structures within the images, something difficult to achieve and computationally expensive. Moreover, the errors of this step can be propagated to the rest of the system. Other methods use outdated machine learning algorithms and image processing techniques and do not provide objective numerical measures with a sufficient number of analyzed images. Regarding the grading of the diabetic retinopathy, the majority of these previous studies were focused on the differentiation between NPDR and PDR. Only a few studies were devoted to the classification of the images into different severity level/stages of DR [3, 4, 25, 27, 33, 40, 43]. In fact, a complete classification into the five stages was only addressed in [33]. Still, those systems required extensive domain expert knowledge to do pre- and post-processing steps on retinal fundus images. Therefore, the detection of severity level in retinal fundus images remains a challenging task that needs to be solved before automatic pathology grading systems. It can be rolled out for general clinical use.

2 Methodology

2.1 Data acquisition

To test and evaluate the performance of a proposed system, the 750 digital retinographs were collected from three online public and one private source. The overall information about this dataset is presented in Table 2. We have requested two experienced ophthalmologists to select the images from these datasets. As shown in this table, the first online dataset of 60 color retinographs is selected from standard diabetic retinopathy database (DIARETDB1). In the DIARETDB1 dataset, we have selected only 60 mild NPDR images among 89 color fundus images. These images were captured using the same 50° field-of-view digital fundus camera with varying imaging settings. The curvelet-based detection of foveal avascular zone (FAZ) dataset contains normal 12 of DR, 4 of moderate NPDR, 12 of severe NPDR and 12 of PDR color retinal images. We have collected 40 images from FAZ dataset in total. Next, the 396 retinographies were selected among 1200 color fundus images from MESSIDOR dataset. In MESSIDOR dataset, we have selected 88 of normal DR, 36 of mild NPDR, 96 of moderate NPDR, 88 of severe NPDR and 88 of PDR color images. The color images were captured in MESSIDOR by using a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a 45° field of view.

The PRV-RDB retinographies were collected from the Private Hospital Universitario Puerta del Mar (HUPM, Cádiz, Spain). This HUPM dataset consists of 250 retinal images that were acquired using a CCD camera of a Topcon mydriatic retinograph model Topcon TRC 50 EX. The HUPM dataset is non-mydriatic retinograph, model TRC NW 200 of Topcom. The private dataset is compounded by images selected from the Diabetic Retinopathy Unit of the Hospital during early detection of diabetic retinopathy program of the "Junta de Andalucía" (regional government). Each image was captured using 8 bits per color plane, and resolution level is (1960 × 1960) pixels. The dataset was divided by two ophthalmological experts into the five

Table 2 Selection of clinical datasets of diabetic retinopathy disease for recognition of five severity level

Datasets	Severity level and images	Download URL
[a]DIARETDB1	Mild NPDR: 60	http://www.it.lut.fi/project/imageret/diaretdb1/
[b]FAZ	Normal DR of 12, moderate NPDR of 4, severe NPDR of 12 and PDR of 12 images	http://www.biosigdata.com/?download=colour-fundus-images-of-healthy-persons-patients-with-diabetic-retinopathy
[c]MESSIDOR	Normal DR of 88, mild NPDR of 36, moderate NPDR of 96, severe NPDR of 88 and PDR of 88 images	http://www.adcis.net/en/DownloadThirdParty/Messidor.html
[d]Prv-DR	Normal DR of 50, mild NPDR of 50, moderate NPDR of 50, severe NPDR of 50 and PDR of 50 images	Private Hospital Universitario Puerta del Mar (HUPM, Cádiz, Spain)
Total images	Normal DR: 150, mild NPDR: 150, moderate NPDR: 150, severe NPDR: 150, PDR: 150 = 750 retinograph images	

[a] DIARETDB1: Kauppi T et al. DIARETDB1 diabetic retinopathy database and evaluation protocol, Technical report

[b] FAZ dataset: Hajeb SH (2012) Analysis of foveal avascular zone for grading of diabetic retinopathy severity 8 based on curvelet transform. Graefe's Archive for Clinical and Experimental Ophthalmology 250 (11) 1607-1614. Available on:

[c] MESSIDOR- Decenci`ere E et al. (2014) Feedback on a publicly distributed image database: the Messidor database. Image Analysis & Stereology 33 (2014) 231–234

[d] Prv-DR: Private dataset contains 250 retinograph images collected from Hospital Puerta del Mar (HUPM, Cádiz, Spain)

common severity levels described as none, mild NPDR, moderate NPDR, severe NPDR and PDR of diabetic retinopathy (DR). In each category, there are 50 images per class. An example of these images is represented in Fig. 2.

2.2 System overview

According to our limited knowledge, we did not find any study that focused on deep-learning algorithm for classification of five levels of severity. Therefore, the identification of five severity levels of DR is performed by the integration of effective visual color features and a deep-learning model [19] to solve the above-mentioned issues without detecting the complex structure of DR-related lesions and performing any pre- or post-processing steps on retinographies. The overall systematic flow diagram of the proposed five severity levels of diabetic retinopathy (SLDR) is shown in Fig. 3.

We used dense color scale-invariant feature transform (DColor-SIFT) [1] and gradient location-orientation histogram (GLOH) [26] to characterize the images. As retinal fundus images possess useful color information, we have considered color-based descriptors to increase the discriminative power of the features. Among the many descriptors proposed in the literature, we have selected DColor-SIFT due to their invariance with respect to light intensity and their widely accepted descriptive performance [39]. Color-SIFT extracts SIFT points on a normalized opponent color space, where intensity information has been eliminated. In order to extract these local features, we need to find a set of interest points (IPs) and compute the Color-SIFT descriptor associated with each of them. In this work, IPs are the gradient strengths locally normalized. In this way, we account for changes in illumination and contrast. Besides DColor-SIFT descriptors, we have also extracted GLOH features at every point of the image. To that purpose, a log-polar location grid is identified and SIFT descriptors are calculated on it. The obtained deep visual features (DVFs) are normalized and combined into a feature vector that characterizes the image.

A deep-learning-based classifier [16, 35, 37] is utilized to learn these deep visual features (DVFs). Moreover, the proposed method introduces a new semi-supervised hierarchical approach to best describe the five stages of diabetes. The performance of the proposed SDLR system has been evaluated using the area under the receiving operating curves (AUC) [13]. We have also compared the method with other state-of-the-art tools and tested that this SDLR system on a dataset consists of 750 (150 images per stage) retinal fundus images.

In summary, the method instead of detecting individual DR-related lesions as in [4, 25, 27, 28, 32, 33, 40, 42, 43], uses a general technique based on local features to classify fundus images into any of the five stages of DR. To this purpose, this SLDR system performs three major tasks such as detection of visual features, prediction of diabetic stage and fine-tuning of deep neural networks that are further explained in the following subsections.

2.3 Detection of visual features

To classify images, detection of features is a fundamental and relevant problem in computer vision and image processing tasks. In the literature, there were many proposed methods for visual feature extraction [6, 12]. However, it is difficult to extract optimal and effective visual features due to the requirement of domain expert knowledge. Moreover,

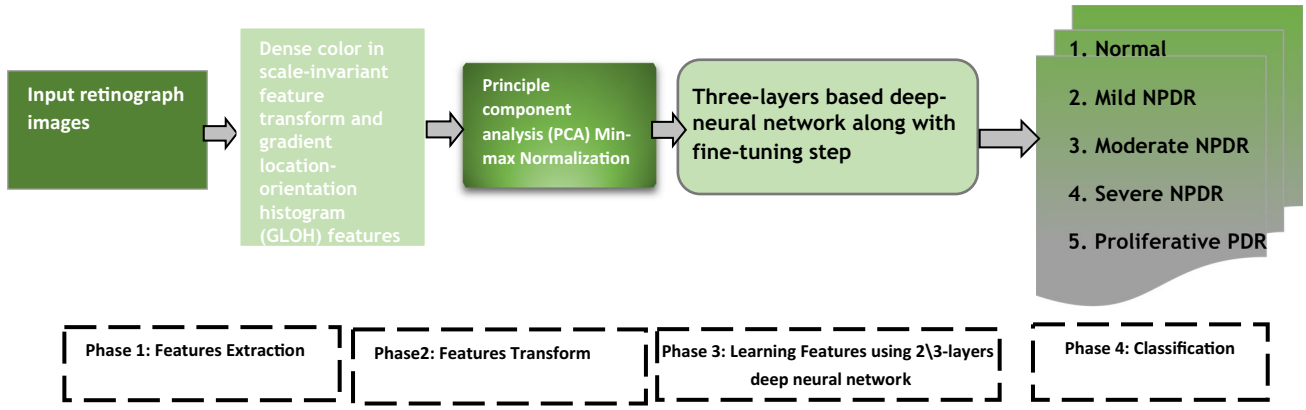

Fig. 3 Systematic diagram of the proposed severity-level recognition system to classify five stages of diabetic retinopathy

the selection of features is also a tedious job for experts. In this paper, we used color features to define properties of retinograph images.

To extract visual features, the DColor-SIFT [1] and gradient location-orientation histogram (GLOH) [26] methods were performed on each retinograph images. In this study, the DColor-SIFT [39] features are used to deliver a description of the color content of an image. In addition to this, it can also describe color variations while being invariant to illumination changes.

DColor-SIFT needs interest points (IPs) to be defined. Each of them is the location where SIFT will be computed. In this paper, we used for IPs selection a Laplacian pyramid, which is approximated by difference of Gaussian (DOG). For each of the input images, IPs are considered as the extrema in a 3D image representation (space coordinates and scale within the pyramid). For Gaussian filter definition, we have used a variance of $\sigma = 1.4$, a value that was experimentally determined.

Instead of computing these gradients in the gray-level version of the image, we consider its color content. Consider a retinography $R_C(x, y)$, which is defined in RGB-color format as $R_C(x, y) \rightarrow (R, G, B)$. The DColor-SIFT features are extracted from a fixed window of size 16×16 pixels around the keypoint g, with an inter spacing of 8 pixels. The color invariance model for Color-SIFT descriptor depends on the Kubelka–Munk theory. By assuming equal energy illumination, the reflection spectrum at the observation position can be modeled as:

$$E(\mu, g) = I(g)\left[1 - \rho_f(g)\right]^2 R(\mu, g) + I(\mu, g) \cdot \rho_f(g) \qquad (1)$$

where μ represents wavelengths, g is the 2D vector defined over the fixed grid, $I(\mu, g)$ is the spectral intensity and $\rho_f(g)$ denotes Fresnel reflection coefficient at the grid g, whereas $R(\mu, g)$ denotes the material reflectance. Let E_μ and $E_{\mu\mu}$ be the first and second derivative of (1) with respect to μ. Then, Color-SIFT $C_{\mathrm{INV}}(x, y)$ invariance features are defined as the ratio of E_μ and $E_{\mu\mu}$ as:

$$C_{\mathrm{INV}}(x, y) = \frac{E_\mu}{E_{\mu\mu}} = \frac{\partial R(\mu, g)}{\partial \mu} \Big/ \frac{\partial^2 R(\mu, g)}{\partial \mu^2} \qquad (2)$$

This descriptor as calculated by Eq. (2) is independent of viewpoint, surface orientation, illumination direction, intensity and Fresnel reflectance coefficient. A linear transformation from the RGB space can be used to obtain an approximation of E, E_μ and $E_{\mu\mu}$ based on the human vision system and CIE 1964 XYZ as follows:

$$\begin{bmatrix} \hat{E}(x, y) \\ \hat{E}_\mu(x, y) \\ \hat{E}_{\mu\mu}(x, y) \end{bmatrix} = - \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & 0.60 & 0.17 \end{pmatrix} \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \qquad (3)$$

In the present approach, we also compute GLOH [29] features to improve classification performance. To this purpose, we calculate a binned histogram of gradients (HOG) structure from local gradient magnitudes. However, instead of Cartesian rectangular bins, we have adopted a variation that uses log-polar bins at selected locations, IPs. Here IPs are the ones selected for DColor-SIFT computing, that is, maximal pixels in space and scale provided by DOG pyramid.

At each of the IPs, the SIFT descriptors are computed. The log-polar location grid [26] has 3 bins in the radial direction (the radius set to 6, 11 and 15) and 8 in angular direction resulting in 17 location bins. Note that the central bin is not divided into angular directions. The gradient orientations are quantized in 16 bins, giving a 272-bin histogram. The size of this descriptor is reduced with the principal component analysis technique (PCA) where the covariance matrix for PCA is estimated on 200 patches collected from various images. The 64 largest eigenvectors are retained for description.

$$h(x,y) = \arg\log - \text{polar} \sum_{k=0}^{n} R_g(x,y)^k \tag{4}$$

and

$$G\_F(x,y) = \arg\max_{\text{PCA}} \left[h(x,y)\right] \tag{5}$$

Afterward, the DColor-SIFT and GLOH features are combined into a vector and normalized using min–max scaling approach. In this approach, the data are scaled to a fixed range usually between ranges [0–1]. This normalized feature vector is then defined as:

$$f(x,y) = \arg_{\min-\max} \{C_{\text{INV}}(x,y), G\_F(x,y)\} \tag{6}$$

This visual feature vector, Eq. (6), constitutes the visual feature vector, which is used subsequently by deep-learning algorithm to form deep visual features.

2.4 Recognition of severity level

Five classes of the severity level of diabetic retinopathy (SLDR) are recognized by using visual features and multilayer deep-learning neural network (DLNN) architecture. In this paper, we have introduced some improvements to DLNN model in terms of features learning and classification tasks. In general, the DLNN algorithm has many layers that are hard to optimize. In addition to this optimization issue, the DLNN algorithms may generalize poorly [16]. The deep-learning algorithms are capable of modeling very complex and highly nonlinear relationships [35] between inputs and outputs through different layers.

The deep-learning idea is to learn one layer of feature detectors at a time with the states of the feature detectors in one layer acting as the data for training the next layer. Moreover, the deep-learning algorithms are used to transfer the input (raw data) pixels [35] into multiple layers for learning the features that are hard to rectify. However, learning direct features from image pixels required lots of learning time. Therefore, the color visual features are extracted and classified using DLNN model along with discriminative "fine-tuning" phase during which backpropagation through the DLNN slightly adjusts the weights found in pre-training [14].

In this paper, three active layers were developed using deep learning such as base layer (Base-L), compressed layer (Comp-L) and predicted layer (Pred-L) along with a fine-tuning step. The Base-L layer was developed by using restricted Boltzmann machines (RBMs) [6]. The RBMs framework was adopted due to its unsupervised power for features learning. In RBMs unsupervised learning model, the weights are often initialized randomly and then used a greedy layer-by-layer pre-training algorithm to initialize the network weights in an unsupervised fashion. Similarly, the RBMs model was used to construct Base-L layer to transform the features vector into activation of the hidden units.

The Comp-L layer is generated by using Shannon entropy constraints (SECs) [15] to obtain effective weights in an unsupervised manner from the outputs of the Base-L layer and, therefore, refining the feature weights from this previous layer. Afterward, the Pred-L layer is added in a supervised fashion to classify among five classes of diabetes. In addition to these three layers (Base-L, Comp-L and Pred-L), a fine-tuning step was also introduced by using the backpropagation neural network step. This weights fine-tuning step was applied in an iterated manner until we achieved the desired classification accuracy and was performed only on Pred-L layer to refine the classification decision. In practice, this fine-tuning step uses gradients from SM-L layer for classification of errors and then it is back-fed into the Comp-L layer of the networks.

The base layer (Base-L) is trained as a bipartite undirected graphical model called restricted Boltzmann machines (RBMs) [6] that model visual features in an unsupervised way. We use an unsupervised greedy layer-by-layer pre-training algorithm to initialize the network weights. The RBMs have been used effectively in the past for several practical applications including handwriting recognition, 3D object recognition, dimensionality reduction and speech recognition, showing significant advantages over convolution neural networks (CNNs). Therefore, to construct Base-L layer, the RMBs model is applied in a two-layer architecture as shown in Fig. 4.

The RMBs have a two-layer architecture [6], in which the visible stochastic units v (typically Bernoulli or Gaussian) are connected to the hidden stochastic units h (typically Bernoulli). Consider the first layer that is trained as an RBM P1 with hidden layer h1 and visible layer v1. Normally, all visible units are connected to all hidden units and there is no visible-to-visible or hidden-to-hidden unit connection. The weights of the connections and the biases of the individual units form a joint probability distribution $P(v, h|\theta)$ over the visible units v and hidden units h given the model parameters θ.

This distribution is computed based on an energy function $E(v, h|\theta)$:

$$P(v,h1,h2;\theta) = P_{\text{RBM}}(h1,h2;w2) \tag{7}$$

where

$$P_{\text{RBM}}(h1,h2;w2) = 1/z(w2)\exp\left(h1^t w2 h2\right) \tag{8}$$

Energy function is defined as:

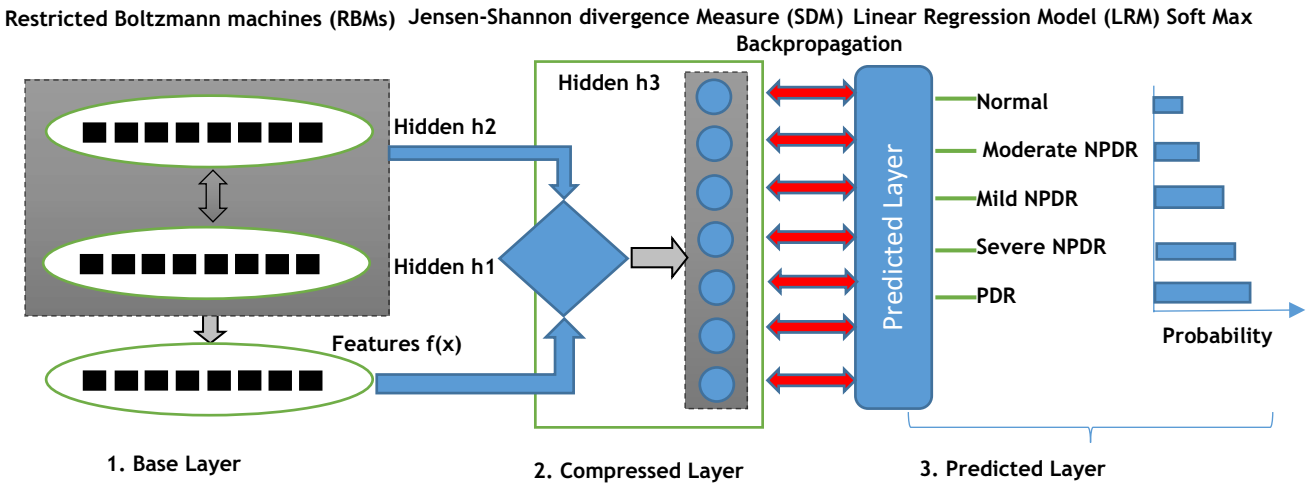

Fig. 4 A proposed deep-learning neural network (DLNN) model for classification of five severity level of diabetic stage with different layers

$$P(v,h|\theta) = \exp\left(-E(v,h|\theta)\right)/Z(\theta) \tag{9}$$

where $Z(\theta)$ is the normalized constant calculated as:

$$Z(\theta) = \sum_{v}\sum_{h}\exp(-E(v,h|\theta)) \tag{10}$$

θ consists on the model parameters $\theta = \{w, b, a\}$ where w_{ij} is the weight between visible unit i and hidden unit j; b_i and a_j are biases for visible unit i and hidden unit j, respectively. V, H are the numbers of visible units and hidden units, respectively. Since there is no visible–visible connection, all of the visible units become independent given the hidden units, and vice versa. The conditional distributions can be effectively derived [10] as

$$P\left(h_j = 1|v,\theta\right) = \text{sigm}\left(\sum_{i=1}^{V} W_{ij}v_i + a_j\right) \tag{11}$$

and

$$P(v_i = 1|h,\theta) = \text{sigm}\left(\sum_{j=1}^{H} W_{ij}H_i + b_i\right) \tag{12}$$

The compressed layer (Comp-L) is generated by reducing the weighted parameters that are obtained from Base-L ($h2$ hidden layer) layer by using the Jensen–Shannon divergence measure (SDM) [24]. The compressed parameters are trained from visual features and the output produced by Base-L in an unsupervised way. The main purpose of the Comp-L layer is to extract the features that having discriminate power based on the divergence measure from the hidden unit ($H2$) produced by Base-L layer. As a result, it can be used to measure the probability difference between the weights generated from H2-layer and the input set of visual features.

$$P = 1/n\sum_{i=1}^{V} W_{ij}v_i + a_j + \sum_{j=1}^{H} W_{ij}H_i + b_i \tag{13}$$

where w_{ij} is the weight between visible unit i and hidden unit j; b_i and a_j are biases for visible unit i and hidden unit j, respectively. V, H are the numbers of visible units and hidden units, respectively. Since there is no visible-to-visible connection, all of the visible units become independent given the hidden units, and vice versa. For any probability distribution $P = \left(p_1, p_2, p_{3,\ldots,}p_n\right) \in \Omega_n$ Shannon introduced the entropy to measure the uncertainty associated with probability distribution P as follows:

$$H(P) = -\sum_{j=1}^{n} p_j \log p_j. \tag{14}$$

And the $Q = (q_1, q_2, q_3, \ldots, q_n)$ is defined as the input visual features and the Shannon entropy is defined for Q as:

$$H(Q) = -\sum_{j=1}^{n} f_j \log f_j \tag{15}$$

Therefore, Jensen–Shannon divergence (JSD) between the two probability distributions, i.e., one for $P = (p_1, p_2, p_3, \ldots, p_n)$ and another for $Q = (q_1, q_2, q_3, \ldots, q_n)$, is defined as:

$$\text{JSD}(P; Q) = H\left(\frac{P+Q}{2}\right) - \left(\frac{H(P)+H(Q)}{2}\right) \tag{16}$$

Finally, the compressed layer (Comp-L) is used to update the weights that are obtained from Base-L, which is derived as:

$$W(t) = W(t) + a(W(t) - \text{JSD}(P; Q)) \tag{17}$$

The weighted features generated by Comp-L layer were used as an input to the predicted layer (Pred_L). The main goal is for the generated data (Pred_L) to be as close as possible to the real-world problems, and this is reflected in the weight update formula.

The predicted layer (Pred_L) is generated by using a linear softmax classifier [14] in a supervised manner according to five classes training labels (Y) that are already determined from the dataset. In fact, the Pred-L layer allows handling the maximum diversity among the class categories. In the Pred_L layer, the softmax classifier is used alone with the backpropagation fine-tuning step to predict the corresponding class. In the past studies, the softmax regression model is also used in various probabilistic multiclass classification problems and defined in general as:

$$P(Y = i|x) = e^{x^T} w_j \Big/ \sum_{k=1}^{K} e^{x^T} w_k \tag{18}$$

where K parameter represents the distinct linear functions with the constant value 5, and the probability for the j class is predicted by given a sample vector x. As a result, the softmax function (where $X^T W$ denotes the inner product of X and W) is used to predict five classes of diabetic severity level and defined by a weighted matrix w and a base vector b as:

$$P(Y = i|x, w, b) = \arg_{\text{soft}\max_i} (wx + b_i) \tag{19}$$

Hence, the layer predicted model is defined as:

$$Y_{pred} = \arg\max_i P(Y = i|x, w, b) \tag{20}$$

This Pred_L layer determined the five classes of diabetes. However, there are classification errors in the Pred-L that minimized using backpropagation in all previous two layers (Comp_L and Base_L). The Pred_L layer is defined by using cross-entropy function due to its easy implementation capability that is associated with the softmax classifier. This cross-entropy loss function can be written as:

$$\arg\{\text{lost}_i\} = \frac{1}{2} \times \log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) \tag{21}$$

The expression in Eq. (21) is low when the correct class probability is high, and the opposite happens when the class probability is low. To get better performance of the deep network, a fine-tuning step with backpropagation is used to change the parameters of the supervised layer according to the training style. The fine-tuning step is then performed as the last step using the gradients from softmax layer during classification errors and back-feeding them into the encoding layers of the networks to enhance the recognition results. The overall layers of DLNN architecture are visually presented in Fig. 4.

3 Experimental results

A statistical analysis was performed to evaluate the suitability of the proposed SLDR system on 750 retinograph fundus images (150 per category). We used the area under (AUC) [14] the receiver operating characteristic (ROC) analysis on the test dataset along with three measures such as sensitivity (SE), specificity (SP) and training errors (E). In the past studies, the AUC was a commonly used index to evaluate the overall discrimination approach of the classifiers. The higher the value of AUC means the greater classification accuracy in which the AUC ranges vary from 0.50 to 1.0. In general, the receiver operating characteristic (ROC) curve is basically defined as a graphical plot to illustrate the performance via a fixed threshold value of 0.5.

This SLDR system for classification of five stages of diabetic retinopathy was implemented in MATLAB R2015a (8.5.0 ® Mathworks, Natick, MA) on a Core i7 64-bit Intel processor system with 8 GB DDR3 RAM, running operating system (Windows 8). The algorithm took on average of 6.46 s per image to extract all of the features based on DColor-SIFT and GLOH techniques. The training of these features was performed by DLNN that took 3.45 s on average, while the final out layer took 1.78 s on average to be created. Therefore, a total time of 12.30 s was consumed for each diabetic class, on a fixed number of 300 iterations, for training the deep-learning neural network (DLNN) classifier. However, when the training has been done and the test is performed, an average time of 7 s is only needed to classify the image. This computation time spent on feature extraction that could be further reduced by using an optimized C/C++ implementation.

To perform a comparison, the SE and SP statistical measures are calculated by deriving the true-positive rate (TPR) against the false-positive rate (FPR), respectively. The TPR metric is known as a sensitivity measure, while FPR is known as (1 − specificity). If the classifier is predicted positive (P) case and the actual (T) is also the same, then this is known as TP. However, if the actual class label is negative (N), then it is known as false positive (FP). On the other hand, a true negative (TN) has followed when both the estimated class label and the actual value are n, and false negative (FN) is when the prediction outcome is n, while the actual value is p. The sensitivity (TPR) and specificity (1-FPR) metrics are calculated by using Eqs. (22)–(24). To show the significance of proposed system, the ROC curve is plotting the sensitivity in the y-axis versus the specificity of the false alarm in the x-axis.

Table 3 Performance of proposed system for classification of five severity level of diabetic retinopathy system (SLDR) on 750 retinograph images

Classes	Severity level	SE	SP	*E*	AUC
1.	Normal	94.10%	97.60%	0.453	0.901
2.	MNPDR	92.23%	96.99%	0.467	0.934
3.	NPDRM	88.45%	93.25%	0.553	0.926
4.	SNPDR	88.43%	92.16%	0.687	0.882
5.	PDR	89.30%	92.21%	0.514	0.880
Total result		92.18	94.50	0.545	0.924

E average standard deviation training error during learning, *SE* sensitivity, *SP* specificity, *AUC* area under the ROC (receiver operating characteristic) curve, *Normal* no diabetes, *MNPDR* mild proliferative diabetic retinopathy, *NPDRM* moderate proliferative diabetic retinopathy, *SNPDR* severe proliferative diabetic retinopathy, *PDR* proliferative diabetic retinopathy

Table 4 Performance comparisons with state-of-the-art three severity-level recognition system on 450 retinograph images

No.	Methods	SE (%)	SP (%)	*E*	AUC
1.	Verma_DR [40]				
	Normal	67.70	72.60	0.653	0.710
	MNPDR	71.30	75.43	0.635	0.722
	SNDPR	78.50	83.34	0.621	0.731
2.	Akram_DR [4]				
	Normal	70.23	75.20	0.521	0.731
	MNPDR	75.21	79.40	0.523	0.754
	SNDPR	80.54	85.14	0.501	0.763
3.	Proposed SLDR				
	Normal	94.70	97.60	0.453	0.901
	MNPDR	95.30	67.43	0.435	0.943
	SNDPR	78.50	86.34	0.321	0.875

E average standard deviation training error during learning, *SE* sensitivity, *SP* specificity, *AUC* area under the ROC (receiver operating characteristic) curve, *Normal* no diabetes, *MNPDR* mild proliferative diabetic retinopathy, *NPDRM* moderate proliferative diabetic retinopathy, *SNPDR* severe proliferative diabetic retinopathy, *PDR* proliferative diabetic retinopathy

$$\text{Senstivity (SE)} = \text{True - Positive Rate (TPR)} = \text{TP}/(\text{TP} + \text{FN}) \tag{22}$$

$$\text{Specificity (SP)} = 1 - (\text{FP}/(\text{FP} + \text{TN})) \tag{23}$$

$$\text{False - Positive Rate (FPR)} = \text{FP}/(\text{FP} + \text{TN}) \tag{24}$$

The performance of SLDR system has been statistically measured in terms of receiving operating characteristic curve (ROC), SP and SE measures. Table 3 illustrates the best-measured performance, i.e., AUC: 0.924. The proposed SLDR system significantly improves in terms of sensitivity (SE) of 92.18% and specificity (SP) of 94.50%. The best performance is measured for severity level of diabetes in case of normal of (SE: 94.10%, SP: 97.60% and AUC: 0.901) and mild NPDR (SE: 92.23%, SP: 96.99% and AUC: 0.934) compared to moderate NDPDR (SE: 88.45%, SP: 93.25% and AUC: 0.926), severe NPDR of (SE: 88.43%, SP: 92.16% and AUC: 0.882) and PDR of (SE: 89.30%, SP: 92.21% and AUC: 0.880). It was also noticed from this table that the training errors (E) are 0.545 on average for every class.

This achieved result of SLDR system was compared with other two state-of-the-art severity level recognition systems such as Akram_DR [4] and Verma_DR [40]. However, these systems have performed experiments that are related to only three classes such as normal, moderate NPDR and severe NPDR. In Verma_DR [40] study, those three stages have been driven using limited DR-related lesions such as blood vessels and hemorrhage from each of the retinographies followed by a random forests (RF) classifier [40]. In Akram_DR [4] study, another set of DR-related lesions such as microaneurysms, hemorrhages and exudates were segmented and classified using an ensemble classifier [4]. According to Akram_DR and Verma_DR systems, we have utilized the same computerized techniques to compare it with our proposed SLDR system. The results of these comparisons with other state-of-the-art systems are presented in Table 4 on 450 retinograph images (150 per category). In Verma_DR system, the area under the receiving operating curve (AUC) is 0.710 of normal, MNPDR of 0.722 and SNDPR of 0.731. Similarly, the sensitivity (SE) values of normal of 67.70%, MNPDR of 71.30% and SNDPR of 78.50% were obtained. Also, the specificity (SP) values were normal of 72.60%, MNPDR of 75.43% and SNDPR of 83.34% calculated. In Akram_DR system, the area under the receiving operating curve (AUC) is 0.731 of normal, MNPDR of 0.731 and SNDPR of 0.763. Moreover, the sensitivity (SE) values of normal of 70.23%, MNPDR of 75.21% and SNDPR of 80.54% were obtained. The specificity (SP) values were also calculated such as normal of 75.20%, MNPDR of 79.40% and SNDPR of 85.14%. In contrast with these two approaches, our proposed SLDR system obtained significant higher results such as AUC is 0.901 of normal, MNPDR of 0.943 and SNDPR of 0.875. Similarly, the sensitivity (SE) values of normal of 94.70%, MNPDR of 95.30% and SNDPR of 78.50% were obtained. Also, the specificity (SP) values were normal of 97.60%, MNPDR of 67.43% and SNDPR of 86.34% calculated.

The results are also displayed in terms of training errors in Table 5. A low learning rate (E) is observed for the proposed DLNN algorithm in case of normal (5%), MNPDR (7%) and SNDPR (6%). These values are lower to the ones obtained with Akram_DR: normal (11%), MNPDR (18%)

Table 5 Performance comparisons with state-of-the-art three severity-level recognition system on 450 retinograph images using ten fold cross-validation test

Methods	State-of-the-art algorithms	Normal	MNPDR	SNDPR
1.	Verma_DR [40]	13	19	20
2.	Akram_DR [4]	11	18	17
3.	Proposed SLDR	5	7	6

Normal no diabetes, *MNPDR* mild proliferative diabetic retinopathy, *SNPDR* severe proliferative diabetic retinopathy

and SNDPR (17%) and Verma_DR system such as normal (13%), MNPDR (19%) and SNDPR (20%). In Fig. 5, we can visually understand the significant improvement of the proposed SLDR system when compared to Akram_DR [4] and Verma_DR [40]. In Table 4, we have achieved a significant improvement compared to the other two systems because of the selection of all DR-related features from each retinal funds image compared to the limited number of features used in [4] and [40]. Also, the segmentation steps are another cause of lower accuracy of the two comparison methods.

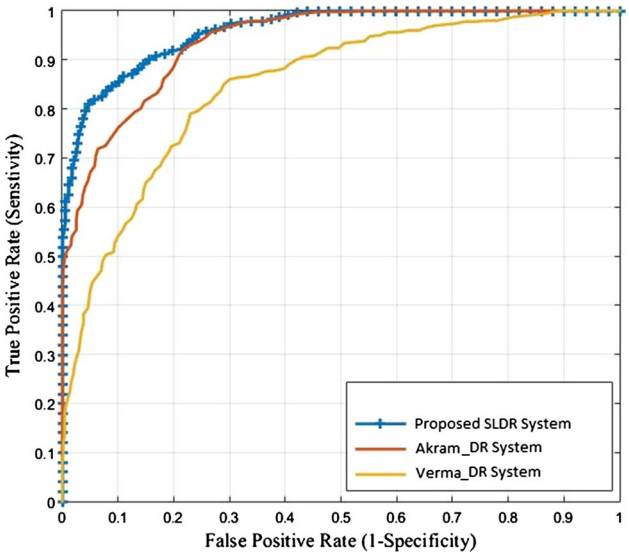

Fig. 5 Performance comparisons of proposed SLDR system with state-of-the-art classification systems in terms of area under the receiver operating curve

4 Discussion

Eye fundus photography is the diagnostic tool for diabetic retinopathy (DR). This examination should be performed periodically throughout the life of the patient. Most guidelines recommend annual screening for those with none or mild retinopathy and repeat fundus examination for moderate DR every six months. This strategy and the increase in the incidence of DR worldwide represent an important workload impossible to be assumed by ophthalmology specialists. In the USA, it is noticed that the half of the population with diabetes was screened annually for diabetic retinopathy. Moreover, in France, the 30% of the diabetic patients have undergone through DR screening each year.

The emergence of new technologies such as non-mydriatic retinography and online data transmission has enabled the implementation of telemedicine programs for the early diagnosis of DR in different countries. The interpretation of these images is evaluated by specially trained family doctors, endocrinologists or non-medical personnel with results of sensitivity (SE) and specificity (SP) comparable or superior to the conventional examination of the ophthalmologist. The overall diagnosis accuracy of telemedicine DR screening is high with sensitivity exceeding 80% and specificity of 90%.

The incidence of diabetes blindness has been also changed as reported between 2000 and 2010 year in England and Wales. The growing demand for care has led to the need to replace these physicians with automated computer-aided diagnosis (CAD) systems. For these CAD systems to replace human evaluators, they must obtain results with high SE and SP values. Diabetes in UK recommendations suggested that the SE of a screening program should be at least 80%. An estimation of a 61.2% reduction in the workload by introducing an automatic analysis system capable of differentiating between normal and diabetic retinopathy (DR) is noticed. Therefore, the automatic systems have a diagnostic capacity similar to that of human evaluators, but the capacity for stage classification and detection of progression is potentially superior. In addition, their analysis is consistent with time without presenting weaknesses due to transient criterion changes, fatigue or an effort to interpret poor-quality retinographies. The median time required to manually staging DR in retinal images with the seven fields on non-mydriatic fundus photography protocol was 12.8 min per patient. In contrast, the ultra-wide fundus image protocol had a median evaluation time of 9.2 min per patient, which represents a 28% reduction of image evaluation time (P, 0.0001).

Accordingly, the CAD systems were developed to provide second opinion to clinical experts for diagnosis of diabetes. However, the development of this CAD system for diagnosis of DR is very complicated due to different DR-related lesions. It is noticed that the state-of-the-art CAD systems did not detect all five stages of diabetes. The detection of five stages of diabetes is required for prediction of type of diabetes. Therefore, this paper focused on extracting color visual features by using DColor-SIFT and GLOH methods and classified them through a deep-learning neural

network (DLNN) architecture that integrates a compressed layer and fine-tuning steps to recognize among five severity level of diabetes (SLDR). The proposed SLDR system effectively differentiates among five stages of diabetes obtaining higher classification results when compared with other commonly used techniques. However, the small diversity of patterns produced for the same class over 10 different runs indicates the diversity of features learned for that class by using proposed DLNN approach.

The proposed SLDR system has been tested on 750 retinal images, and the results demonstrate the viability of an automatic grading system of severity level recognition based on it. On average, the sensitivity and specificity measures indicate that many retinal images are correctly graded into one of the five classes. However, there are some retinal images that are not correctly classified by the proposed SLDR system. To highlight this issue, we have visually displayed three retinal images of PDR, severe NDPR and moderate NPDR in Fig. 6. It can be observed that these three images have particular patterns that are difficult to characterize. To overcome this issue, we will further research in using other color spaces or color appearance models that better represent the color content of the images. As future work, we will also implement these visual features extraction and DLNN algorithms in C++ to increase the speed of SLDR system. Moreover, the speed of the system can be increased by implementing it through GPU-based parallel technique. We will also provide the five severity level of diabetes dataset online as a resource so that the other authors can evaluate their classification results. Also, we will try to implement SLDR system using perceptual-oriented color space to classify the diabetes retinopathy instead of using non-uniform color space such as RGB.

Regarding other comparisons with state-of-the-art methods for severity assessment into the five commonly used levels, as we have shown in the related work section, only a few efforts have been devoted to the recognition task [4, 27, 33, 40] of a limited number of diabetes stages. The majority of the methods rely on the precise segmentation of some elements of the images and were tested on private image databases. In the last few decades, the other studies focused more on pre- and post-processing techniques to classify limited diabetic severity level from color fundus images. The pre-processing steps rely on segmentation or detection of DR-related lesions. As a post-processing step, the features are extracted, selected and classified using outdated machine learning methods. In practice, these techniques cannot apply to detect real-time diabetic severity level. We have observed that there is no study that detected visual features and utilized deep-learning architecture for classification of the severity level for diabetes. In the past studies, the one form of the deep-learning model is a convolutional neural network (CNN) for classification of different biomedical image analysis tasks without extracting features. In practice, the CNN model is hard to train on image pixels and therefore in this study, we have extracted visual features to provide assistance to deep-learning architecture.

Furthermore, the issue of image databases availability, only [39] database provides 4 levels of DR severity and not five as it would be desirable. Therefore, we were forced to utilize a new private dataset for training and testing the proposed automatic SLDR system. In future works, we aim at building a free access fundus image database with DR grading provided by human experts according to the clinical classification. This dataset could then be used to train and validate an automatic grading system such as the one proposed in this study.

Our automated system for diabetic retinopathy detection and classification offers several advantages with respect to human evaluation including consistency of interpretation, high sensitivity and specificity and very fast results analysis. When training done, the SLDR system took an anverage time of 2.12 seconds to classify the severity level of diabetes.

Under these conditions, we can eliminate the need for the examination of normal images by trained professionals (SE: 92.180%, SP: 94.50% and AUC: 0.924). Also, we can repeat the examination of at-risk patients more often by

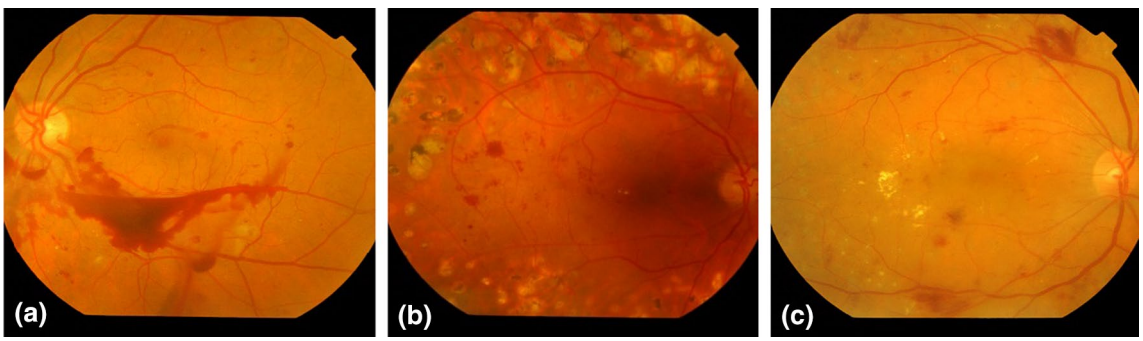

Fig. 6 An example of misclassified diabetic severity levels where **a** RDP, **b** severe NDPR and **c** moderate NPDR

lessening the possibility of misinterpretation. Although the diagnosis and classification have been performed using a single image of the retina, several studies have shown that the sensitivity of the single photograph camera in detecting moderate to severe non-proliferative and proliferative diabetic retinopathy is greater than 90%.

Two limitations that we intend to overcome in future studies are the evaluation of diabetic maculopathy and the use of more than one image for classification. Diabetic maculopathy, together with vitreous hemorrhage due to neovascularization, is one of the main causes of visual loss in the diabetic patient and one of the criteria for referring the patient to an ophthalmologist's visit. The international classification of diabetic retinopathy uses the criteria of the four quadrants for the diagnosis of severe retinopathy, and we can only detect the four quadrants with at least two fundus images.

5 Conclusions

In this paper, we presented a novel system known as SLDR to recognize five severity level stages of diabetic retinopathy by using visual features and deep-learning neural network (DLNN) model. The primarily aim of this paper is to focus on the recognition of five severity levels without performing pre- and post-processing steps on retinal fundus images. To the best of our knowledge, although there are a few studies devoted toward the recognition of severity level of DR, none of them focused on the five classes commonly used in practice by the experts. Moreover, these literature techniques were dedicated to the detection of limited DR-related lesions trying to classify the images on the base of the so-obtained regions' parameters. However, it is very difficult and time-consuming to correctly extract all of the DR-related lesions with image processing techniques due to the diversity of possible appearances. In contrast with state-of-the-art approaches, we have mainly focused on the use of a powerful classifier along with visual features extracted at several points of the images. Therefore, we do not need to identify characteristic elements, such as blood vessels or microaneurysms, on the images saving time and reducing error propagation. We have look for adaptability and accuracy in this work with an implementation based on visual features (DC-SIFT and GLOH) and a trilevel optimization technique on the deep-learning neural network method.

The proposed SLDR system has been evaluated using the area under the receiving operating curves (AUC) and tested on 750 retinograph images (150 per category) that were acquired from different sources. The overall performance of the proposed algorithm has been measured in terms of effectiveness and time. The sensitivity (SE) of 92.18%, specificity (SP) of 94.50% and area under the receiver operating curve (AUC) of 0.924 have obtained on average for all severity levels of DR, which constitutes a better value than other methods. SLDR system has achieved significantly higher classification rate with less computational time than other state-of-the-art methods. An AUC of 90.4% is obtained on average for all severity level of DR demonstrating that this system can be useful for its integration of a CAD tool for diabetes grading and prediction of type.

Acknowledgements This research was supported by the Research and Development Program of Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia (Grant No. 2006AA02Z347). We would also like to thank two expert ophthalmologists for providing us new private dataset.

Compliance with ethical standards

Conflict of interest The authors disclosed no potential conflicts of interest.

References

1. Abdel-Hakim AE and Farag AA (2006) CSIFT: A SIFT descriptor with color invariant characteristics. IEEE computer society conference on computer vision and pattern recognition 1978–1983
2. Acharya UR et al (2016) Novel risk index for the identification of age-related macular degeneration using radon transform and DWT features. Comput Biol Med 73:131–140
3. Ahmad Fadzil MH et al (2011) Analysis of retinal fundus images for grading of diabetic retinopathy severity. Med Biol Eng Comput 49(6):693–700
4. Akram MU et al (2014) Detection and classification of retinal lesions for grading of diabetic retinopathy. Comput Biol Med 45:161–171
5. Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. Computer vision. Lect Notes Comput Sci 3951:404–417
6. Bertolini D et al (2013) Texture-based descriptors for writer identification and verification. Expert Syst Appl 40(6):2069–2080
7. Bhaskaranand M et al (2016) Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis. J Diabet Sci Technol 10(2):254–261
8. Datta NS, Dutta HS, Majumder K (2016) Brightness-preserving fuzzy contrast enhancement scheme for the detection and classification of diabetic retinopathy disease. J Med Imaging 3(1):1–10
9. Early Treatment Diabetic Retinopathy Study Research Group (1991) Grading diabetic retinopathy from stereoscopic color fundus photographs-an extension of the modified Airlie House classification. ETDRS report number 10. Ophtalmology 98:776–806
10. Faust O et al (2012) Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. J Med Syst 36(1):145–157
11. Ganesan K et al (2014) Computer-aided diabetic retinopathy detection using trace transforms on digital fundus images. Med Biol Eng Comput 52(8):663–672
12. Guo Y, Zhao G, Pietikinen M (2012) Discriminative features for texture description. Pattern Recogn 45(10):3834–3843

13. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36
14. Hinton GE (2002) Training products of experts by minimizing contrastive divergence. Neural Comput 14:1771–1800
15. Hinton GE (2010) A practical guide to training restricted boltzmann machines. Tech. Rep. UTML TR 2010-003, Department of Computer Science, University of Toronto. https://www.cs.toronto.edu/~hinton/absps/guideTR.pdf. Accessed 4/1/2015
16. Hinton GE et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 29(6):82–97
17. Ibrahim S et al (2015) Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier. Med Biol Eng Comput 53(12):1345–1360
18. Kandemir M, Hamprecht FA (2015) Computer-aided diagnosis from weak supervision: a benchmarking study. Comput Med Imaging Graph 42:44–50
19. Keshavan MS (2017) Sudarshan M (2017) Deep dreaming, aberrant salience and psychosis: connecting the dots by artificial neural networks. Schizophr Res S0920–9964(17):30029–33034
20. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene Categories. IEEE computer society conference on computer vision and pattern recognition, pp. 2169–278
21. Lee J, Zee BC, Li Q (2013) Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy. PLoS ONE 8(12):e75699
22. Li B, Li HK (2013) Automated analysis of diabetic retinopathy images. Curr Diab Rep 13(4):453–459
23. Li Y et al (2015) A survey of recent advances in visual feature detection. Neurocomputing 149:736–751
24. Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory Arch 37(1):145–151
25. ManjulaSri R, Raghupathy RM, Rao KMM (2014) Image processing for identifying different stages of diabetic retinopathy. Int J Recent Trends Eng Technol 11:83–92
26. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 10(27):1615–1630
27. Mishra PK et al (2014) A computational modeling for the detection of diabetic retinopathy severity. Bioinformation 10(9):556–561
28. Mookiah MRK et al (2013) Computer aided diagnosis of diabetic retinopathy using multi-resolution analysis and feature ranking frame work. J Med Imaging Health Inform 3(4):598–606
29. Mookiah MRK et al (2013) Computer-aided diagnosis of diabetic retinopathy: a review. Comput Biol Med 43(12):2136–2155
30. Mookiah MR et al (2014) Decision support system for age-related macular degeneration using discrete wavelet transform. Biol Eng Comput 52(9):781–796
31. Nayak J et al (2008) Automated identification of diabetic retinopathy stages using digital fundus images. J Med Syst 32(2):107–115
32. Pires R et al (2014) Advancing bag-of-visual-words representations for lesion classification in retinal images. PLoS ONE 9(6):e96814
33. Prakash NB, Selvathi D, Hemalakshmi GR (2014) Development of algorithm for dual stage classification to estimate severity level of diabetic retinopathy in retinal images using soft computing techniques. Int J Elect Eng Inform 6(4):717–739
34. Rodriguez-Poncelas A et al (2015) Prevalence of diabetic retinopathy in individuals with type 2 diabetes who had recorded diabetic retinopathy from retinal photographs in Catalonia (Spain). Br J Ophthalmol 99:1628–1633
35. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
36. Teng T, Lefley M, Claremont D (2002) Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. Med Biol Eng Comput 40(1):2–13
37. Thomas S et al (2013) Deep neural network features and semi-supervised training for low resource speech recognition. In: Proceeding of IEEE international conference on acoustics, speech and signal processing, Vancouver, BC, pp. 6704–6708
38. Ting DS, Cheung GC, Wong TY (2016) Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol 44(4):260–277
39. Van de Sande KE, Gevers T, Snoek CG (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32(9):1582–1596
40. Verma K, Deep P, Ramakrishnan AG (2011) Detection and classification of diabetic retinopathy using retinal images. In: Proceeding of 2011 annual IEEE India conference (INDICON), pp 1–6
41. Washington RE et al (2014) All-cause mortality in a population-based type 1 diabetes cohort in the U.S. Virgin Islands. Diabetes Res Clin Pract 103(3):504–509
42. Welikala RA et al (2015) Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy. Comput Med Imaging Graph 43:64–77
43. Wong LY et al (2008) Identification of different stages of diabetic retinopathy using retinal optical images. Inf Sci 178(1):106–121

Dr. Qaisar Abbas received Doctor of Engineering Degree (Ph.D.) from the huazhong university of science and technology at Wuhan, China, in 2011. He is currently working as an Assistant Professor in the College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432 Saudi Arabia. He has published more than 30 research papers in both reputable journals and conferences. His research interests include image processing, medical image analysis and computer vision.

Dr. Irene Fondon received the B.E., M.E. and Ph.D. degrees in telecommunication engineering at the University of Seville, Seville, Spain, where she is currently working as an assistant professor in the Department of Signal Theory and Communication. Her research interests include image processing, color theory and texture perception usually applied to the development of computer-aided diagnosis tools.

Dr. Auxiliadora Sarmiento received the Telecommunication Engineering degree in 2002 and the Ph.D. degree in 2011, from the University of Seville, Spain, where she is currently an Assistant Professor. Her research interests focus on the area of component analysis and blind signal separation, especially in audio signals and biomedical images.

Dr. Soledad Jiménez received her Ph.D. degree from University of Cádiz, Cádiz, Spain, where she is currently Senior Lecturer in the Department of Surgery. She is also in the Glaucoma Unit from the Ophthalmological Management Unit in the University Hospital Puerta del Mar. Her research interests include retinal diseases such as diabetic retinopathy and glaucoma.

Dr. Pedro Alemany received his Ph.D. degree from University of Cádiz, Cádiz, Spain. Since then, he has been Senior Lecturer in the Department of Surgery. She is also in the Glaucoma Unit from the Ophthalmological Management Unit in the University Hospital Puerta del Mar. Her research interests include retinal diseases such as diabetic retinopathy and glaucoma.