Detection of severity level of diabetic retinopathy using Bag of features model

Mona Leeza[1] ✉, Humera Farooq[1]

[1]*Department of Computer Science, Bahria University Karachi Campus, 13 National Stadium Road, Karachi, Pakistan*
✉ *E-mail: monaleeza.bukc@bahria.edu.pk*

Abstract: Diabetic retinopathy is a vascular disease caused by uncontrolled diabetes. Its early detection can save diabetic patients from blindness. However, the detection of its severity level is a challenge for ophthalmologists since last few decades. Several efforts have been made for the identification of its limited stages by using pre- and post-processing methods, which require extensive domain knowledge. This study proposes an improved automated system for severity detection of diabetic retinopathy which is a dictionary-based approach and does not include pre- and post-processing steps. This approach integrates pathological explicit image representation into a learning outline. To create the dictionary of visual features, points of interest are detected to compute the descriptive features from retinal images through speed up robust features algorithm and histogram of oriented gradients. These features are clustered to generate a dictionary, then coding and pooling are applied for compact representation of features. Radial basis kernel support vector machine and neural network are used to classify the images into five classes namely normal, mild, moderate, severe non-proliferative diabetic retinopathy, and proliferative diabetic retinopathy. The proposed system exhibits improved results of 95.92% sensitivity and 98.90% specificity in relation to the reported state of the art methods.

1 Introduction

Diabetes mellitus (DM) is regarded as a challenging disease for public health worldwide. According to epidemiological studies aging, longer duration of DM and cardiovascular complications may lead to diabetic retinopathy (DR) [1, 2]. DR could be a major reason for blindness in the working-age population [3] having no clear clue at its early stage until it progresses and affects vision badly. Its progression may cause retinal damage and loss of vision or blindness. Patients diagnosed with type-I and type-II diabetes have a chance to suffer DR. In the first five years of diagnosis, type-I patients have almost no chance of DR but one in every five patients with newly diagnosed type-II diabetes have DR [4]. Chance of DR increases with time, almost all type-I diabetic patients have DR after 15 years of diagnosis of diabetes while this ratio is one-third for type-II diabetes in the same period of time [5].

Currently, some ophthalmologists use computer-aided diagnosis (CAD) systems to diagnose DR and its severity level but this detection of severity level is based on the number of DR-lesions present in the retina. DR can be divided into two main classes namely non-proliferative DR (NPDR) and proliferative DR (PDR), [6] and NPDR can be further classified into three classes namely mild, moderate, and severe.

Microaneurysm (M), hard exudates (HEs), soft exudates or cotton wool spots (CWS), hemorrhage (H), and neovascularisation are the lesions of DR [6]. M is small swelling on the walls of blood vessels inside retina that is caused due to loss of pericyte. Capillaries are not observable from conventional fundus images, Ms become visible like isolated red dots not attached to any blood vessel. Diabetic patients may get this abnormality in the retina, moreover, it is the initial detectable sign of DR. Hs lie in the inner part of the retina and are formed when Ms or walls of capillaries become fragile and get burst. It is similar to the M when small in size. HEs are formed when protein leaks from blood vessels. HEs are waxy and yellow or white deposits of protein and lipid which leaks from the arteries when arteries become weak due to Ms. CWS or soft exudates are formed when leakage of blood vessels block the vessels. They are fluffy and are white in colour. As the capillary break down progresses, the retina becomes ischemic and triggers the growth of the new cells as an attempt for revascularisation of the tissues deprived of oxygen. Neovascularisation is caused due to the abnormal progression of tiny and leaky blood vessels. These abnormal blood vessels are tenuous and can grow anywhere inside the retina. Figure 1 shows a pathological image having some of these lesions.

NPDR is caused when blood vessels are damaged inside the retina resulting in the leakage of blood or fluid. It soaks the retina and hence swells macula which affects the function of the retina. The lesions present at this stage of DR are M, HEs, soft exudates or CWS, and H [7]. Among the three types of NPDR, in mild NPDR, a few Ms are present inside the retina and some loss of vision is experienced by the patients. However, moderate NPDR can be detected by the presence of HE, CWS, and H, whereas in severe NPDR, HE and leakage of blood and fluid severely affect the retina.

PDR is an advanced stage of DR; at this stage blood vessels inside retina are obstructed. As a result, the retina is deprived of nutrition and thus sends signals for the nourishment; hence new blood vessels grow inside the retina. Although the birth of infant vessels is not harmful, however, due to their fragile nature, these may lead to leakage, loss of vision, or even blindness.

DR lesions can appear anywhere in the retina and this complication makes the detection of five DR-levels a difficult and tedious task for ophthalmologists and hence motivates researchers to design efficient CAD systems. Literature revealed that many researchers proposed efficient methods [8–13] for DR-lesion detection and CAD systems [14–25] for identification of severity levels of DR. Some of these studies proposed the automated methods to detect severity of DR on the basis of DR-lesions which is quite similar to the one usually practiced by ophthalmologists. Detection of DR-lesions depends on the expert domain knowledge as well as pre and post processing of images. In contrast, some of the methods [21, 22, 26, 27] used visual features without pre- and post-processing steps to diagnose DR and its severity levels.

Detection of exudates was proposed in [8, 10] using fuzzy *C*-means clustering after applying colour normalisation and local contrast enhancement as pre-processing steps. Segmented patches were classified as exudate and non-exudate using a neural network (NN) classifier. Comparative classification of exudates was also proposed by [11] using support vector machines (SVMs) and NNs.
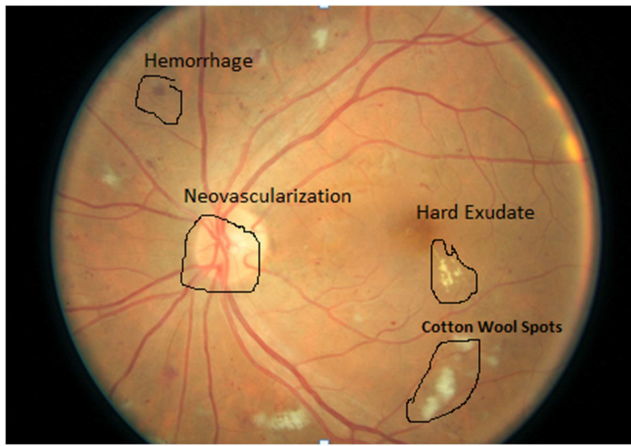
Fig. 1 *Pathological images with labelled anomalies*

SVM, nearest neighbor, and Naive Bayes classifiers were used to detect exudates candidates in [12]. Fifteen features were used in the investigation with no former segmentation for the detection of the candidates, instead of this the pixel-based features were computed including intensity, hue, number of edge pixels, the difference of Gaussian filter responses, and standard deviation of intensity. A pixel classification method was used to introduce a system for the extraction of red lesions [13]. K-nearest neighbour (KNN) classifier was used to classify the pixels of vessels and red lesions.

Detection of four severity levels of DR namely normal, moderate NPDR, severe NPDR, and PDR was proposed in [25]. Six features based on area and perimeter of red, green, and blue (RGB) layers were extracted from 120 retinal images after applying contrast enhancement method as pre-processing. Three-layered feed-forward NN (FFNN) was used for classification and achieved sensitivity and specificity of 90 and 100%, respectively. Although, the authors have shown an effort to achieve better efficiency of classification they could not detect mild NPDR. Retinal images were classified only into three classes of DR namely normal, moderate NPDR, and severe NPDR using a tree-type classifier Random Forests in [14]. Adaptive histogram equalisation was applied for contrast enhancement and median filters for removal of noise. Blood vessels were segmented using the matched filter. The global threshold was used to convert the filtered matched image into its binary image. Normal blood vessels were eliminated by using bounding box technique before H candidate detection which was identified by transforming the bright values in hue saturation value space and then applying gamma correction on RGB to highlight brown regions. Authors reported 90% accuracy in normal case and 87.5% accuracy in moderate and severe NPDR cases. However, the authors did not identify mild NPDR and PDR and therefore, the algorithm is not based on the characteristics of blood vessels.

The bag of words approach with scale-invariant feature transform (SIFT) and SVM classifier was used in [21] to detect only three stages of DR namely normal, NPDR, and PDR. In total, 64-bin histograms were created and neighbourhood of 3×3 for each pixel was considered to form a feature vector of 64-D. In total, 425 retinal images were manually assembled from publicly available well-known databases DIARETDB0, DIARETDB, STARE, and MESSIDOR for experiments and achieved 87.61% mean accuracy. An automatic screening system was proposed to classify retinal images into three classes namely normal, NPDR, and PDR [24]. The authors proposed a system that involved processing of fundus images for extraction of abnormal signs, such as the area of HEs, the area of blood vessels, bifurcation points, texture, and entropies. Thirteen statistically significant features were used to feed into Decision Tree, SVM and probabilistic NN (PNN). The proposed algorithm achieved 96.15% accuracy, 96.27% sensitivity, and 96.08% specificity using PNN classifier. Similarly, classification of DR into three stages such as normal, NPDR, and PDR was proposed using SVM in [15]. Morphological techniques and texture analysis methods were applied as image processing techniques. The detected features such as HEs,

homogeneity, and contrast and area of blood vessels were fed to the SVM. The reported results of classification were accuracy = 93%, sensitivity = 90%, and specificity = 100%. However, the proposed algorithms did not distinguish among levels of NPDR such as mild, moderate, and severe.

Identification of four stages of DR such as no DR, mild NPDR, moderate NPDR, and severe NPDR on the basis of the number of Ms was proposed in [16]. Contrast adjustment method was used in the inverted green channel for reduction of non-uniform illumination and for the contrast enhancement. Then Median filter was applied to pre-processed images for removal of noise. This was followed by an extended minima transformation. Ten images were tested to investigate the performance of the designed algorithm and compared the result with hand-drawn ground-truth images by ophthalmologists. Sensitivity and predictive values were used for evaluation and reported as 98.89 and 89.70%, respectively. The authors did not consider PDR cases in the study. DR lesions such as Ms, HE, and CWS were detected using the bag of visual words (BoVW) [26]. Speed up robust feature (SURF) and mid-soft coding with max-pooling were applied. DR1, DR2, and MESSIDOR databases were used and achieved an area under the curve (AUC) of 97.8% for exudates detection and AUC of 93.5% while detecting red lesions. However, the authors used different data sets to validate the proposed method, but the method could only be used to detect DR lesions and after that, the results could be used for identification of severity levels manually on the basis of type and count of lesions.

An automated system for the detection and classification of DR was proposed in [17]. The proposed algorithm was designed to discriminate normal and abnormal images, then abnormal images were further classified into three classes of NPDR and reported 98.52% accuracy. Mean- and variance-based techniques were applied for subtraction of background and removal of noise using saturation, hue, and intensity channels. Images obtained by applying the adaptive contrast enhancement method were used as the input for Gabor filter banks to detect the DR lesions. A feature vector based on colour, intensity, shape, and statistical features was designed for the classification of NPDR stages using modified m-Mediods-based classifier with the Gaussian mixture model. Although authors reported better accuracy of classification, they did not consider PDR level. Similarly, several DR lesions (such as fovea region), the thickness of vessels, and area of blood clotting were identified to detect normal, mild, moderate, and severe NPDR using KNN in [18]. However, the accuracy of the proposed system was not demonstrated, as the authors did not show numerical results. In [20], an automated method was proposed to detect four subsets of DR grades. In this study, normal, mild NPDR, moderate and severe NPDR, and PDR were identified by reporting 85% accuracy. Since the authors did not discriminate between severe and moderate classes, the proposed algorithm could only characterise blood vessels.

Bag of words was implemented to classify retinal images as normal and abnormal in [27]. Identification of severity levels of DR was not considered in this case. SVM, Naïve Bayes, FFNN, Decision Tree, and OR Logic classifiers were used. Similarly, the classification of retinal images into five stages of DR using NN and SVM was proposed with 80% accuracy in [19]. Authors used the modified region growing method for segmentation of optic disk and morphological operations for blood vessels. In this study, normal and abnormal images were first classified using NN with the features based on mean, variance, area, and entropy then SVM was used for classification of abnormal images into four stages. Since the proposed method used pre- and post-processing steps the technique was computationally expensive for the classification of DR into five stages.

An automated method for diagnosis of five severity levels of DR was proposed in [23]. Retinal images were taken from Kaggle data set and were pre-processed for colour normalisation using OpenCV package. Images were resized before applying convolutional NN (CNN) for classification task and reported the results with sensitivity = 30%, specificity = 95%, and accuracy = 75%. However, the authors had put great efforts for the classification but achieved low sensitivity. Grading of DR on two
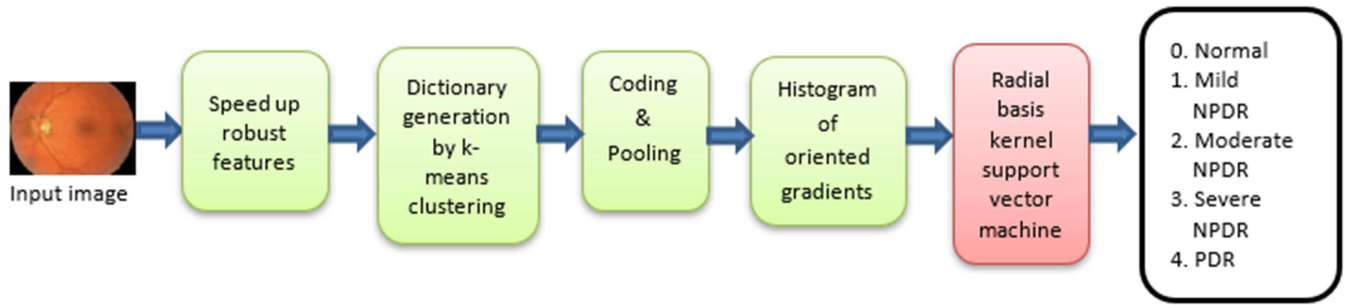
Fig. 2 *Architecture of proposed methodology*

privately collected datasets (8788 images and 1745 images) was reported in [28]. Retinal images were pre-processed and used to train CNN for multiple binary classification. The algorithm was designed to predict whether the retinal image belonged to (i) moderate or worse DR only, (ii) severe or worse DR only, (iii) referable diabetic macular edema (DMO) only, or (iv) fully gradable. The reported sensitivity for moderate or worse DR was 90 and 87%, respectively, whereas 98% specificity was reported for both data sets. In total, 84 and 88% of sensitivity and 99 and 98% of specificity were achieved for severe or worse DR and 91 and 90% of sensitivity and 98 and 99% of specificity were achieved for DMO only, respectively.

In [29], the authors applied CNN to discriminate stages of NPDR such as normal, mild, moderate, and severe. Data set was obtained by collecting images from Kaggle and Messidor. The authors designed classification models of secondary, tertiary, and quaternary stages. The transfer learning-based approach was applied after pre-processing, data augmentation, and training. Sensitivity was recorded as 85, 29, and 75% for no DR, mild, and severe DR, respectively. However, in quaternary classification, the authors described that the deep CNN was unable to discriminate multiclass classification. It is also noticeable that the PDR case was not considered in this study.

Deep visual features (DVF) were used [22] for classification of DR into five stages namely normal, mild NPDR, moderate NPDR, severe NPDR, and PDR. The authors used dense Color-SIFT to extract points of interest and associated SIFT descriptors, then, gradient location–orientation histogram was applied to it. The log-polar location grid method was used to compute SIFT descriptors. After normalising these features DVF and deep NN were used to classify 750 images (150 for each class) and reported 92.18% of sensitivity and 94.50% of specificity on an average. In this study, the retinal data set of various stages were assembled from different databases: 60 and 36 mild NPDR from DIARETDB1 and MESSIDOR, respectively, 40 and 396 images of foveal avascular zone from MESSIDOR in which 12 and 88 were from normal, 4 and 96 were from moderate, 12 and 88 were from severe NPDR, and 12 and 88 were from PDR, respectively. In total, 250 images (50 for each class) were collected from Private Hospital Universitario Puerta del Mar (HUPM, Cádiz, Spain). However, the authors reported a better accuracy of classification for five severity levels of DR, but 92.18% of sensitivity and 94.5% of specificity still need to be improved as sensitivity and specificity have great significance in medical diagnosis.

From the above discussion of the current progress of automatic methods for grading of DR, some facts can be emphasised. Some of the previous methods rely more on the precise segmentation of DR lesions which is hard to achieve and moreover, expensive computationally. Furthermore, errors in the segmentation process can affect the performance of the CAD systems. In addition to the above discussion, it is noticeable that previous methods proposed the techniques to distinguish between NPDR and PDR. Grading of severity levels of DR was proposed by only a few studies including [19, 22, 23]. These techniques were developed for the classification of five severity levels of DR. The method proposed in [19] used pre-processing methods and reported 80% accuracy; CNN was used in [23] and 30% sensitivity was reported, which was considerably low; and in [22] a visual features-based approach without pre-processing was proposed and achieved significant

results (sensitivity = 92.18% and specificity = 94.50%) but it needs to be improved since sensitivity and specificity have high significance in medical diagnosis. Therefore, the task of identification of severity levels remains to be a challenge.

2 Methodology

Automated severity detection of DR (SDDR) is proposed in the current research through bag of features (BoF) technique. It is an adaptive approach to represent the image in a robust way and is used for the classification of images in computer vision. Adaptiveness of BoF is one of its advantages as it allows image collection to be processed and it also identifies visual patterns of the whole image collection [30]. The approach is proposed to detect five severity levels of DR and its architecture is shown in Fig. 2. The details of each step are given in the following subsections.

2.1 Creation of codebook/dictionary

Feature extraction is the key step in image classification problems. Moreover, classification results depend on the selection of features which is a difficult task. In the current approach, BoF containing visual features is used for classification of retinal images into five stages. Local features are extracted from the retinal images using SURF [31] descriptor. SURF quickly computes distinctive descriptors, which is the main advantage of it. In addition, it is invariant to image transformations such as image rotation, illumination changes, scale changes, and minor change in the viewpoint. Moreover, SURF is a good feature descriptor of low-level representation. The construction process of SURF includes interest point detection, major point localisation in scale space, and orientation assignment.

Laplacian of Gaussian (LoG) approximations with box filters are used to estimate second-order Gaussian kernel $(\partial^2/\partial x^2)g(\sigma)$. To calculate intensities of rectangles within the input image LoG filters of size 9×9 with $\sigma = 1.2$ are used. Grid selection method having grid step of [8 8] and the block width of [32 64 96 128] is used in the computation of points of interest (PoI) for SURF descriptor. Haar wavelet responses of size 4σ in the direction of x and y are calculated to compute the primary direction of features. A square window descriptor of size $20 \times \sigma$ is constructed around each PoI. Each square window is divided into 4×4 subregions. Haar wavelets of 2 s are calculated within each subregion, hence the total length of each feature descriptor is $4 \times 4 \times 4 = 64$. In order to generate a codebook, K-means clustering algorithm is used and all the local features are clustered together independently using $K = 500$. Here K indicates the size of dictionary/codebook, however, codebook size is not significant in medical images [32]. An average of 32,770 vectors was identified for all lesions as well as for normal features to form 500 clusters.

2.2 Feature encoding and pooling

The idea of creating BoF has some cons: one is some valuable information can be lost during quantisation of visual words and other is the loss of spatial information. In order to solve such problems, coding and pooling are applied in the current approach.

Midlevel features are designed using coding for compact representation of local features and to preserve relevant

IET Comput. Vis., 2019, Vol. 13 Iss. 5, pp. 523-530

525

information. Consider F to be the set of descriptors with P-dimension for each image in the training set such that $F = \{f_1, f_2, \ldots, f_N\} \in R^{P \times N}$ and a visual dictionary of codewords is $C = \{c_1, c_2, \ldots, c_M\} \in R^{P \times M}$. Purpose of encoding is to calculate code for F with C. As a result, each descriptor f_i is assigned to the nearest visual feature within the dictionary by using $\mathrm{argmin}_j f_i - c_j^2$. Thus a vector U is formed that contains corresponding words of each descriptor; it is usually termed as hard assignment coding [33]. Visual dictionary is represented as $C = C\{j\}$, where $j = 1, 2, \ldots, M$.

A process to accumulate several local descriptor encodings into a single representation is termed as pooling. It is considered as one of the crucial steps in the BoVW representation and is followed by coding $g\{\alpha_j\}; j = 1, 2, \ldots, N$, where α indicates the assigned codewords to local feature vectors. Pooling is attained by two methods; one is summation and other is taking the maximum response, but max-pooling is an effective choice [34]. In the present approach, the max-pooling method is used in which the largest value is selected from the midlevel features that are corresponding to the codewords (here the max-pooling corresponds to the number of words with high frequencies).

$$g\{\alpha_j\} = Z, \quad Z = \mathrm{Max}(\alpha_{mj}), \quad (1)$$

Histograms of oriented gradient (HOG) is also constructed for each image to combine with the SURF visual features, Z in (1) represents the bag of visual features. The feature vector used in the present work is of dimension 390×500.

2.3 Identification of severity level

The classification of severity levels of DR into five classes namely normal, mild NPDR, moderate NPDR, severe NPDR, and PDR are identified using bag of visual features with two classifiers SVM and artificial NN (ANN).

2.3.1 Support vector machine:
A supervised learning algorithm SVM [35] is used for the classification of input images. If the classification problem was linear then SVM generated two hyperplanes (margin = $2/|w|$), such that no sample points lie between these two planes. If training data is a set of points f_i and if y_i is their labels, then the hyperplane's equation is $y_i(wf_i + b) = 0$, where w is weight vector while b denotes bias. All samples with $y = 1$ lie on one side of the plane and belong to one class and samples with $y = -1$ lie on the other side of the hyperplane and belong to the other class. In the present case, the classification problem is not linear, the images are classified into five classes, i.e. no DR, mild DR, moderate NPDR, severe NPDR, and NPDR. For classification in higher dimensions, SVM with radial basis function (RBF) kernel were used, which can be defined as (2).

$$k(f, f_i) = \exp\left(-\gamma\|f - f_i\|^2\right) \quad (2)$$

$k(f, f_i)$ is the kernel for samples (f, f_i), $\gamma = 1/2\sigma^2$ with $\sigma = 1$, and $\|f - f_i\|^2$ is the square of Euclidean distance between two feature vectors. Experiments are repeated using a ten-fold cross-validation method.

2.3.2 Artificial neural network:
Nonlinear four-layered ANN with backpropagation is used, which consists of one input layer, two hidden layers, and one output layer. ANN configuration consists of 500 input nodes, 50 hidden units on each hidden layer, and five output nodes that correspond to normal, mild NPDR, moderate NPDR, severe NPDR, and PDR. Backpropagation algorithm is used for the training of network. Gradient descent is implemented to reduce mean squared error between actual error rate and network output. The network is used until one of the following conditions is satisfied.

(i). Maximum gradient.

(ii). Maximum epoch.
(iii). Maximum goal.

An activation function 'log sigmoid' is implemented on the first hidden layer. The second hidden layer is connected to the output layer and a transfer function 'softmax' is used for the generation of output. Only layers and their connections are made during the construction of ANN. The Nguyen–Widrow method is used to initialise the values of these connections and biases. It distributes the active region of each neuron according to the input space. Although the values are assigned randomly in the active region, slightly different results were achieved on each iteration. Then according to the data sets, these randomly assigned values of parameters are trained.

The scaled conjugate backpropagation technique is applied for the training of ANN according to the obtained visual features. For avoiding overfitting, validation check is performed; it also monitors the performance of updated parameters. Results are compiled at six validation checks and 19 epochs with a gradient equal to 0.055387 and values of bias and weights are saved, in which the minimum validation error has occurred. Performance of the network is measured by the accurate classification of test samples.

3 Experimental setup

In the current method visual dictionary is created by extracting visual features of each image through SURF and HOG using grid selection of [8 8], block width of [32 64 96 128] with the standard deviation of $\sigma = 1.2$, then the extracted visual features are grouped using K-means clustering algorithm with $k = 500$. Each centre of the cluster is considered as a codeword and the collection of these codewords form a dictionary. These visual features are used for classification of retinal images into five classes through SVM and ANN. SVM with RBF kernel using $\sigma = 1$ is applied to discriminate the visual features. For validation of results, ten-fold cross-validation checks are performed by preserving the same ratio of images on each fold. The proposed approach is also tested using ANN and compared with the results achieved by SVM and ANN. The structure of ANN includes one input layer with 500 nodes, two hidden layers with 50 active nodes and one output layer with five nodes. The results of ANN are compiled on 6 validation checks and 19 epochs with a gradient equal to 0.055387. Confusion matrices are computed to show the statistical results.

To test the proposed approach, data set is taken from the Kaggle National Data Science Bowl, Kaggle is a platform founded by Anthony John Goldbloom in April 2010 for analytical competitions of machine learning and for predictive modelling.

The data set for detection of DR containing 35126 images was released by Kaggle [36] for a competition announced in 2015. Retinal images were having a high resolution of around 6 megapixels in 24-bit depth, annotated with patient ID and left and right eye. In the current study, 390 images (78 of each class) from 35,126 were selected using a random sampling method. The selected set of 390 retinal images, include 78 normal and 312 pathological images. From the selected data set, 70% images were used for training, 15% for testing, and 15% for validation. An example of a healthy retina and retinal images having four severity levels of DR are shown in Fig. 3.

The proposed SDDR system was implemented in MATLAB R2015b on a running operating system (Windows 10) Intel processor system with 8 GB RAM, Core i3 64-bit. The feature extraction part took 6.32 s per image on an average, and the training of extracted features that was performed by SVM took an average of 2.57 s. However, when the test was performed the image needed an average of 6.44 s time for classification.

3.1 Experimental results

In this section, the detailed quantitative analysis of the proposed SDDR system is given. For performance evaluation and comparison of proposed SDDR system with state of the art methods, sensitivity, specificity, positive predictive value (PPV),
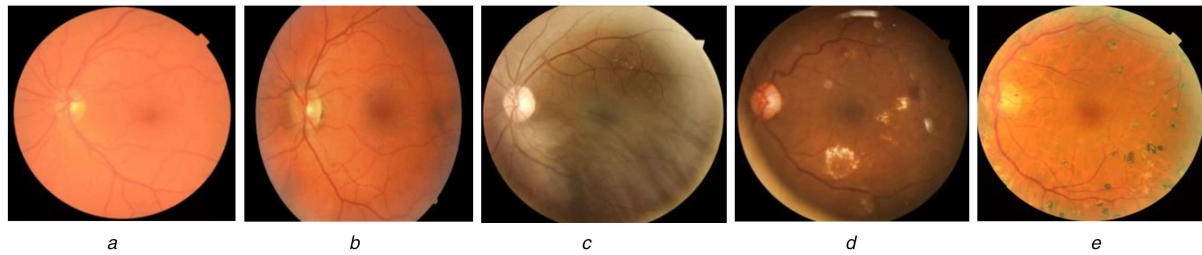
526

IET Comput. Vis., 2019, Vol. 13 Iss. 5, pp. 523-530

Fig. 3 *Example of pathological images arranged in increasing severity levels of DR*
(a) No DR, *(b)* Mild NPDR, *(c)* Moderate NPDR, *(d)* Severe NPDR, and *(e)* PDR

Fig. 4 *Classification of DR grading using SVM*

Fig. 5 *Classification of DR grading using ANN*

Table 1 Description of TP, FP, TN, and FN for classification of retinal images

Description	Normal image in classification	Image affected by DR in classification
normal image in actual	TP	FN
image affected by DR in actual	FP	TN

classes are classified correctly, and 68 images of moderate NPDR class are classified correctly. The results computed from the confusion matrix mentioned in Fig. 4 are shown in Table 2. These results show TP, TN, FP, and FN of sensitivity, specificity PPV, and accuracy of the proposed SDDR system though SVM. The calculated indices are sensitivity = 95.92%, specificity = 98.90%, PPV = 95.74%, and accuracy = 98.30%.

The output of ANN is shown in Fig. 5; it shows that 62 images are correctly classified in the normal case, 76 images in the mild NPDR class, 62, 69, and 58 images in the moderate NPDR, severe NPDR, and PDR cases, respectively. TP, TN, FP, and FN are given in Table 3 and are used to calculate sensitivity, specificity, PPV, and accuracy of the proposed SDDR system through ANN: sensitivity=83.83%, specificity=95.97%, PPV=83.82% and accuracy=92.92%.

It can be noticed from Table 4 that the proposed SDDR system achieved better results (shown in bold) while using visual features + SVM. By using visual features + SVM, the proposed system gives the best results in PDR case (sensitivity = 100%, specificity = 99%) and then in normal cases (sensitivity = 100%, specificity = 98.7%), in mild NPDR cases (sensitivity = 96.2%, specificity = 100%), in severe NPDR cases (sensitivity = 96.2%, specificity = 97.8%), and in moderate cases (sensitivity = 95.92% and specificity = 99%). On an average, the proposed SDDR system achieved sensitivity of 95.92% and specificity of 98.90% using visual features + SVM.

3.2 Comparative study

It is significant to note that all the existing automated systems have used different databases, but the purpose of comparison is to show that the system proposed in the present study has performed well when the authors compared the sensitivity and specificity with other studies. For the comprehensive comparative study, they compared the results achieved by the current approach with the previous methods of pre- and post-processing methods, CNN classification method, and visual dictionary-based approach. The authors compared the sensitivity and specificity of each severity level of DR (normal/no DR, mild, moderate and severe NPDR, and PDR) achieved by the proposed SDDR system with some state of the art methods. Table 5 shows the comparison of the results of the proposed and the previous methods.

It is observable that the comparison in Table 5 is made on the basis of achieved sensitivity and specificity of individual severity level and the average time that an image needed to get classified. The techniques of pre- and post-processing for the classification of retinal images into only three classes were proposed in [14], retinal images were classified into five severity levels using CNN in [23], Carson Lam *et al.* [28] proposed secondary, tertiary, and quaternary classification using deep CNN while the classification of retinal images into five severity levels using visual features with deep

and accuracy are computed. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) values are calculated for each severity level of DR. In the present study, TP means that a person is affected by the disease and is diagnosed correctly, TN means that a person's actual and estimated values are negative. FP means that the person does not have the disease but is diagnosed positive. However, FN is followed when the person is diagnosed negative but actually has the disease, this description is summarised in Table 1. Results achieved using visual features with SVM and ANN are computed in the form of confusion matrices and shown in Figs. 4 and 5, respectively. These confusion matrices are used to calculate the sensitivity, specificity, PPV, and accuracy indices. In the conventional analysis for a diagnostic test, sensitivity and specificity are considered as primary indices of accuracy [37].

It is defined earlier that there are 390 retinal images containing 78 images of each class. Figure 4 shows the output of SVM classifier which indicates that all the normal and PDR images are classified correctly, 75 images of each mild and severe NPDR

IET Comput. Vis., 2019, Vol. 13 Iss. 5, pp. 523-530

527

learning NN was proposed in [22], therefore, the authors compared the sensitivity and the specificity of each class individually. It can also be noticed from Table 5 that our proposed method is taking less time to classify a retinal image as compared with [22]. The algorithm proposed in [23] took lesser running time per image but the achieved sensitivity was very low.

It is necessary to discuss that in MESSIDOR database [38], some of the images were marked wrong such as image Base 11/20051020_63045_0100_PP.tif was marked as an image with severe NPDR while according to the latest updates, it is a normal image. Similarly, image Base 11/20051020_64007_0100_PP.tif belongs to severe NPDR instead of mild NPDR, image Base 11/20051020_63936_0100_PP.tif relates to mild NPDR instead of severe NPDR, and image Base 13/20060523_48477_0100_PP.tif belongs to the class severe NPDR instead of moderate NPDR.

Similarly, image Base 20051202_55626_0400_PP.tif is now marked as moderate NPDR and 20051205_33025_0400_PP.tif is marked as severe NPDR. This update is available on MESSIDOR webpage and can affect the results of the previously proposed methodologies that were tested on these images. This given update can increase or decrease the efficiency of an automated system.

4 Discussion

People with uncontrolled diabetes fall under the condition of DR, therefore, its diagnosis at the earlier stage is essential, as it impairs the retina if it remains to be undiagnosed. Hence, it requires an immediate need to consult an ophthalmologist. Since eye examinations are expensive, therefore, a large number of people are deprived of adequate treatment. Currently, the ophthalmologists use manual methods for screening of DR; these methods are expensive and require medical experts who have extensive domain knowledge. In contrast, automatic methods for diagnosis of DR use image processing and machine learning techniques to yield better and consistent results. The technique of automatic diagnosis is exclusively used nowadays through more established and refined CAD systems. Medical images are used in CAD systems to detect lesions of the retina for the diagnosis. It can provide many advantages, for instance, it reduces time, manpower, and cost while analysing a large set of images. Identification of five stages of DR is essential to detect the exact type of DR. Therefore, the present work focused on the detection of five stages of DR.

The viability of an automated detection and recognition system can be demonstrated by its results, therefore, to test the proposed SDDR system the authors used 390 retinal images. Sensitivity and specificity achieved by the proposed SDDR system show that on an average, many retinal images are correctly classified. However, the proposed SDDR system misclassified a few images. It can be noticed from Table 2 that the normal and PDR classes achieved the highest sensitivity of 100% and moderate class achieved the lowest sensitivity of 87.2%. The reason is that the approach used in the current work makes a dictionary of PoI. As normal class consisted of all healthy retinas having only normal features, the detected PoI of all healthy images were the same and had almost the same frequency. PDR has all lesions in abundance including neovascularisation, thus, the frequency of each PoI is high for all PDR images and it is made easy for the proposed automated system to recognise a subject. In the case of moderate NPDR, the

Table 2 Evaluation performance for DR grading using visual features + SVM

DR stage	TP	FN	FP	TN	Sensitivity, %	Specificity, %	PPV, %	Accuracy, %
normal	78	0	4	308	100	98.7	95.1	99
mild NPDR	75	3	0	312	96.2	100	100	99.2
moderate NPDR	68	10	3	309	87.2	99	95.8	96.7
sever NPDR	75	3	7	305	96.2	97.8	91.5	97.4
PDR	78	0	3	309	100	99	96.3	99.2
overall result					95.92	98.90	95.74	98.30

Table 3 Evaluation performance for DR grading using visual features + ANN

DR stage	TP	FN	FP	TN	Sensitivity, %	Specificity, %	PPV, %	Accuracy, %
normal	62	12	19	297	83.78	93.99	76.54	92.05
mild NPDR	76	14	7	293	84.44	97.67	91.57	94.62
moderate NPDR	62	11	12	305	84.96	96.21	83.78	91.03
sever NPDR	69	14	14	293	83.13	95.44	83.13	92.82
PDR	58	12	11	309	82.86	96.56	84.06	94.10
overall result					83.83	95.97	83.82	92.92

Table 4 Performance comparison of SVM and ANN

Proposed method	Sensitivity, %	Specificity, %	PPV, %	Accuracy, %
visual features + ANN	83.83	95.97	83.82	92.92
visual features + SVM	**95.92**	**98.90**	**95.74**	**98.30**

Table 5 Performance and time per image comparison of the proposed and previous state of the art methods on the basis of DR severity levels

Methods	Verma et al. [14]		Carson Lam et al. [28]		Pratt et al. [23]		Abbas et al. [22]		Proposed method	
Severity levels	Sn, %	Sp, %	Sn, %	Sp, %	Sn, %	Sp, %	Sn, %	Sp, %	Sn, %	Sp, %
normal/no DR	67.70	72.60	85	—	95	61.78	94.10	97.60	**100**	**98.7**
mild NPDR	71.30	75.43	29	—	0	100	92.23	96.99	**96.2**	**100**
moderate NPDR	—	—	—	—	22.95	94.92	88.45	93.25	87.2	**99**
severe NPDR	78.50	83.34	75	—	7.81	99.82	88.43	92.16	96.2	**97.8**
PDR	—	—	—	—	44.33	98.22	89.30	92.21	**100**	**99**
total Result	—	—	—	—	30	95	92.18	94.50	**95.92**	**98.90**
time taken per image	—		—		0.04 s		07 s		6.44 s	

Sn, sensitivity; Sp, specificity.

reason for achieving the lowest sensitivity is that since the number of Ms in moderate NPDR should be >5 and <15 and the number of Hs ranges from 0 to 4, while in the case of severe NPDR the number of Ms should be >15 and H is greater than or equal to 5. Thus, Ms can be misclassified as H if large in size and can misclassify the image from moderate to severe NPDR. Owing to this reason, seven images of moderate NPDR were misclassified as severe NPDR, which decreased the sensitivity of this class.

In the past decades, the researchers relied on the segmentation of some components of the retinal images and focused more on the methods of pre- and post-processing to detect the limited stages of DR. These methods consider segmentation as a pre-processing step and extraction, selection, and classification of features as post-processing steps. Furthermore, CNN is considered as one of the deep learning architectures and was used for classification of medical images in the field of medical diagnosis, but it is hard to train its model on image pixels, in practice. Therefore, in this study, the authors used visual features to detect five severity levels of DR using Kaggle data set. It was also noticed that MESSIDOR is the only database for DR that provides images for four stages of DR, images for PDR are not provided. Kaggle data set provided retinal images for five severity levels of DR, which are used in a few studies including [23, 28]. It is shown in the literature review that many researchers [14–25] have devoted their efforts for the detection of the stages of DR. Detection of five stages of DR was proposed in [23] using Kaggle data set and CNN and in [22] through visual features and deep NN using private data collected from a local hospital.

The authors proposed an improved severity detection system which identifies five severity levels of DR through visual features and SVM and achieved better results (sensitivity = 95.92%, specificity = 98.90%) when compared with the automated system for identification of five severity levels of DR proposed in [22, 23]. The advantage of the proposed SDDR system with regard to the evaluation of severity levels of DR by medical experts is its high sensitivity and specificity. It can eradicate the need of medical experts in the examination of diabetic retinal images and a CAD system can be designed to recognise the DR cases according to its severity.

Dictionary-based approach is easy to apply due to its simplicity, and adaptive nature of BOF approach is one of its important advantages. Another advantage is robustness to obstruct, affine transfiguration and its efficiency of computation. All these characteristics are useful in the analysis of medical images. However, a huge set of features is a disadvantage of this algorithm since an increase in the number of features can proceed to be overfitting.

Misclassification of few images is the limitation of the proposed approach which the authors intend to overcome in the future. They have planned to use different colour schemes and to select the one which signifies the better representation of lesions in the images. They did not consider the DR with maculopathy in this study, therefore, the method can be applied for the classification of DR with and without maculopathy, i.e. the severity levels of DR as normal and mild NPDR with and without maculopathy, moderate NPDR with and without maculopathy, severe NPDR with and without maculopathy, and PDR with and without maculopathy. The method of the visual dictionary may help in the diagnosis of other diseases such as for detection of malignant melanoma, to classify colorectal tumours and for detection of polyps.

5 Conclusions

In the present work SDDR system is proposed for classification of five severity levels of DR using visual features and a radial basis kernel SVM. This study focuses on the identification of severity levels of DR without applying pre- and post-processing on the images. It can be noticed from the literature that many studies were devoted to the identification of limited severity levels. They emphasised more on the detection of DR lesions and the resultant severity level. This method of identification is still used by medical experts. In addition, the correct detection of DR lesions is difficult due to the miscellany of lesion appearance. In contrast, few researchers proposed automated methods for identification of five severity levels of DR, among those few used pre- and post-processing methods, while others worked on visual features approach. In contrast with these state of the art methods, the present study used visual features with a simple classifier. Hence, this study did not focus on the characterisation of lesions such as Ms, exudates, or blood vessels on the images, which ultimately reduced computational time and error propagation.

The present research mainly emphasised on accuracy and adaptability of the use of visual features (SURF and HOG) and radial basis kernel SVM. The evaluation of the proposed SDDR system was calculated using sensitivity, specificity, and accuracy and tested it on 390 images (78 per class). The obtained sensitivity is 95.92%, specificity is 98.90%, and accuracy is 98.30%. The specificity of 98.90% is considered as good in automated detection, specially, it is used for triaging. Table 5 provides the evidence that the proposed SDDR system has achieved better specificity than the previously proposed methods.

Finally, the proposed SDDR system is greatly significant for the automated diagnosis of DR and its five-stage detection. As detection of all stages of DR is a vital requirement in the highly growing rate of the disease, the sensitivity of 95.92% and specificity of 98.90% achieved by the proposed SDDR system can be used to integrate a CAD system with this one.

6 Acknowledgments

7 References

[1] Ahmed, R.A., Khalil, S.N., Al-Qahtani, M.A.: 'Diabetic retinopathy and the associated risk factors in diabetes type 2 patients in Abha, Saudi Arabia', *J. Family. Community. Med.*, 2016, **23**, (1), p. 18

[2] Alam, U., Arul-Devah, V., Javed, S., *et al.*: 'Vitamin D and diabetic complications: true or false prophet?', *Diabetes. Ther.*, 2016, **7**, (1), pp. 11–26

[3] Abràmoff, M.D., Garvin, M.K., Sonka, M.: 'Retinal imaging and image analysis', *IEEE Rev. Biomed. Eng.*, 2010, **3**, pp. 169–208

[4] Solomon, S.D., Chew, E., Duh, E.J., *et al.*: 'Diabetic retinopathy: a position statement by the American Diabetes Association', *Diabetes Care*, 2017, **40**, (3), pp. 412–418

[5] Yau, J.W., Rogers, S.L., Kawasaki, R., *et al.*: 'Global prevalence and major risk factors of diabetic retinopathy', *Diabetes Care*, 2012, **35**, (3), pp. 556–564

[6] Bhaskaranand, M., Ramachandra, C., Bhat, S., *et al.*: 'Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis', *J. Diabetes Sci. Technol.*, 2016, **10**, (2), pp. 254–261

[7] Welikala, R.A.: 'Automated detection of proliferative diabetic retinopathy from retinal images'. PhD thesis, Kingston University, 2014

[8] Osareh, A., Mirmehdi, M., Thomas, B., *et al.*: 'Automatic recognition of exudative maculopathy using fuzzy *C*-means clustering and neural networks'. Proc. Medical Image Understanding Analysis Conf, UK, November 2001, pp. 49–52

[9] Osareh, A., Mirmehdi, M., Thomas, B., *et al.*: 'Classification and localisation of diabetic-related eye disease'. European Conf. Computer Vision, London, UK, May 2002, pp. 502–516

[10] Osareh, A., Mirmehdi, M., Thomas, B., *et al.*: 'Automated identification of diabetic retinal exudates in digital colour images', *Br. J. Ophthalmol.*, 2003, **87**, (10), pp. 1220–1223

[11] Osareh, A., Mirmehdi, M., Thomas, B., *et al.*: 'Comparative exudate classification using support vector machines and neural networks'. Int. Conf. Medical Image Computing and Computer-Assisted Intervention, London, UK, September 2002, pp. 413–420

[12] Sopharak, A., Dailey, M.N., Uyyanonvara, B., *et al.*: 'Machine learning approach to automatic exudate detection in retinal images from diabetic patients', *J. Mod. Opt.*, 2010, **57**, (2), pp. 124–135

[13] Niemeijer, M., Van Ginneken, B., Staal, J., *et al.*: 'Automatic detection of red lesions in digital color fundus photographs', *IEEE Trans. Med. Imaging*, 2005, **24**, (5), pp. 584–592

[14] Verma, K., Deep, P., Ramakrishnan, A.: 'Detection and classification of diabetic retinopathy using retinal images'. Proc. Int. Conf. India Conf. (INDICON), Hyderabad, India, December 2011, pp. 1–6

[15] Du, N., Li, Y.: 'Automated identification of diabetic retinopathy stages using support vector machine'. Proc. Int. Conf. Chinese Control Conf., Xi'an, China, July 2013, pp. 3882–3886

[16] Alaimahal, A., Vasuki, D.S.: 'Identification of diabetic retinopathy stages in human retinal image', *Int. J. Adv. Res. Comput. Eng. Technol.*, 2013, **2**, (2), pp. 808–811

[17] Akram, M.U., Khalid, S., Tariq, A., *et al.*: 'Detection and classification of retinal lesions for grading of diabetic retinopathy', *Comput. Biol. Med.*, 2014, **45**, pp. 161–171

[18] Mishra, P.K., Sinha, A., Teja, K.R*., et al.*: 'A computational modeling for the detection of diabetic retinopathy severity', *Bioinformation*, 2014, **10**, (9), p. 556

[19] Prakash, N.B., Selvathi, D., Hemalakshmi, G.R.: 'Development of algorithm for dual stage classification to estimate severity level of diabetic retinopathy in retinal images using soft computing techniques', *Int. J. Electr. Eng. Inf.*, 2014, **6**, (4), p. 717

[20] Sri, R.M., Reddy, M.R., Rao, K.: 'Image processing for identifying different stages of diabetic retinopathy', *Int. J. Recent Trends Eng. Technol.*, 2014, **11**, (1), p. 83

[21] Venkatesan, R., Chandakkar, P., Li, B*., et al.*: 'Classification of diabetic retinopathy images using multi-class multiple-instance learning based on color correlogram features'. Proc. Int. Conf. Engineering in Medicine and Biology Society, San Diego, CA, USA, August 2012, pp. 1462–1465

[22] Abbas, Q., Fondon, I., Sarmiento, A*., et al.*: 'Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features', *Med. Biol. Eng. Comput.*, 2017, **55**, (11), pp. 1959–1974

[23] Pratt, H, Coenen, F., Broadbent, D.M.: 'Convolutional neural networks for diabetic retinopathy', *Proc Comput. Sci.*, 2016, **90**, pp. 200–205

[24] Mookiah, M.R.K., Acharya, U.R., Martis, R.J*., et al.*: 'Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach', *Knowl.-Based Syst.*, 2013, **39**, pp. 9–22

[25] Yun, W.L., Acharya, U.R., Venkatesh, Y.V*., et al.*: 'Identification of different stages of diabetic retinopathy using retinal optical images', *Inf. Sci.*, 2008, **178**, (1), pp. 106–121

[26] Pires, R., Jelinek, H.F., Wainer, J*., et al.*: 'Advancing bag-of-visual-words representations for lesion classification in retinal images', *PloS One*, 2014, **9**, (6), p. e96814

[27] Mukti, F.A., Eswaran, C., Hashim, N.: 'Detection and classification of diabetic retinopathy anomalies using Bag-of-words model', *J. Med. Imaging. Health. Inform.*, 2015, **5**, (5), pp. 1009–1019

[28] Gulshan, V., Peng, L., Coram, M*., et al.*: 'Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs', *JAMA*, 2016, **316**, (22), pp. 2402–2410

[29] Lam, C., Yi, D., Guo, M*., et al.*: 'Automated detection of diabetic retinopathy using deep learning', *AMIA Summits Translational Sci. Proc.*, 2018, **2017**, p. 147

[30] Law, M.T., Thome, N., Cord, M.: 'Bag-of-words image representation: key ideas and further insight', in Ionescu, B. '*Fusion in computer vision*' (Switzerland: International publishing Springer, 2014), pp. 29–52

[31] Bay, H., Ess, A., Tuytelaars, T*., et al.*: 'Speeded-up robust features (SURF)', *Comput. Vis. Image Underst.*, 2008, **110**, (3), pp. 346–359

[32] Tommasi, T., Orabona, F., Caputo, B.: 'CLEF2007 image annotation task: an SVM-based cue integration approach'. Proc. ImageCLEF 2007-LNCS, 2007

[33] Wang, X., Wang, L., Qiao, Y.: 'A comparative study of encoding, pooling and normalization methods for action recognition'. Proc. Int. Conf. Asian Conf. on Computer Vision, Daejeon, Korea, November 2012, pp. 572–585

[34] Avila, S., Thome, N., Cord, M*., et al.*: 'Pooling in image representation: the visual codeword point of view', *Comput. Vis. Image Underst.*, 2013, **117**, (5), pp. 453–465

[35] Anzai, Y.: '*Pattern recognition and machine learning*' (Elsevier, USA, 2012, 1st edn. 2012)

[36] Kaggle: 'Diabetic retinopathy detection'. Available at https://www.kaggle.com/c/diabetic-retinopathy-detection/data, accessed 1 June 2015

[37] Hajian-Tilaki, K.: 'Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation', *Caspian. J. Intern. Med.*, 2013, **4**, (2), p. 627

[38] MESSIDOR: 'Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology'. Available at http://www.adcis.net/en/Download-Third-Party/Messidor.html, accessed 20 February 2018

530

IET Comput. Vis., 2019, Vol. 13 Iss. 5, pp. 523-530