Journal of Ambient Intelligence and Humanized Computing (2021) 12:2649–2670
https://doi.org/10.1007/s12652-020-02426-9

ORIGINAL RESEARCH

Hierarchical severity grade classification of non-proliferative diabetic retinopathy

Charu Bhardwaj[1] · Shruti Jain[1] · Meenakshi Sood[2]

Received: 17 March 2020 / Accepted: 25 July 2020 / Published online: 3 August 2020

Abstract
Curability of diabetic retinopathy (DR) abnormalities highly rely on regular monitoring, early-stage diagnosis and timely treatment. Detection and analysis of variation in eye images can help the patient to take the early action before progression of the disease. Vision loss can be effectively prevented by automated diagnostic system that assist the ophthalmologists who otherwise practice manual lesion detection processes which are tedious and time-consuming. This paper proposes a hierarchical severity level grading (HSG) system for the detection and classification of DR ailments. The retinal fundus images in the proposed HSG system are categorized as grade 0 (indicating Non-DR class) and DR severity grades 1, 2, 3 depending upon the number of anomalies; microaneurysms and haemorrhages in the fundus images. The challenge of retinal landmark segmentation, DR retinal discrimination and DR severity grading have been addressed in this work contributing to the novelty of the proposed approach. For non-DR and DR classification, the proposed system achieves an overall accuracy of 98.10% by SVM classifier and 100% by kNN classifier. Hierarchal discrimination into further grades of abnormalities resulted in accuracy values of 95.68% and 92.60% with SVM classifier using Gaussian kernel and, 97.90% and 95.30% employing fine kNN classifier. The HSG system demonstrates a clear improvement in accuracy with significantly less computational time comparative to the other state-of-the-art methods when applied to the MESSIDOR dataset. IDRiD dataset is also evaluated for performance validation of the proposed HSG system yielding a maximum of 94.00% classification accuracy using a kNN classifier with a computational time of 0.67 s.

Keywords Diabetic retinopathy · Gray- level co-occurrence matrix features · Statistical features · Support vector machine · k-nearest neighbour

1 Introduction

Retinal image analysis is an active research area in diabetic retinopathy (DR). It is the principal cause of severe eye complications or even blindness in the developed as well as developing countries. Several risk factors related to this disease are unhealthy lifestyle, obesity and aging. However, early diagnosis along with continuous periodic examination is a determinant factor in diminishing the menace of severe visual impairments. At the time of DR screening, retinal fundus images are captured by ophthalmic experts for detection and identification of the DR severity level (Yen and Leong 2008; Habib et al. 2017). There is a need for an automated DR diagnostic system, as fundus image evaluation becomes tedious due to the increased number of DR cases causing a burden on the ophthalmologists (Yen and Leong 2008).

DR can be graded into different grades depending upon the level of severity of the disease. It is broadly categorized into proliferative DR (PDR) and non-proliferative DR (NPDR). NPDR is further classified into three prominent categories: mild non-proliferative, moderate non-proliferative and severe non-proliferative stages. Healthy fundus images having no symptoms of diabetic retinopathy are graded as grade 0 images in DR severity grading. Severity level 1 is indicated by a mild non-proliferative stage in which microaneurysms (MAs) develop and appear as dark red lesions near the blood vessels. The moderate NPDR stage indicates severity level 2. In this stage, the number of MAs increases along with blood leakage into the retinal

✉ Charu Bhardwaj
cbcharubhardwaj215@gmail.com

[1] Department of Electronics and Communication Engineering, JUIT Waknaghat, Solan, HP, India

[2] Department of CDC, NITTTR, Chandigarh, India

surface causing haemorrhages (HMs). MAs and HMs are identified as red lesions as they are dark red in color having a similar intensity range. Exudates (EXs) or cotton wools (CWs) indicated as yellow lesions may also appear in severity grade 2. The severe non-proliferative DR stage graded as grade 3, consists of an increased number of DR lesions along with intra-retinal microvascular abnormality (IRMA) which leads to blockage in retinal blood vessels (Vaishnavi et al. 2016). Patients suffering from severe NPDR stage have higher chances of developing PDR in the near future. PDR is characterized by abnormal blood vessel overgrowth which leaks blood leading to vision loss ultimately causing blindness.

There have been numerous advances in the automated DR classification system to address different DR abnormalities, some of which are reported in this section. Roychowdhury et al. (2012) proposes a two-stage DR detection system in which optical disc (OD) localization is accomplished using minimum intensity maximum solidity overlap (*MinIMaS*) algorithm and Gaussian mixer model (GMM) is used to detect the presence of lesions. This system outperforms the state-of-the-art algorithms by diminishing the number of false positives in lesion classification providing area under the curve (AUC) of 0.9593 for bright lesion detection and 0.8663for red lesion detection. Habib et al. (2017) introduced an MA classification approach which detects an initial set of candidates using GMM and false positives are reduced in this approach employing tree ensemble classifier. The algorithm was evaluated on DIARETDB2 and MESSIDOR Dataset providing receiver operating characteristics (ROC) score of 0.415 outperforming the best available techniques. An automated DR grading system for the public database was introduced by Seoud et al. (2015) in which lesion probability map is generated for red lesion detection and a set of 35 features was fed to random forest classifier for classification. The system achieves 74.1% accuracy rate providing performance comparable to human experts. An automated assessment system was developed by Vaishnavi et al. (2016) using the SVM classifier and it provides the highest accuracy in detecting DR lesions. A scheme to detect retinal hemorrhage was proposed by Inbarathi and Karthikeyan (2014) for automated DR screening using the support vector machine (SVM) classifier. MESSIDOR dataset was considered utilizing Splat and GLCM features to obtain improved classification accuracy. Habib et al. (2016), also introduces MA detection approach evaluated on a subset of MESSIDOR dataset and extracted features were classified using random forest classifier. It provides improved sensitivity for MA detection as compared to other techniques reported in the literature. Lachure et al. (2015) proposed an MA and EX detection technique using SVM and K-nearest neighbor (kNN) classifier considering the MESSIDOR dataset. SVM provides better performance over kNN while considering structural and GLCM features for DR classification. Authors in Roychowdhury et al. (2013) presented a CAD system to grade DR severity using a machine learning approach on fundus images with a varying field of view and illumination. Several classifiers like SVM, kNN, AdaBoost and GMM are analysed to classify retinopathy lesions from non- lesions. It lead to the reduction in the number of features used for lesion detection and severity classification. A two-step hierarchical approach was validated using MESSIDOR database for classification. In the first step, lesion and non- lesion classification is done and in the second step, yellow lesions are classified into exudates and cotton wools and red lesions are classified into HMs and MAs. The system achieves sensitivity, specificity and area under the curve (AUC) of 100%, 53.16% and 0.904 respectively, which is better than the already reported method for image classification for DR. Authors in Goatman et al. (2010) presented a technique for automatic detection of proliferative DR by detecting vessel like segments using ridge strength measurement and watershed line technique. Fifteen features were selected for different candidate segments based on the shape, position, brightness, contrast, orientation and line density. SVM method was used for classification purposes to categorize between non-DR and DR affected retinal structures. AUC of 0.911 was obtained for new vessels detected originating from the optical disc. MA detection technique was proposed in Wang et al. (2016) which employs a dark object filtering process for MA candidate detection and the candidate cross-sectional profile was processed using singular spectrum analysis. A statistical feature set for candidate profiles is then extracted to be fed as a training set for k-nearest neighbour classification. The approach yields better results providing a great tool for DR screening and analysis. Texture features of retinal images were exploited in Morales et al. (2015) to distinguish between the healthy and pathological images. In this paper, local binary pattern (LBP) descriptors for retinal images are exploited and compared to the other descriptors to validate the effectiveness of the approach and it was proved a robust algorithm for DR screening and diagnosis using retinal texture features. Koh et al. (2018) developed a DR screening system to differentiate between non-DR and DR affected fundus image. Feature set after canonical correlation analysis was subjected to tenfold cross-validations and kNN classification achieving 96.21% accuracy. Karthikeyan and Alli (2018) presented an efficient DR disease classification method utilizing SVM parameters optimized using glowworm swarm optimization (GSO) and genetic algorithm (GA). SVM with hybrid GSO feature selection provides higher accuracy compared to existing DR classification techniques. This paper lags in considering computational complexity, therefore, the future work in this area will be focused on applying metaheuristic approaches. A machine learning bagging ensemble classifier (ML-BEC)

employing machine learning and ensemble classification method was designed to identify DR features (Somasundaram and Alli 2017). This system comprises two stages; the first stage extracts the candidate object features from retinal images and the second stage utilizes ensemble classifiers features for accurate analysis. Ensemble classifier achieves better classification accuracy with reduced DR classification time than single classification models. A microaneurysm monitoring system was presented by You et al. (2016) to enhance the DR screening process for a large population. The retinal images are subjected to analysis via cloud computing. Multi-orientation sum of matched filter method was used for microaneurysm detection and the classification is accomplished using the SVM approach. The results obtained demonstrate the feasibility of the systems which provides improved accuracy, speed and convenience. Bandyopadhyay et al. (2018) proposed an algorithm to detect the finest retinal blood vessels and the features of extracted blood vessels are fed to the kNN classifier for DR detection. Navarro et al. (2016) developed an automated image processing system for the detection of microaneurysms by transforming images into L*a*b* color space. MA candidate features are sent to the kNN classifier stage for final assessment which provides an accuracy of 84%. Although, the extensive literature has been done by the authors, some of the main research papers are tabulated in Table 1 depicting the algorithms proposed by various researchers for addressing DR severities.

The exhaustive literature survey reveals that there exist several research gaps in the existing work done in this area. The set of handcrafted features utilized does not provide significant improvement in the performance. Also, there is the effect of size of dataset in providing the desired performance as utilization of small dataset for automated detection does not provide satisfactory results. Therefore, the experimentation can be improved by increasing the annotated dataset and utilizing a different feature vector selection strategy as well as changing the choice of the kernel for SVM classifier.

1.1 Contribution

A hierarchical severity grading (HSG) system for DR classification is proposed in this work to help the ophthalmologists in expeditious and easy diabetic retinopathy detection which is capable of classifying the fundus images based on different DR severity levels. The retinal fundus images are graded as; Grade 0, Grade 1, Grade 2 and Grade 3, where Grade 0 indicates the non-DR class and the DR extremities are classified into three severity levels following the guidelines formulated by ophthalmic experts. This work extensively overcomes the challenge of retinal landmark segmentation, DR retinal discrimination and DR severity grading contributing to the novelty of the proposed approach. The major contribution of our HSG DR grading system lies in the critical analysis of multiple classifiers to find an optimal classification method for DR lesion detection and grading. The novelty of our model lies in the statistical analysis of the derived feature set for optimal feature selection. Classification outcomes are obtained for the statistically optimized feature set, to validate the significance of statistical analysis in DR severity classification and grading. Testing of the proposed system on the benchmark MESSIDOR dataset reveals its outperformance over the state-of-the-art methods for NPDR severity classification. The HSG approach is validated on another IDRiD dataset and uniformity in the experimental outcomes is observed, justifying the generalization capabilities of the proposed system for DR severity grading.

Table 1 Literature survey of different NPDR screening techniques

Algorithm proposed by	Image processing techniques	Database	Type of severity grading
Wulandari et al. (2019)	Statistical region merging segmentation along with convolutional neural network (CNN)	MESSIDOR	NPDR
Harangi et al. (2019)	Hand-crafter features and CNN	IDRiD	DR and diabetic macular edema (DME)
Al-Jarrah and Shatnawi (2017)	Morphology-based algorithm and artificial neural network (ANN)	DIARETDB1	NPDR
Ashraf et al. (2014)	Local binary pattern and support vector machine (SVM)	DIARETDB1	NPDR
Dupas et al. (2010)	Fundus image analysis algorithms	MESSIDOR	NPDR
Aptel et al. (2008)	Single-field non-mydriatic, single-field mydriatic, three-field non-mydriatic, three-field mydriatic	Local database having 79 patients	Both NPDR and PDR
Kahai et al. (2006)	Decision support system (DSS)	Local database having 143 images from Louisiana State University Eye Center	NPDR

The research article is organized as follows: Material and methods are described in Sect. 2 elaborating the dataset employed and proposed methodology. The results and discussions regarding experimentations performed is discussed in Sect. 3 along with the concluding remarks and future perspective of the proposed system in Sect. 4.

2 Materials and methods

The methodology adopted for the proposed HSG system employs the implementation strategy comprising retinal landmark segmentation, pathology identification and detection, feature extraction, selection and DR severity classification. Gray-level co-occurrence matrix features are extracted corresponding to each candidate object for lesion classification. These extracted features are statistically analysed for optimal feature selection and the optimal features are fed to the hierarchical classification stage.

2.1 Dataset

The authors have incorporated DRIVE dataset for OD localization and blood vessel extraction (Bhardwaj et al. 2018a, b; Bhardwaj et al. 2019). STARE dataset images are used for anomaly detection (Bhardwaj et al. 2020). The proposed grading system is evaluated on the benchmark standard MESSIDOR (Staal et al. 2004; Decencière et al. 2014) dataset for NPDR abnormality classification and severity grading, as it consists of a varying number of images with different resolutions. The feasibility of the proposed system is also validated using the latest Indian Diabetic Retinopathy Image Dataset (IDRiD) by increasing the number of patient samples to help physicians realize the progress of the disease. The list of publicly available fundus image datasets utilized for this work are detailed in Table 2.

MESSIDOR (Staal et al. 2004; Decencière et al. 2014) dataset comprises a total of 1200 fundus images acquired using a 3CCD camera at 45° FOV by 3 different ophthalmologic departments. These images were of standard sizes 1440×960, 2240×1488, or 2304×1536 pixels captured at 8 bits per color plane (Sisodia et al. 2017). The image size available in the MESSIDOR dataset comprises different sizes and resolution images. To make it resolution-independent, pre-processing operations are applied which includes image scaling, normalization, image denoising and contrast enhancement to normalize all the images to fit the same scale. MESSIDOR is a labelled dataset indicating four degrees of severities ranging from 0 to 3. Fundus images with no DR symptoms are graded as severity level 0; mild, moderate and severe DR grades are indicated by severity levels 1, 2 and 3 respectively.

The retinal fundus images in the IDRiD (Porwal et al. 2018) dataset were acquired from an Eye Clinic located in Nanded, (M.S.), India captured using a Kowa VX-10a digital fundus camera with 39 mm distance between lenses and the eye. This dataset was formed using 516 images of size 800 kB approximately with 4288×2848 pixel resolution and 50° FOV. The dataset consists of 454 fundus images containing NPDR severity symptoms and the remaining 62 contains PDR symptoms. This dataset contains clinically relevant images of the adequate quality representative of DR and diabetic macular edema tests performed on thousands of patients during the period 2009–2017.

The International Clinical Diabetic Retinopathy Disease Severity Scale (Wilkinson et al. 2003) referred in this work categorized the diabetic retinopathy from 0 to 4 severity levels. The DR severity grade 0 is denoted for the non-DR category, grade 1 denotes mild NPDR symptoms, moderate NPDR severity level is indicated by grade 2, grade 3 denotes severe NPDR symptoms and PDR symptoms are indicated by grade 4. The International Clinical Diabetic Retinopathy

Table 2 Dataset description

Dataset	Image size	FOV (degrees)	Number of images	Usage
DRIVE (Staal et al. 2004)	565×584	45	40	Structural analysis of retina
STARE (Hoover et al. 2000)	605×700	35	400	Abnormal blood vessel, exudates, haemorrhages and microaneurysms detection
DIARETDB1 (Kauppi et al. 2007)	1500×1152	50	89	Position and detailed labelling of retinal abnormalities
MESSIDOR (Staal et al. 2004; Decencière et al. 2014)[a]	1440×960, 2240×1488, 2304×1536	45	1200	DR severity grading
IDRiD (Porwal et al. 2018)[b]	4288×2848	50	516	DR severity grading

MESSIDOR: Methods for Evaluating Segmentation and Indexing technique Dedicated to Retinal Ophthalmology, 2004 [Online]. Available: https://www.adcis.net/en/Download-Third-Party/Messidor.html. Accessed: Feb. 10, 2019

IDRiD: Indian Diabetic Retinopathy Image Dataset, 2018 [Online]. Available: https://dx.doi.org/10.21227/H25W98. Accessed: May. 11, 2020

Disease Severity Scale ranging from 0 to 4 is detailed in Table 3.

This tabular representation reveals that the occurrence of NPDR severity relies on the presence of microaneurysms, haemorrhages and intra-retinal microvascular abnormalities (IRMA) and no signs of proliferative diabetic retinopathy (PDR) are seen till the third stage of International Clinical Diabetic Retinopathy Disease Severity grading. However, the PDR stage occurs due to retinal neovascularization or pre-retinal vitreous haemorrhage. This research article uses the NPDR severity grading scale range of the International Clinical Diabetic Retinopathy Disease Severity Scale ranging from 0 to 3.

Both MESSIDOR and IDRiD datasets used for the evaluation of the proposed (HSG) system provide DR severity level grading for individual images in the dataset in terms of retinopathy grade and risk of macular edema. For this research work, we have used retinopathy grades as a reference and their explanation along with the number of images present in the dataset are provided in Table 4.

Table 4 presents the number of images in each DR severity grade and also describes the retinopathy grades depending upon the number of microaneurysms and haemorrhages. As per International Clinical Diabetic Retinopathy Disease Severity Scale dataset guidelines, DR severity grading depends upon the number of MAs and HMs and therefore these lesions are combined to generate the DR grading. For the classification-stage 2, the feature vector of the lesion and non-lesion objects is scaled in the range of 0 and 1. Lesion objects are manually annotated as class label 1 and non-lesion objects are labelled as class label 0 depending upon the results of automated lesion detection as well as ground truth annotations. Feature vectors annotations for different classification stages are given in Table 5.

The complete set of images is separated into training and testing set using 70%-30% training and validation criteria. tenfold cross-validations are used to divide the training set into 10 distinct classes and every subset of training data is used to train the classifier against the other (10–1) validation set for the entire training phase. This validation scheme yields better training which provides the favourable outcomes for the classification task.

Table 3 International Clinical Diabetic Retinopathy Disease Severity Scale considering NPDR severities

Disease severity scale	Diabetic retinopathy grade	Description
0	Non-DR grade	No abnormalities
1	Mild NPDR grade	Microaneurysms only
2	Moderate NPDR grade	More number of microaneurysms but less than severe NPDR
3	Severe NPDR grade	More than 20 intra-retinal microaneurysms or hemorrhages Prominent intra-retinal microvascular abnormalities (IRMA)
4	PDR grade	Neovascularization or pre-retinal vitreous hemorrhage

Table 4 NPDR grades distribution for MESSIDOR and IDRiD datasets

Retinopathy grades	Grading description	Number of images in MESSIDOR dataset	Number of images in IDRiD dataset
Grade 0	$(N_{MA}=0)$ and $(N_{HM}=0)$	546	168
Grade 1	$(0<N_{MA}\leq 5)$ and $(N_{HM}=0)$	153	25
Grade 2	$(5<N_{MA}<15)$ or $(0<N_{HM}\leq 5)$	247	168
Grade 3	$(N_{MA}\geq 15)$ or $(N_{HM}\geq 5)$	254	93

N_{MA} number of microaneurysms, N_{HM} number of hemorrhages

Table 5 Annotations for different classification stages

Classification stage 1		Classification stage 2		Classification stage 3	
Grades	Annotations	Grades	Annotations	Grades	Annotations
Grade 0	0	Grade 0	0	Grade 0	0
Grade 1, 2, 3	1	Grade 1 and 2	1	Grade 1	1
		Grade 3	2	Grade 2	2
				Grade 3	3

2.2 Performance metrics

The performance of classification is analysed using various performance metrics which are defined in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) which are specified in the following equations [from Eq. (1) to Eq. (6)].

Positive prediction value (PPV) is the measure of statistical variability in data which computes the deviation of DR affected class from its true value. It is expressed by Eq. (1).

$$Positivepredictionvalue(PPV) = \frac{TP}{TP + FP} \tag{1}$$

Sensitivity (Sen.) is defined as the ratio of TP to the sum of TP and FN. It provides the measure of the ability of the classification method to correctly identify the true DR affected class. Its range is between 0 to 100% and more its value tending towards 100% more is the ability to correctly detect the DR affected pixels. Sensitivity is expressed by Eq. (2).

$$Sensitivity(Sen.) = \frac{TP}{TP + FN} \tag{2}$$

Specificity (Sp.) provides the measure of the ability of the classification method to correctly identify the non-DR class. It is defined as the ratio of TN to the sum of TN and FP. Its range lies in between 0 to 100% and more its value tending towards 100% more is the ability to correctly detect the non-DR pixels. Specificity is expressed by Eq. (3).

$$Specificity(Sp.) = \frac{TN}{TN + FP} \tag{3}$$

Accuracy (Acc.) is termed as the proportion of the total number of truly identified non-DR and true DR affected fundus images out of the total number of fundus images considered for classification. Accuracy is expressed by Eq. (4).

$$Accuracy(Acc.) = \frac{TP + TN}{TP + FN + TN + FP} \tag{4}$$

The performance metric area under the curve (AUC) indicates the overall quality of classification performance.

The two parameters evaluated to validate the correctness of the OD segmentation approach are optical disc overlap and dice metric. Optical disc overlap is defined as the ratio of true positives (TP) to the sum of true positive (TP), false negative (FN) and false positive (FP) given by the Eq. (5).

$$Opticaldiscoverlap = \frac{TP}{TP + FN + FP} \tag{5}$$

TP represents the area that is correctly overlapped between the manual ground truth and automatically obtained OD segmentation. FN represents those pixels that are classified only in the manual ground truth and FP denotes the pixels which are only classified by the automated segmentation approach. Another parameter used for performance evaluation of the OD segmentation approach is dice metric which is expressed in Eq. (6). Dice metric is the measure for comparing the similarity between two segmentations and it is defined as the ratio of $2 \times$TP to the sum of TP, FN, FP and TN.

$$Dicemetric = \frac{2 \times TP}{TP + FN + FP + TN} \tag{6}$$

2.3 Methodology

The proposed system is implemented using MATLAB2018b environment on a computer system equipped with Intel Core i5 processor, 3 GHz and 8 GB RAM. The authors have implemented the command based SVM and inbuilt kNN classifier in this research work. A novel hierarchical severity grading (HSG) classification method comprising three hierarchal stages is proposed in this work that is capable of generating DR severity grade of every fundus image and its framework is depicted in Fig. 1.

Figure 1 illustrates the complete framework of the proposed HSG system for accurate DR severity grading. The proposed system comprises of three main stages: Retinal landmark segmentation, DR lesion discrimination and DR severity grade classification. Retinal landmark segmentation facilitates in the identification of non-DR and DR affected fundus images. The second stage of the proposed method enables the discernment of DR lesions (red and yellow lesion) employing lesion discrimination strategy. The third stage classifies the fundus images into different grades combining the guidelines provided by International Clinical Diabetic Retinopathy Disease Severity grading. The complete algorithm of the proposed HSG system is illustrated in Fig. 2, however, the block by block description of the proposed methodology is detailed in the following subsections.

2.3.1 Retinal Landmark Segmentation

The preliminary steps involved in automated detection of DR are image pre-processing and masking of retinal landmark background portions including optical disc (OD) and vasculatures. Fundus image pre-processing involves variation attenuation, green channel conversion, image denoising and contrast enhancement employing the contrast limited adaptive histogram equalization approach (CLAHE).

In this paper, Grayworld normalization is used for image variation attenuation (Bhardwaj et al. 2018a, b). RGB values of the original image are initially scaled and the change in illumination is mitigated by averaging the scaled RGB values by applying three constant scaling factors α, β and γ

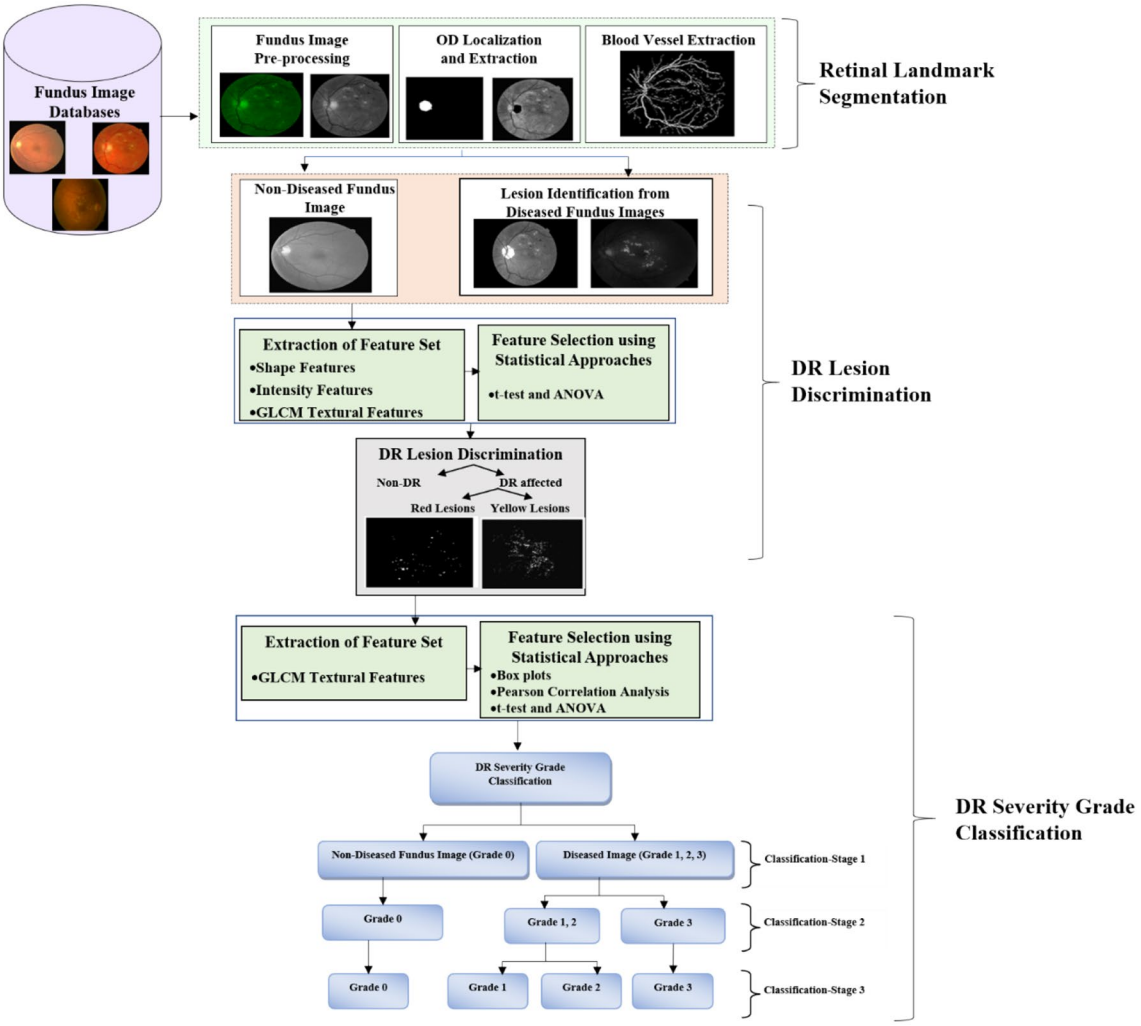

Fig. 1 Framework of the proposed HSG system

in R, G and B color channels respectively. The equation for grayworld normalization is expressed in Eq. (7).

$$(\alpha R, \beta G, \gamma B) = \left(\frac{\alpha R}{\frac{\alpha}{n}\sum_i R}, \frac{\beta G}{\frac{\beta}{n}\sum_i G}, \frac{\gamma B}{\frac{\gamma}{n}\sum_i B} \right) \tag{7}$$

This normalization step is used to eliminate the image variations and fit all the fundus images to the same scale.

Further, green channel conversion is done and this converted image is used for information extraction as this channel has maximum contrast out of all the three RGB channels. Filtering is done to suppress the isolated noise in the fundus image without blurring the edges. CLAHE is used for contrast enhancement and the contrast of each region is enhanced in such a way that the histogram of the output region approximately matches the specific histogram distribution (Niemeijer et al. 2004; Bhardwaj et al. 2018a, b). Let I_{den} be the denoised fundus image represented by a matrix of integer pixel intensities ranging from a dynamic range of 0 to $L-1$, L is the maximum value of intensity range (256). The normalized histogram (h_n) of the denoised image is denoted by Eq. (8),

$$h_n = \frac{number\ of\ pixels\ with\ intensity\ n}{Total\ number\ of\ pixels} \tag{8}$$

for n = 0,1,…, L − 1.

Thus, the histogram equalized image ($f_{i,j}$) is expressed by Eq. (9),

$$f_{i,j} = floor((L-1)\sum_{n=0}^{f_{i,j}} h_n) \tag{9}$$

where floor function rounds off to the nearest integer value.

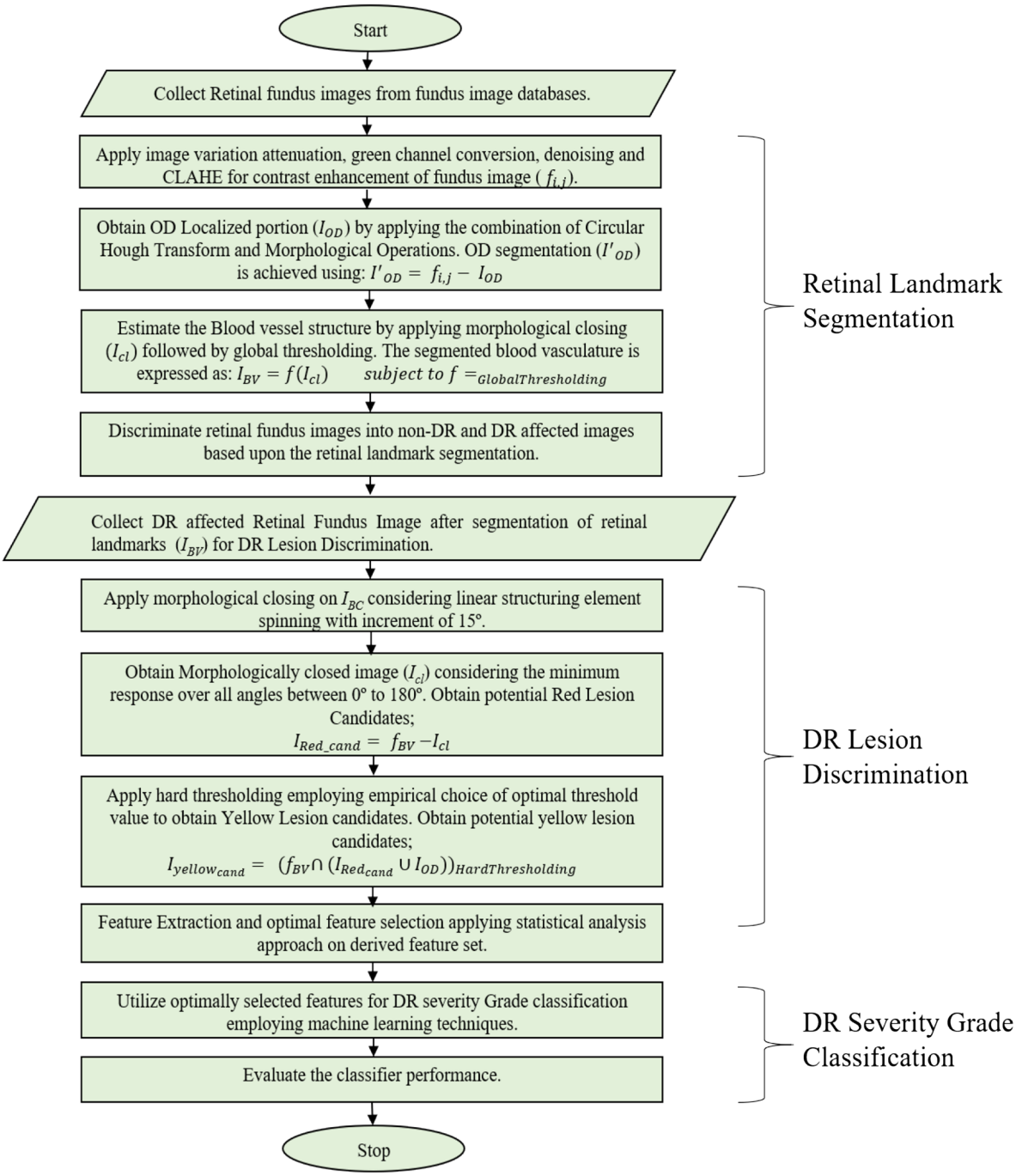

Fig. 2 Algorithm for the proposed HSG system

For the segmentation of retinal landmarks, optical disc (I_{OD}) and blood vasculatures (I_{BV}) are segmented and removed from the histogram equalized image.

2.3.1.1 OD localization OD localization is achieved employing the HSI color model considering only the intensity values (Bhardwaj et al. 2019). Circular Hough transform is used to locate the circular regions in the fundus image by considering the radius range of 70–140 pixels. The morphological closing operation is performed for the exact boundary localization of the OD portion and the largest circular region in the intensity plane is indicated as the optical disc (I_{OD}). OD segmentation (I'_{OD}) is achieved using Eq. (10).

$$I'_{OD} = f_{i,j} - I_{OD} \tag{10}$$

The block diagram of the OD segmentation approach used in the proposed HSG system is depicted in Fig. 3.

Table 6 depicts the comparison of the OD segmentation approach with another approach i.e. region growing OD segmentation method (Bhardwaj et al. 2018b). This method initializes the seed point and thereby growing the region by appending the similar property neighborhood pixels to the seed. In this approach, the optical disc centre is considered as the seed point and the absolute difference between the seed and the pixels is used as the stopping criteria for region growing.

The comparative analysis of both the approaches validates the segmentation efficiency of the proposed approach over the other in terms of optical disc overlap and dice metric parameters. Therefore, the viability of the proposed OD segmentation method is demonstrated from this performance validation.

Table 6 Performance evaluation of OD Segmentation approach

Approaches	Region growing OD segmentation approach (Bhardwaj et al. 2018b)	Proposed morphology based OD segmentation approach
Parameters		
Optical disc overlap	0.9565	0.9872
Dice metric	0.94.89	0.9931

2.3.1.2 Blood vessel segmentation Blood vessel segmentation in this work utilizes a two-fold approach in which morphological closing followed by global thresholding for vessel structure estimation. The maximum contrast green channel is utilized for blood vasculature segmentation and morphological closing operation is applied on the pre-processed green channelled fundus image. The morphologically closed image (I_{cl}) is subjected to vessel structure estimation by setting a global threshold by taking the absolute difference of the original image and the estimated retinal background.

The segmented blood vasculature is expressed in Eq. (11).

$$I_{BV} = f(I_{cl}) subject\ to\ f = _{Global\ Thresholding} \tag{11}$$

Post-processing steps are applied on estimated vessel structure for local contrast enhancement of the blood vasculatures and to remove the false detected isolated regions. The fundus image is labelled into different components (vessel pixel or non-vessel pixels) depending upon the pixel connectivity among four adjacent and four diagonal neighbors (8-connectivity) to obtain the true vessel pixels (Giancardo

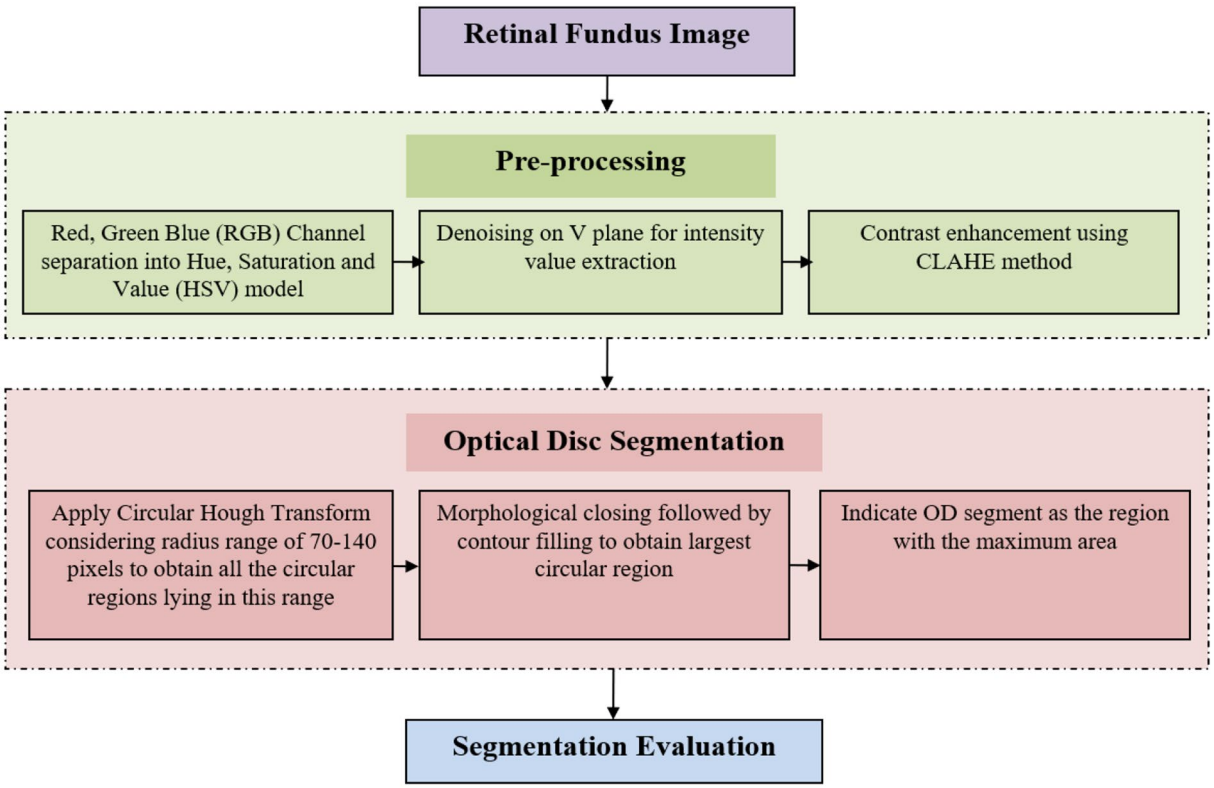

Fig. 3 Block diagram of OD segmentation approach

et al. 2012). If these background regions are not masked at the initial stage, the automated system may detect exudates as false OD region and blood vessels may falsely be detected as red lesions (microaneurysms). Candidate regions corresponding to lesion candidates are identified as foreground regions.

2.3.2 DR lesion discrimination

2.3.2.1 Pathology identification and detection In the pathology identification process, the foreground candidates are categorized into true lesions (red or yellow) and non-lesions. For lesion identification and detection, morphological operations are used to distinguish red and yellow lesions depending upon their intensity values. Red lesions correspond to dark intensity regions whereas yellow lesions correspond to brighter intensity regions.

Red lesion candidates are detected by subtracting the morphologically closed image from pre-processed histogram equalized image as expressed in Eq. (12).

$$I_{Red\_cand} = f_{BV} - I_{cl} \tag{12}$$

For yellow lesion detection, all the background regions and the identified red lesion candidate regions are considered as background and they are subtracted from the pre-processed image. Yellow lesions are detected using a hard thresholding method by selecting a threshold value considering histogram properties of the lesion. The expression for yellow lesion candidate detection is given in Eq. (13).

$$I_{yellow_{cand}} = (f_{BV} \cap (I_{Red_{cand}} \cup I_{OD}))_{Hard\ Thresholding} \tag{13}$$

Red lesions have the dark intensity and therefore are found on the extreme left of the histogram; however, yellow lesions have the bright intensity and hence found on the right side of the histogram (Bhardwaj et al. 2020). The automatically classified regions using morphological operations are compared to manually annotated ground-truth regions to obtain the true lesions and discard non-lesion candidates as it discards some false candidates which are neither red nor yellow lesion candidates. False red lesion candidates correspond to some crossing points or branches in the blood vessels, black spots or scars due to previous clinical records. Nerve fibre layer reflections or some other bright structures appear as false exudates candidates. Post-processing steps are used in the proposed approach which reduces the chances of false candidate detection up to a great extent. Hierarchical categorization of detected pathologies divides true red lesions into microaneurysms and hemorrhages whereas true yellow lesions into exudates and cotton wool spots.

2.3.2.2 Feature set description For the identification of DR lesions, a feature set comprising shape features, intensity features and textural features have been investigated. The geometrical properties of DR lesions are exploited using 11 shape features including lesion area, convex area, perimeter, major axis length, minor axis length, eccentricity of the lesion, lesion orientation, diameter, solidity, extent, and compactness (Seoud et al. 2016). The another set of features utilized for lesion discrimination comprises of pixel-based intensity features. This feature set consists of nine attributes including maximum and minimum pixel intensities, mean, median, standard deviation (SD), inter-quartile range (IQR), mean absolute difference (MAD), skewness and kurtosis (Harini and Sheela, 2016; Selvathi et al. 2012). The third set of features exploit the image textural properties of the detected lesions considering the pixel spatial relationships. Gray-level co-occurrence matrix (GLCM) textural features used in this work includes lesion autocorrelation, correlation of lesions, lesion contrast, lesion energy, lesion entropy, lesion homogeneity, lesion dissimilarity, cluster shade of lesions, cluster prominence of lesions, maximum probability, inverse difference normalized, inverse different moment normalized, information measure of correlation1 and information measure of correlation2 (Clausi 2002; Sood 2017; Acharya et al. 2009).

2.3.2.3 Statistical analysis Statistical analysis methods adopted in this paper using the SPSS package are based on box plots, Pearson correlation coefficient analysis, *t* test and ANOVA.

Box plots Feature vector can be graphically represented as a box or whisker diagram presenting more information in terms of the degree of dispersions, skewness in data and identifying the outliers. Box-plots are used for graphical normality analysis of feature set indicating the quartile ranges specifically pin-pointing the outliers. Five significant points in the data distribution are depicted by box plots: minimum, first quartile, median, third quartile and maximum data point.

Pearson correlation coefficient analysis Pearson correlation measures the linear correlation between two sets of samples and its value ranges from − 1 to 1. Pearson correlation coefficient value 1 shows a total positive linear correlation between the samples and 0 shows no correlation. The total negative correlation is indicated by the value of the Pearson correlation coefficient as − 1.

Levene's test Levene's inferential statistic test is used to assess the equality of variances for two or more than two groups. Its result depends upon the p value and if the significant p value is less than 0.05 then it is considered significant otherwise not.

t test and ANOVA t test is a parametric test helpful in evaluating the sets of continuous data by comparing their means. Analysis of variance (ANOVA) inferential statistical tool analyzes the difference or variation among the group means.

This is basically the generalization of the *t* test for more than two groups as it provides evidence of whether the means of several groups are equal or unequal. The significance of *t* test and ANOVA are depicted by the p value taken as less than 0.05 otherwise the test depicts insignificant value. Selected features based on statistical analysis done using the above-mentioned tools are provided for the classification stage.

2.3.3 DR severity grade classification

Red lesion candidates are grouped to obtain DR severity level grading per fundus image in the dataset. MESSIDOR dataset annotations are graded per image, where grade 0 indicates non-DR fundus image, grade 1 implies mild DR symptoms, moderate DR stage is indicated as grade 2 and grade 3 implies severe DR stage.

To classify the DR grades into different stages, various classifiers available in the literature were used to obtain higher accuracy (Habib et al. 2016; Lachure et al. 2015; Roychowdhury et al. 2013). We have used support vector machine (SVM) and k-nearest neighbour (kNN) classifiers for the classification task in this research work.

2.3.3.1 Support vector machine (SVM) classifier SVM discriminative supervised learning method separates the hyperplane using labelled training data and optimally categorizes the testing data. This approach is basically applicable for binary classification but its practical applications can be used to solve multi-class pattern recognition problems.

For a given set of training data of n points denoted by Eq. (14);

$$\left(\vec{i}_1, j_1\right), \ldots\ldots\ldots, (\vec{i}_n, j_n) \tag{14}$$

where output j_i may either be 1 or -1 denoting individual class for binary classification to which point $\vec{i}_i$ belongs. So the hyperplane can be represented as a set of points satisfying the Eq. (15).

$$\vec{w} \bullet \vec{i} - b = 0 \tag{15}$$

where $\vec{w}$ indicates weight vector, $\vec{i}$ denotes input vectors and b is the bias. The hyperplane representation for the multi-class classification problem is shown in Fig. 4.

For the multi-class DR classification problem, we have used "one-against-rest" approach in which different classifiers are prepared for each class. Let the number of classifiers trained for multiclass classification be 'n' and for a particular jth classifier to be trained, the whole fundus image dataset is used for training to classify the members of class J against the rest. Two tuning parameters regularization parameter (C) and gamma (γ) are varied to achieve better accuracy

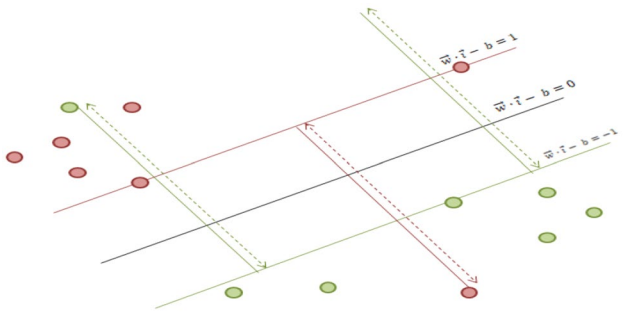

Fig. 4 Hyperplane representation for the multi-class classification problem

for non-linear classification. Regularization Parameter is responsible for avoiding misclassification of each training example and gamma determines the influence of each training example in the classification process. The error function for SVM classifier is expressed by Eq. (16).

$$C \sum_{i=1}^{L} \left(\varepsilon_i^+ + \varepsilon_i^-\right) + \frac{1}{2}\|w\|^2 \tag{16}$$

This error function is needed to be minimized depending upon regularization parameter C and error metric ε subject to minimization constraint, $\varepsilon^+ \geq 0$, $\varepsilon^- < 0$.

SVM classification relies on kernel selection as there are different SVM kernels like linear, polynomial, Gaussian radial basis function (RBF) kernel, etc. For the DR classification problem, Linear and Gaussian radial basis function kernels are considered as polynomial kernels and do not provide favourable outcomes in terms of accuracy. For an input vector x, support vector x_i, bias b and weight vector w, linear Kernel is given by Eq. (17).

$$K(x, x_i) = \sum \vec{w} \bullet \overrightarrow{x, x_i} + b \tag{17}$$

The polynomial kernel is denoted by Eq. (18):

$$K(x, x_i) = \left(1 + \sum \vec{w} \bullet \overrightarrow{x, x_i}\right)^d \tag{18}$$

where d indicates the degree of the polynomial.

Gaussian radial basis function kernel is expressed by Eq. (19) where γ denotes the Gaussian envelope width for a high-dimensional feature space.

$$K(x, x_i) = exp(-\gamma \times \sum \|\overrightarrow{x - x_i}\|^2) \tag{19}$$

2.3.3.2 k-nearest neighbour (kNN classifier) The other classifier used in this paper is the kNN classifier which classifies k- nearest neighbors depending upon the Euclidean distance between the dataset samples. The dataset contains a total of

'*n*' samples with '*f*' number of feature vectors where input sample with *f* features is denoted by Eq. (20).

$$x_i(x_{i1}, x_{i2}, x_{i3} \ldots \ldots \ldots \ldots, x_{if}) \tag{20}$$

The Euclidean distance between sample x_i and x_j (for $j = 1, 2, 3, \ldots, n$) is expressed by Eq. (21).

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{if} - x_{jf})^2} \tag{21}$$

Depending upon the Euclidean distance between samples, k-nearest training data points are considered for multiclass classification and assign the most frequently occurring class to the test data (Lachure et al. 2015). For DR classification problem, fine and weighted kNN algorithms are used (Venkatesan et al. 2012; Rahim et al. 2016). Fine kNN classifies the nearest neighbor by making a finely detailed distinction between the classes whereas weighted kNN uses distance weighting to determine the members of a particular class.

3 Results and discussion

The performance of the proposed HSG system is analysed in terms of visual, statistical and performance metrics. The results of the background and foreground segmentation, pathology identification and detection are depicted in Fig. 5. OD region and blood vessels are considered as the background portions whereas lesions from the foreground segmentation are extracted as pathological regions.

The separation of background and foreground portions from the original fundus image is necessary for lesion identification. The field of view part is separated by masking and the optical disc portion is located and removed as this portion of the retina is undesirable for DR pathology diagnosis. Blood vessels have a similar intensity as those of microaneurysms and hemorrhage and regarded as background regions and thus removed. Further from the foreground portion, DR pathologies are identified as lesion candidates. As per the grading guidelines provided in the MESSIDOR dataset,

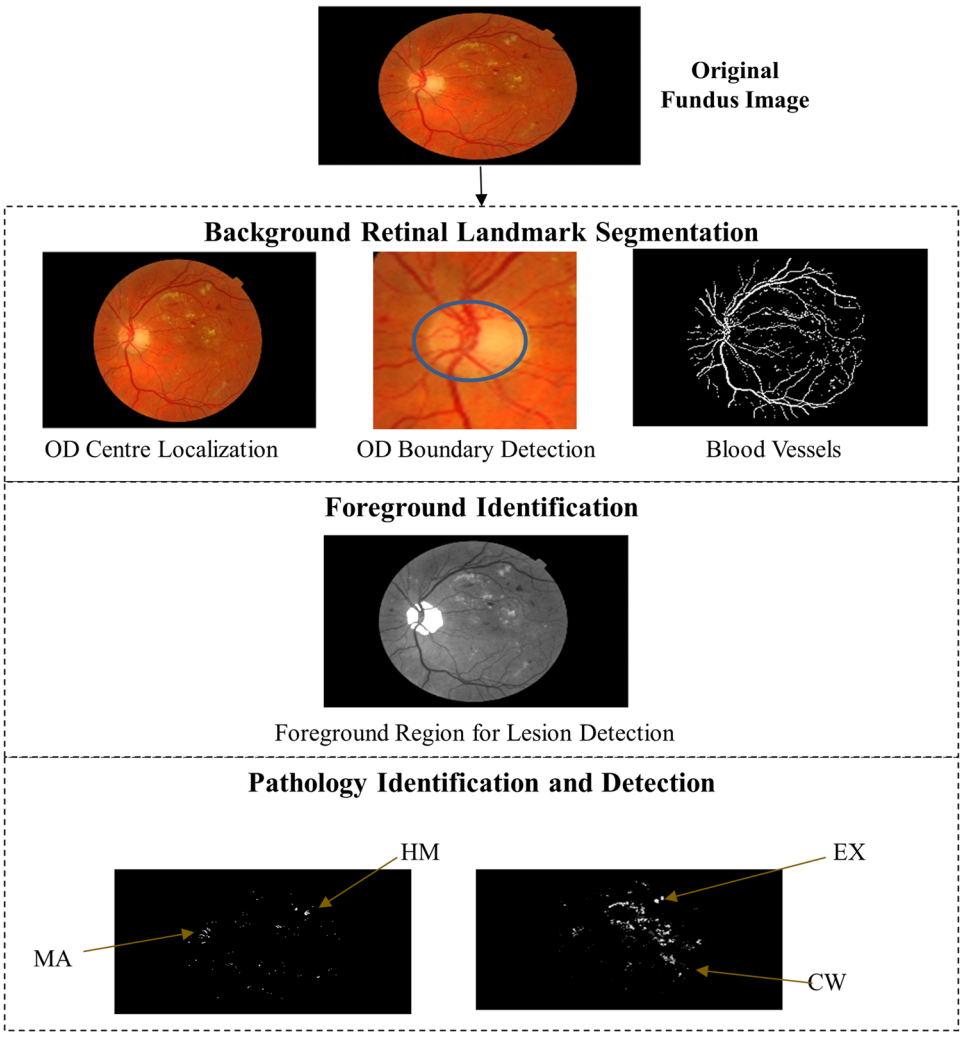

Fig. 5 Background retinal landmark segmentation, foreground identification and pathology detection

MA and HM lesion candidates are combined after pathology identification to classify images based on DR severity grading. GLCM features are extracted for lesion candidates in grade 1, grade 2, grade 3 and non-lesion candidates for grade 0 fundus images. Extracted features are analysed statistically using various statistical tools.

3.1 Statistical analysis of shape and intensity feature set

The shape and intensity-based features are of significant importance for lesion discrimination and statistical analysis of these extracted features is done to obtain most prominent features reducing the dimensional complexity. The feature set analysis is done using the significance value (p-value) investigation and the observations are depicted in Table 7.

Table 7 reveals the statistical significance of shape and intensity features for lesion discrimination, in terms of significant p-values. All the intensity features and shape features except eccentricity and orientation provides significant p-value indicating their statistical relevance to lesion discrimination strategy.

The feature set investigating and literature suggests that GLCM based textural features provides the promising outcomes for DR severity grading problem and therefore they are considered for further experimentations involved in this research.

3.2 Statistical analysis of GLCM feature set

The statistical analysis of extracted GLCM feature set for non-DR and DR affected fundus images is done using various SPSS statistical tools. Statistical feature set analysis using box plots reveals the normality visualization of GLCM features indicating the data distribution inside a rectangular box and pinpointing the outliers. Box plots of different GLCM features are shown in Fig. 6 which indicates 25% to 75% of the data distribution inside a rectangular box. Individual points outside the rectangular box are specified as outliers.

The relationship between GLCM feature variables is quantified using a pair-wise correlation test as provided by Pearson correlation coefficients tabulated in Table 8.

The Pearson correlation coefficient values shown in Table 8 reveal that GLCM features except *cluster shade* and *energy* are prominent but further analysis using other statistical techniques is required to check the usefulness of less related features for DR classification. Further investigation is done in terms of descriptive statistical analysis using Levene's test, t test and ANOVA for optimal feature selection. The statistical significance of GLCM features is determined by Levene's test for equality of variance (σ) and t test for equality of means (μ). Levene's test is performed on the null hypothesis (H_0) considering that there is no significant difference between the variance of GLCM features extracted from different DR grade classes as expressed by Eq. (22). Null hypothesis rejection H_1 given by Eq. (23), states that there is a significant difference between the variance of GLCM features of non-DR and DR affected class.

$$H_o : \sigma_0 = \sigma_1 \tag{22}$$

$$H_1 : \sigma_0 \neq \sigma_1 \tag{23}$$

The null hypothesis ($H_0^{'}$) for the t test states there is no significant difference between the mean values of GLCM features for grade 0 and grade 1–3 which is expressed as Eq. (24). Rejection of the null hypothesis is indicated as $H_1^{'}$ which states that there is a substantial difference between the mean values of GLCM features for non-DR and different DR severities and it is given in Eq. (25).

Table 7 Statistical analysis of extracted shape and intensity features

Shape features			Intensity features		
Features	ANOVA (p value)	t test (p value)	Features	ANOVA (p value)	t test (p value)
Lesion area	0.019	0.019	Minimum intensity	0.000	0.003
Lesion perimeter	0.033	0.040	Maximum intensity	0.000	0.000
Major_axis_length	0.003	0.033	Mean intensity	0.000	0.000
Minor_axis_length	0.005	0.011	Median intensity	0.000	0.000
Lesion eccentricity	0.062	0.062	MAD intensity	0.000	0.000
Convex_area	0.002	0.005	SD intensity	0.000	0.000
Lesion orientation	0.102	0.102	IQR	0.000	0.000
Lesion equiv_dia	0.002	0.002	Lesion skewness	0.015	0.012
Lesion solidity	0.007	0.015	Lesion kurtosis	0.005	0.005
Lesion extent	0.030	0.033			
Lesion compactness	0.000	0.002			

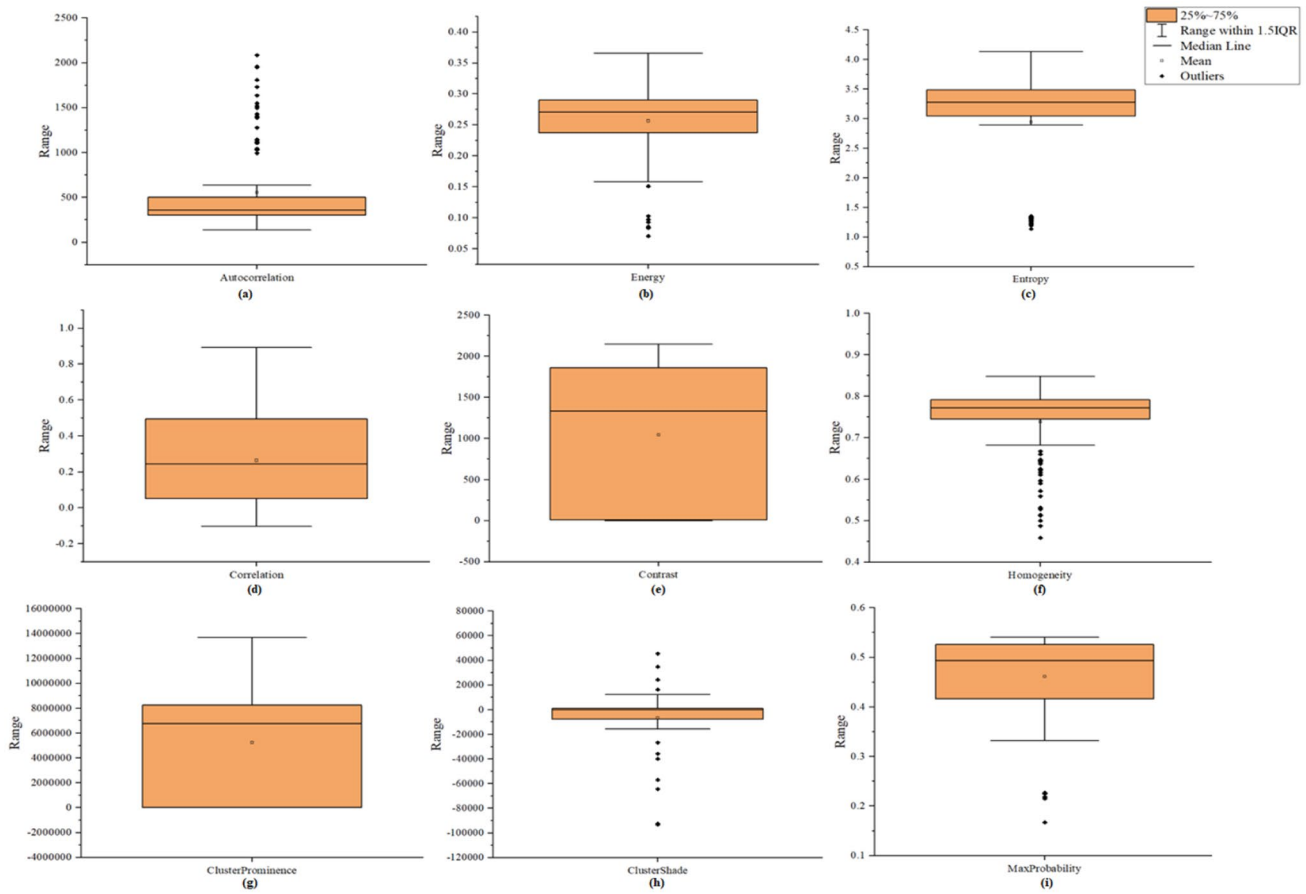

Fig. 6 Box plots of some of the extracted GLCM feature set

$$H_0^{'} : \mu_0 = \mu_1 \tag{24}$$

$$H_1^{'} : \mu_0 \neq \mu_1 \tag{25}$$

The null hypothesis is accepted if the significance values for the feature set are greater than 0.05. The results for Levene's test and t test are summarized in Table 9 for the GLCM feature set for grade 0 and grade 1–3 distinguishing between non-DR and different DR severities.

The tabular representation of Levene's test and *t* test for GLCM features reveals that all the GLCM features are significant having a p value less than 0.05 except *cluster shade* and *energy*. *Cluster shade* and *energy* features show insignificance as the p value deviates from 0.05 significant level. Therefore, the null hypothesis is rejected for GLCM features except for *cluster shade* and *energy* indicating that there is a significant difference between the means and variances of both the DR abnormality grades in terms of GLCM features (H_1 and H'_1).

t test reveals the descriptive statistics of the features but for more statistical evidence in terms of inferential statistics and brings out the inference about the DR phenomenon concerning the selected sample, one-way analysis of variance (ANOVA) is used. The significant values obtained for the one-way ANOVA test applied on GLCM features are tabulated in Table 10.

Table 10 shows the significance values of features using the ANOVA test and the GLCM textural features except *cluster shade* and *energy* are proven prominent as their p value is less than 0.05. *Cluster shade* and *energy* features deviate from significant p value thus showing statistical insignificance.

From the statistical and inferential analysis using Levene's test, *t* test and ANOVA, it is revealed that among the feature set consisting of 14 GLCM features, two features are insignificant for our research problem and the rest of the features are prominent for DR severity grade classification. Further all the computations are done using these prominent 12 GLCM features.

Table 8 Pearson correlation coefficients of GLCM feature set

GLCM features	Pearson correlation coefficients													
	Auto Corr	Corr	Contrast	Cluster-Shade	Cluster Prom	Energy	Entropy	Homo	Inv.Diff. Norm	Inv.Diff. Mom	Diss	Max Prob	Inf. Corr. 1	Inf. Corr. 2
Auto corr	1.000	−0.865**	0.860**	−0.745**	0.929	0.338**	−0.869**	−0.777**	−0.857	−0.860**	0.859**	−0.492**	0.900	−0.900**
Corr	−0.865**	1.000	−0.999**	0.454**	−0.933**	−0.403	0.958**	0.957**	0.999**	0.999	−0.999**	0.532**	−0.972**	0.986
Contrast	0.860**	−0.999**	1.000	−0.442**	0.934**	0.394**	−0.957	−0.960**	−1.000**	−1.000**	1.000	−0.542**	0.973**	−0.987**
Cluster-Shade	−0.745**	0.454**	−0.442**	1.000	−0.512**	−0.260**	0.518**	0.325	0.435**	0.441**	−0.438**	0.180	−0.490**	0.506**
Cluster-Prom	0.929**	−0.933**	0.934**	−0.512**	1.000**	0.385**	−0.940**	−0.849**	−0.931**	−0.934**	0.933**	−0.517**	0.974**	−0.970**
Energy	0.338**	−0.403**	0.394**	−0.260**	0.385**	1.000**	−0.572**	−0.265**	−0.388**	−0.393**	0.391**	0.534**	0.393**	−0.414**
Entropy	−0.869**	0.958**	−0.957**	0.518**	−0.940**	−0.572**	1.000**	0.843**	0.951**	0.956**	−0.954**	0.364**	−0.946**	0.970**
Homo	−0.777**	0.957**	−0.960**	0.325**	−0.849**	−0.265**	0.843**	1.000**	0.965**	0.960**	−0.962**	0.594**	−0.924**	0.922**
Inv.Diff. Norm	−0.857**	0.999**	−1.000**	0.435**	−0.931**	−0.388**	0.951**	0.965**	1.000**	1.000**	−1.000**	0.545**	−0.972**	0.985**
InvDiff. moment	−0.860**	0.999**	−1.000**	0.441**	−0.934**	−−0.393**	0.956**	0.960**	1.000**	1.000**	−1.000**	0.543**	−0.973**	0.987**
Diss	0.859**	−0.999**	1.000**	−0.438**	0.933**	0.391**	−0.954**	−0.962**	−1.000**	−1.000**	1.000**	−0.544**	0.973**	−0.986**
MaxProb	0.900	−0.900**	−0.972**	0.986	−0.517	0.534	0.364	0.594	0.545	0.543	−0.544	1.000	−0.520	0.520
Inf. Corr.1	0.973**	−0.987**	−0.490**	0.506**	0.974	0.393	−0.946	−0.924	−0.972	−0.973	0.973	−0.520	1.000	−0.992
Info. Corr.2	0.974**	−0.970**	0.393**	−0.414**	−0.970	−0.414	0.970	0.922	0.985	0.987	−0.986	0.520	−0.992	1.000

**Correlation is significant at the 0.01 level

Table 9 Levene's test and t test for GLCM Features

Features	Levene's test		*t* test	
	F	Sig	t	Sig. (2-tailed)
Auto correlation	55.124	0.000	−4.826	0.000
Correlation	108.862	0.000	3.443	0.001
Contrast	109.651	0.000	−3.480	0.001
ClusterShade	13.601	0.000	0.744	0.458
ClusterProminance	97.118	0.000	−4.214	0.000
Energy	7.595	0.007	1.050	0.295
Entropy	73.125	0.000	2.414	0.017
Homogenity	40.663	0.000	4.290	0.000
InverseDiffNorm	104.124	0.000	3.560	0.000
InvDiffMoment	109.226	0.000	3.491	0.001
Dissimilarity	106.816	0.000	−3.523	0.001
MaxProb	34.019	0.000	4.078	0.000
InfoCorr1	91.816	0.000	−3.880	0.000
InfoCorr2	121.068	0.000	3.479	0.001

Table 10 ANOVA results for GLCM feature set

GLCM features	
Features	One way ANOVA (significant value)
Auto-correlation	0.001
Correlation	0.000
Contrast	0.001
ClusterShade	0.458
ClusterProminance	0.000
Energy	0.295
Entropy	0.017
Homogeneity	0.001
InverseDiffNorm	0.000
InvDiffMoment	0.001
Dissimilarity	0.001
MaxProb	0.001
InfoCorr1	0.000
InfoCorr2	0.001

3.3 Hierarchal severity grade (HSG) classification

The reduced feature set comprising 12 GLCM features are used for various classifiers for DR severity grading. After exhaustive literature review and empirical experimentation, SVM classifier with linear and Gaussian RBF kernels and kNN classifiers (fine and weighted) are reported in this research work. Classification results of the HSG system for DR demonstrate the feasibility of the proposed system for stage 1, stage 2 and stage 3 classification of abnormalities. The entire set of images is divided randomly into training and validation set for testing and validation using 70–30% criteria, considering 840 images for training and 360 for testing, out of total 1200 images from the MESSIDOR dataset. The reported results are the average of the tenfold cross-validations which divides the training set into 10 distinct classes to obtain better training of the classification task. The composition of sample images taken with their explicit number is shown in Table 11.

Different cases considered for DR severity grading and classification detailed in the following section are tabulated in Table 12.

3.3.1 Classification-stage 1 (non-DR and DR affected fundus images)

Classification stage 1 classifies the fundus images into non-DR and DR affected class, Grade 0 and Grade 1, 2, 3 (collectively) respectively. The results of this classification stage using SVM and kNN classifiers with different kernels are shown in Table 13.

For non-DR and DR affected fundus images classification, the features determining the DR severity levels are graded correctly providing accuracy of 98.10% employing SVM classifier and 100% for kNN classifier. SVM classifier using linear kernel provides an average sensitivity, specificity and positive prediction rate of 100%, 95.23% and 97.06% respectively. This is a linear problem, thus better results are observed for linear SVM kernel comparative to Gaussian RBF SVM kernel.

3.3.2 Classification-stage 2

Grade 1 and grade 2 severity levels comprise mild and moderate DR abnormality symptoms which are graded into one category and the grade 3 category includes severe DR abnormality symptoms. Grades 1 and 2 are less severe and have similar lesion properties and are therefore considered collectively for classification-stage 2. The classification results are tabulated in Table 14.

For the DR classification-stage 2, the SVM classifier provides 91.50% overall accuracy using the linear kernel and 95.68% using the Gaussian radial basis function kernel. Positive prediction value, sensitivity and specificity of 90%, 93.77% and 95.56% respectively is achieved by the linear SVM classifier. SVM classifier employing Gaussian RBF

Table 11 Training and Validation set distribution for DR severity classification

Grade	Grade 0	Grade 1	Grade 2	Grade 3
Image set				
Training set	240	200	200	200
Testing/validation set	160	60	60	60

Table 12 Cases for HSG classification

		Non-DR	DR affected		
Case 1	Classification-Stage 1	Grade 0		Grade 1, 2, 3	
Case 2	Classification-Stage 2	Grade 0	Grade 1 and 2		Grade 3
Case 3	Classification-Stage 3	Grade 0	Grade 1	Grade 2	Grade 3

Table 13 Performance measure of HSG system for classification-stage 1

Features	Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	AUC
GLCM Features	Linear SVM	100	95.23	98.10	97.06	1.00
	Gaussian RBF SVM	93.93	90.47	92.60	93.93	0.99
	Fine kNN	100	100	100	100	1.00
	Weighted kNN	100	100	100	100	1.00

PPV positive prediction value, *AUC* area under the curve

Table 14 Performance metrics of HSG system for classification-stage 2

Classifier	SVM		kNN	
	Linear kernel	Gaussian RBF kernel	Fine kNN	Weighted kNN
Parameter				
PPV				
PPV_0	87.50%	100%	100%	100%
PPV_1	86.67%	100%	100%	87.50%
PPV_2	95.83%	92.86%	96.29%	100%
Sen				
Sen_0	100%	100%	100%	100%
Sen_1	92.86%	85.71%	92.86%	100%
Sen_2	88.46%	100%	100%	92.31%
Sp				
Sp_0	97.50%	100%	100%	100%
Sp_1	93.93%	100%	97.06%	93.94%
Sp_2	95.24%	90.47%	100%	100%
Acc				
Acc_0	97.87%	100%	100%	100%
Acc_1	93.62%	95.74%	97.87%	95.92%
Acc_2	91.49%	95.74%	97.87%	95.74%
Overall accuracy	91.50%	95.68%	97.90%	95.70%
AUC	1.00	1.00	1.00	1.00

PPV positive prediction value, *Sen.* sensitivity, *Sp.* specificity, *Acc.* accuracy, *AUC* area under the curve

kernel provides 97.62%, 95.24% and 96.82% of PPV, sensitivity and specificity values respectively. Gaussian RBF kernel employed for SVM classification provides better results in terms of all the performance indices as classification-stage 3 is a non-linear problem. kNN classifier provides an overall accuracy of 97.90% for fine kNN and 95.70% weighted kNN classifier. Fine kNN classifier achieves positive prediction value, sensitivity and specificity of 98.76%, 97.62%, 99.02% respectively and weighted kNN classifier provide 95.83%, 97.44%, 97.98% of PPV, sensitivity and specificity values respectively. Further classification of abnormality grades reduces the grading efficiency as the features of different abnormality classes are somewhat correlated. Gaussian RBF kernel for SVM classifier and fine kNN classifier provides better performance in terms of different indices due to the robustness of these classifiers for non-linear classification problems.

3.3.3 Classification-stage 3

Further, to classify DR into distinct abnormalities, Grade 0, 1, 2 and 3 are separated into different classes in classification-stage 3. The results for classification-stage 3 in terms of PPV, Sensitivity, specificity and accuracy are provided in Table 15.

SVM classifier employing linear and Gaussian kernel provides overall accuracy of 91.30% and 92.60% respectively for the DR classification-stage 3. The positive prediction value, sensitivity and specificity of 94.27%, 91.58% and 97.24% respectively are achieved by the linear SVM classifier. SVM classifier employing Gaussian RBF kernel provides 92.62%, 92.65% and 97.36% of PPV, sensitivity and specificity values respectively. Gaussian RBF kernel for SVM classifier performs better than linear kernel due to its competence for non-linearity. kNN classifier provides 95.30% accuracy for fine kNN and 87.90% accuracy is observed for the weighted kNN classification method. Fine kNN classifier achieves positive prediction value, sensitivity and specificity of 96.35%, 94.41%, 98.25% respectively and weighted kNN classifier provide 90.12%, 90.38%, 95.45% of PPV, sensitivity and specificity values respectively. The highest accurate results providing 95.30% accuracy are observed for fine kNN classifier as unlike SVM it does not need higher training examples to distinguish between the classes and decides its class labels depending

Table 15 Performance metrics of HSG system for classification-stage 3

Classifier	SVM		kNN	
	linear kernel	Gaussian RBF kernel	Fine kNN	Weighted kNN
Parameters				
PPV				
PPV_0	96.36%	98.07%	100%	88.89%
PPV_1	85.71%	87.28%	91.67%	81.13%
PPV_2	95.00%	95.65%	100%	95.45%
PPV_3	100%	89.47%	93.75%	95.00%
Sen				
Sen_0	92.98%	89.47%	91.43%	84.21%
Sen_1	96.00%	96.00%	100%	86.00%
Sen_2	82.61%	95.65%	86.21%	91.30%
Sen_3	94.74%	89.47%	100%	100%
Sp				
Sp_0	97.83%	98.91%	100%	93.48%
Sp_1	91.92%	92.93%	94.68%	89.89%
Sp_2	99.20%	97.62%	100%	99.21%
Sp_3	100%	100%	98.32%	99.23%
Acc				
Acc_0	94.91%	95.30%	97.98%	92.39%
Acc_1	93.20%	93.96%	96.64%	88.59%
Acc_2	98.31%	97.31%	97.32%	97.99%
Acc_3	97.33%	98.66%	98.66%	99.33%
Overall accuracy	91.30%	92.60%	95.30%	87.90%
AUC	0.99	0.99	0.99	0.89

PPV positive prediction value, *Sen.* sensitivity, *Sp.* specificity, *Acc.* accuracy, *AUC* area under the curve

upon its immediate neighbors making it best suitable for DR classification problem.

Grade 1 DR severity images have similar features as that of Grade 0 and therefore can be falsely detected as non-DR fundus images. In literature, fewer attempts have been made for automated grading of DR severity and desirable accuracy is not achieved by the methods provided in the literature (Giancardo et al. 2012).

3.4 Validation of the proposed system using Indian Diabetic Retinopathy Image Dataset (IDRiD) dataset

The generalization ability of the proposed HSG system is validated using the IDRiD dataset to increase much more patient samples helping the physicians to realize the progress of the disease. To get an oversight of different classifier performance, we intend to check the performance metrics utilizing KNN and SVM classifiers for classification-stage 3. The proposed methodology is applied on the IDRiD dataset and the performance metrics analysis of observed for HSG approach comparing SVM and kNN classifiers is given in Table 16.

The performance metrics evaluated in Table 16 reveals the feasibility of the proposed HSG system for the IDRiD dataset providing a maximum of 94.00% accuracy for fine kNN classifier with 0.99 AUC. The proposed system provides 90.60% accuracy for the linear SVM classifier, 91.90% accuracy for the Gaussian SVM classifier and 93.31% accuracy for the weighted kNN classification method. In this work, kNN classifier have made the preeminent attempts to address this problem and generalize the grading ability of the proposed HSG system achieving better accuracies.

The two datasets involved in the experimental implementation of the proposed HSG approach are compared in terms of accuracy and computational time in Fig. 7.

It is observed from Fig. 7 that higher accuracy values with comparably less computational time are achieved using the HSG system with kNN classifier approach. It is revealed that the kNN classifier takes 0.73 s to train which is less compared to the SVM classifier, taking 1.13 s exploiting the MESSIDOR dataset. However, training of IDRiD dataset takes 0.67 s to train kNN classifier and 1.10 s for SVM classifier training. The reduced computational time of kNN with increased system accuracy justifies the selection of kNN classifier for DR severity classification. The proposed HSG system provides uniform results, irrespective of the dataset being used, thus establishing its generalization capabilities.

A comparative analysis of the HSG system for separating non-DR and DR affected fundus images is done in Table 17 whereas the comparison of classification-stage 3 HSG DR classification with the existing approaches is made in Table 18.

The comparative study reveals that our system for separating non-DR and DR affected images provide 100% average accuracy using kNN classifier that is maximum among the existing methods using the same dataset. Our system employing fine kNN classifier provides an accuracy of 95.30% and AUC of 0.99 and it is compared to other state of the art approaches in Table 18. The comparison reveals that the proposed HSG method performs better in terms of accuracy and AUC among other existing approaches for classification-stage 3 of DR severity grading.

It was observed from Tables 17 and 18 that our proposed approach outperforms the state-of-the-art methods in terms of accuracy and AUC. The system proposed is less complex in terms of run-time providing cost-effective mass screening solutions for the detection of diabetic retinopathy. For the proposed HSG system, SVM and kNN classifiers are utilized to assess classification performance for DR severity grading. Fine kNN classifier

Table 16 Performance metrics of HSG system for classification-stage 3 (non- DR grade and DR affected grades 1–3) utilizing IDRiD dataset

Classifier	SVM		kNN	
	Linear kernel	Gaussian RBF kernel	Fine kNN	Weighted kNN
Parameters				
PPV				
PPV_0	96.36%	94.12%	97.06%	96.96%
PPV_1	84.21%	92.59%	93.10%	91.38%
PPV_2	94.12%	82.75%	88.46%	88.46%
PPV_3	90.00%	96.87%	96.77%	96.87%
Sen				
Sen_0	92.98%	91.42%	94.28%	91.42%
Sen_1	96.00%	90.91%	98.18%	96.36%
Sen_2	89.56%	92.31%	88.46%	88.46%
Sen_3	94.74%	93.93%	90.91%	93.93%
Sp				
Sp_0	97.83%	98.24%	99.12%	99.12%
Sp_1	90.91%	95.74%	95.74%	94.68%
Sp_2	99.21%	95.93%	97.56%	97.56%
Sp_3	98.46%	99.13%	99.13%	99.13%
Acc				
Acc_0	95.97%	96.64%	97.98%	97.31%
Acc_1	92.62%	93.96%	96.64%	95.30%
Acc_2	94.63%	95.30%	95.97%	95.97%
Acc_3	97.99%	97.98%	97.31%	97.98%
Overall accuracy	90.60%	91.90%	94.00%	93.31%
AUC	0.91	0.94	0.99	0.98

PPV positive prediction value, *Sen.* sensitivity, *Sp.* specificity, *Acc.* accuracy, *AUC* area under the curve

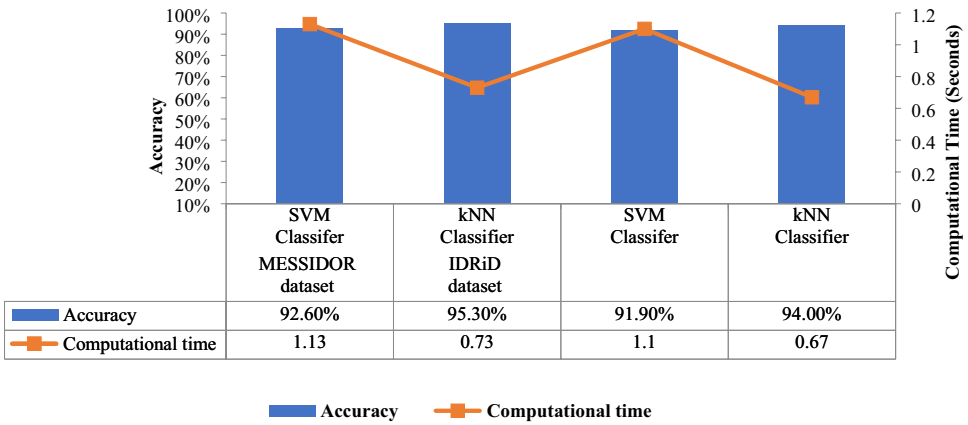

Fig.7 Comparison of MESSIDOR and IDRiD datasets in terms of accuracy and computational time

provides the maximum accuracy of 95.30% among all the classifier combinations used in this paper for abnormality grade classification. SVM classifier results into biased classification outcomes as it classifies each test sample belonging to the majority class (Niemeijer et al. 2004). kNN decides its class label depending upon its immediate neighbors, therefore, providing robust classification. However, a new technological advent in the field of medical sciences has developed with deep neural network (DNN) based screening methods. Abbas et al. (2017) have developed a deep visual feature-based DR severity classification method derived using deep learning-based multilayer semi-supervised technique. Wang and Yang (2018) yields better classification outcomes by introducing a deep learning technique employing Regression Activation Maps (RAMs) after pooling layer for severity level grading

Table 17 Comparison table for DR severity grading for separating non-DR and DR affected images using MESSIDOR database

Dataset used MESSIDOR	Accuracy (%)
Technique	
Lachure et al. (2015)	90.00
Rahim et al. (2016)	93.00
Bandyopadhyay et al. (2018)	95.30
Proposed HSG system (classification-stage 1)	100

Table 18 Comparative analysis of proposed HSG system (classification-stage 3) with the existing methods utilizing MESSIDOR database

Dataset used MESSIDOR	Accuracy (%)	AUC
Technique		
Roychowdhury et al. (2013)	78.12	0.90
Seoud et al. (2015)	74.10	0.73
Navarro et al. (2016)	84.00	–
Thammastitkul and Uyyanonvara (2016)	87.00	0.85
Xiao et al. (2019)	90.50	0.85
Proposed HSG system (Classification-stage 3)	95.30	0.99

based region of interest localization. Yu et al. (2017) uses the combination of CNNs and saliency maps to obtain both supervised as well as unsupervised features which act as input for SVM classifier. A deep CNN model was utilized by Gao et al. (2018), for DR severity grading utilizing their self built labelled dataset. The accuracy of 88.72% was achieved for severity level grading of DR fundus images using this particular model. These neural network-based methods have demonstrated revolutionary performance for various image classification and object recognition applications even surpassing the manual systems. Despite many advances that have been made for using DNNs in DR diagnosis, still these systems pose challenges in their practical implementation. The authors are working on the deep learning-based implementation of DR screening and a transfer learning based CNN module have been tested by the authors utilizing various pre-trained CNN models. Performance evaluation of CNN based classification model in terms of accuracy exploiting AlexNet, GoogleNet, ResNet, Vgg16, Vgg19 and InceptionV3 pre-trained models is depicted in Table 19.

From this tabular interpretation, it is revealed that our proposed kNN classifier based approach provides remarkable performance even when it is compared to the deep learning based approach. This is due to the better learning of machine learning classifier for small set of data and deep leaning methods require large amount of annotated data for proper convergence. However, while increasing the number of data samples for deep learning the network performance may improve. This is the future perspective of this research work.

Table 19 Performance evaluation of CNN based DR severity grade classification model

Pre-trained CNN models	Accuracy obtained for CNN based DR severity grade classification (%)
AlexNet	73.33
GoogleNet	65.56
ResNet	65.83
Vgg16	82.14
Vgg19	80.76
Inception V-3	87.50

The ultimate goal of our HSG system is to classify the fundus images with no retinopathy symptoms as non-DR and the fundus images with retinopathy lesions as abnormal irrespective of database, FOV and resolution.

4 Conclusion

This paper proposes a hierarchical severity level grading (HSG) system for the detection and classification of DR disease employing SVM and kNN classifiers. A total of 14 GLCM features were identified from pathologies obtained after foreground and background retinal landmark segmentation for the 400 fundus images. The statistical analysis resulted in 12 optimal features which are further fed to the respective classifiers. The variation in the classification was brought by selecting the different kernels for SVM as well as kNN classifiers. The proposed HSG system employs retinal fundus image categorization as grade 0 (indicating non-DR class) and DR severity grades 1, 2, 3 depending upon the number of DR anomalies present in the fundus images. The result for non-DR and DR affected fundus image classification provides 100% specificity, sensitivity, accuracy and positive prediction value. The classification-stage 3 of the proposed HSG approach utilizing kNN classifier achieved good classification results in terms of five statistical indices: accuracy, 95.30%; sensitivity, 94.41%; specificity, 98.25%; positive prediction value 96.35%; and AUC, 0.99. MESSIDOR dataset takes 1.13 s for SVM classifier training and computational time of 0.73 s for training kNN classifier. Performance validation of the proposed HSG system on the IDRiD dataset provides a maximum of 94.00% accuracy using kNN classifier providing the computational time of 0.67 s. Therefore, the selection of kNN classifier for classification of DR anomalies is justified in terms of accuracy as well as computation time. The proposed system can be utilized for the screening of DR abnormalities as it provides an

alternative solution for automated DR screening by showing statistically significant improvement in the results obtained. In the future, this work will be extended by utilizing deep learning and optimization algorithms for DR severity grade classification. Additionally, future work of this research will be focused on addressing neovascularization and blood vessel bleeding problems causing proliferative DR which may lead to retinal detachment resulting in acute blindness.

References

Abbas Q, Fondon I, Sarmiento A, Jiménez S, Alemany P (2017) Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features. Med Biol Eng Comput 55(11):1959–1974

Acharya UR, Lim CM, Ng EYK, Chee C, Tamura T (2009) Computer-based detection of diabetes retinopathy stages using digital fundus images. Proc Inst Mech Eng 223(5):545–553

Al-Jarrah MA, Shatnawi H (2017) Non-proliferative diabetic retinopathy symptoms detection and classification using neural network. J Med Eng Technol 41(6):498–505

Aptel F, Denis P, Rouberol F, Thivolet C (2008) Screening of diabetic retinopathy: effect of field number and mydriasis on sensitivity and specificity of digital fundus photography. Diabetes Metab 34(3):290–293

Ashraf MN, Habib Z, Hussain M (2014) Texture feature analysis of digital fundus images for early detection of diabetic retinopathy. In: Proceedings of the 11th IEEE international conference on computer graphics, imaging and visualization (CGIV '14). IEEE, pp 57–62

Bandyopadhyay S, Choudhury S, Latib SK, Kole DK, Giri C (2018) Gradation of diabetic retinopathy using KNN classifier by morphological segmentation of retinal vessels. In: International proceedings on advances in soft computing, intelligent systems and applications. Springer, Singapore, pp 189–198

Bhardwaj C, Jain S, Sood M (2018a) Appraisal of pre-processing techniques for automated detection of diabetic retinopathy. In: 2018 Fifth international conference on parallel, distributed and grid computing (PDGC). IEEE, pp 734–739

Bhardwaj C, Jain S, Sood M (2018b) Automated optical disc segmentation and blood vessel extraction for fundus images using ophthalmic image processing. In: International conference on advanced informatics for computing research. Springer, Singapore, pp 182–194

Bhardwaj C, Jain S, Sood M (2019) Automatic blood vessel extraction of fundus images employing fuzzy approach. Indones J Electr Eng Inform (IJEEI) 7(4):757–771

Bhardwaj C, Jain S, Sood M (2020) Diabetic retinopathy lesion discriminative diagnostic system for retinal fundus images. Adv Biomed Eng 9:71–82

Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens 28(1):45–62

Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, Charton B (2014) Feedback on a publicly distributed image database: the Messidor database. Image Anal Stereol 33(3):231–234

Dupas B, Walter T, Erginary A (2010) Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer assisted diagnostic system for grading diabetic retinopathy. Diabetes Metab 36(3):213–220

Gao Z, Li J, Guo J, Chen Y, Yi Z, Zhong J (2018) Diagnosis of diabetic retinopathy using deep neural networks. IEEE Access 7:3360–3370

Giancardo L, Meriaudeau F, Karnowski TP, Li Y, Garg S, Tobin KW Jr, Chaum E (2012) Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal 16(1):216–226

Goatman KA, Fleming AD, Philip S, Williams GJ, Olson JA, Sharp PF (2010) Detection of new vessels on the optic disc using retinal photographs. IEEE Trans Med Imaging 30(4):972–979

Habib MM, Welikala RA, Hoppe A, Owen CG, Rudnicka AR, Barman SA (2016) Microaneurysm detection in retinal images using an ensemble classifier. In: 2016 sixth international conference on image processing theory, tools and applications (IPTA). IEEE, pp 1–6

Habib MM, Welikala RA, Hoppe A, Owen CG, Rudnicka AR, Barman SA (2017) Detection of microaneurysms in retinal images using an ensemble classifier. Inform Med Unlocked 9:44–57

Harangi B, Toth J, Baran A, Hajdu A (2019) Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 2699–2702

Harini R, Sheela N (2016) Feature extraction and classification of retinal images for automated detection of diabetic retinopathy. In: Second international conference on cognitive computing and information processing (CCIP). IEEE, pp 1–4

Hoover AD, Kouznetsova V, Goldbaum M (2000) Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging 19(3):203–210

Inbarathi R, Karthikeyan R (2014) Detection of retinal hemorrhage in fundus images by classifying the splat features using SVM. Int J Innov Res Sci Eng Technol 3:1979–1986

Kahai P, Namuduri KR, Thompson H (2006) A decision support framework for automated screening of diabetic retinopathy. Int J Biomed Imaging 2006(45806):1–8

Karthikeyan R, Alli P (2018) Feature selection and parameters optimization of support vector machines based on hybrid glowworm swarm optimization for classification of diabetic retinopathy. J Med Syst 42(10):195

Kauppi T, Kalesnykiene V, Kamarainen JK, Lensu L, Sorri I, Raninen A, Pietilä J et al (2007) The diaretdb1 diabetic retinopathy database and evaluation protocol. In: BMVC, vol 1, pp 1–10

Koh JE, Ng EY, Bhandary SV, Laude A, Acharya UR (2018) Automated detection of retinal health using PHOG and SURF features extracted from fundus images. Appl Intell 48(5):1379–1393

Lachure J, Deorankar AV, Lachure S, Gupta S, Jadhav R (2015) Diabetic Retinopathy using morphological operations and machine learning. In: 2015 IEEE international advance computing conference (IACC). IEEE, pp 617–622

Morales S, Engan K, Naranjo V, Colomer A (2015) Retinal disease screening through local binary patterns. IEEE J Biomed Health Inform 21(1):184–192

Navarro PJ, Alonso D, Stathis K (2016) Automatic detection of microaneurysms in diabetic retinopathy fundus images using the L* a* b color space. JOSA A 33(1):74–83

Niemeijer M, Staal J, van Ginneken B, Loog M, Abramoff MD (2004) Comparative study of retinal vessel segmentation methods on a new publicly available database. In: Medical imaging 2004: image processing, vol 5370. International Society for Optics and Photonics, pp 648–656

Porwal P, Pachade S, Kamble R, Kokare M, Deshmukh G, Sahasrabuddhe V, Meriaudeau F (2018) Indian diabetic retinopathy image

dataset (idrid): a database for diabetic retinopathy screening research. Data 3(3):25
Rahim SS, Palade V, Shuttleworth J, Jayne C (2016) Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing. Brain Inform 3(4):249–267
Roychowdhury S, Koozekanani DD, Parhi KK (2012) Screening fundus images for diabetic retinopathy. In: 2012 conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR). IEEE, pp 1641–1645
Roychowdhury S, Koozekanani DD, Parhi KK (2013) DREAM: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform 18(5):1717–1728
Selvathi D, Prakash NB, Balagopal N (2012) Automated detection of diabetic retinopathy for early diagnosis using feature extraction and support vector machine. Int J Emerg Technol Adv Eng 2(11):103–108
Seoud L, Chelbi J, Cheriet F (2015) Automatic grading of diabetic retinopathy on a public database
Seoud L, Hurtut T, Chelbi J, Cheriet F, Langlois JMP (2016) Red lesion detection using dynamic shape features for diabetic retinopathy screening. IEEE Trans Med Imaging 35(4):1116–1126
Sisodia DS, Nair S, Khobragade P (2017) Diabetic retinal fundus images: preprocessing and feature extraction for early detection of diabetic retinopathy. Biomed Pharmacol J 10(2):615–626
Somasundaram SK, Alli P (2017) A machine learning ensemble classifier for early prediction of diabetic retinopathy. J Med Syst 41(12):201
Sood M (2017) Performance analysis of classifiers for seizure diagnosis for single channel EEG data. Biomed Pharmacol J 10(2):795–803
Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B (2004) Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23(4):501–509
Thammastitkul A, Uyyanonvara B (2016) Diabetic Retinopathy Stages Identification Using Retinal Images, p 20
Vaishnavi J, Ravi S, Devi MA, Punitha S (2016) Automatic diabetic assessment for diabetic retinopathy using support vector machines. IJCTA 9(7):3135–3145
Venkatesan R, Chandakkar P, Li B, Li HK (2012) Classification of diabetic retinopathy images using multi-class multiple-instance learning based on color correlogram features. In: 2012 annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 1462–1465
Wang Z, Yang J (2018) Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. In: Workshops at the thirty-second AAAI conference on artificial intelligence
Wang S, Tang HL, Hu Y, Sanei S, Saleh GM, Peto T (2016) Localizing microaneurysms in fundus images through singular spectrum analysis. IEEE Trans Biomed Eng 64(5):990–1002
Wilkinson CP, Ferris FL III, Klein RE, Lee PP, Agardh CD, Davis M, Group, G. D. R. P. et al (2003) Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 110(9):1677–1682
Wulandari CD, Wibowo SA, Novamizanti L (2019) Classification of diabetic retinopathy using statistical region merging and convolutional neural network. In: IEEE Asia pacific conference on wireless and mobile (APWiMob). IEEE, pp 94–98
Xiao D, Bhuiyan A, Frost S, Vignarajan J, Tay-Kearney ML, Kanagasingam Y (2019) Major automatic diabetic retinopathy screening systems and related core algorithms: a review. Mach Vis Appl 30(3):423–446
Yen GG, Leong WF (2008) A sorting system for hierarchical grading of diabetic fundus images: a preliminary study. IEEE Trans Inf Technol Biomed 12(1):118–130
You J, Li Q, Guo Z (2016) Automatic mobile retinal microaneurysm detection using handheld fundus camera via cloud computing. Electron Imaging 2016(11):1–5
Yu F, Sun J, Li A, Cheng J, Wan C, Liu J (2017) Image quality classification for DR screening using deep learning. In: 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 664–667