IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-4, May 2018 E-ISSN: 2347-2693

HM-SPCA: Hybrid Method for Automatic Detection of MA Using MinIMas with Sparse PCA in Diabetic Retinopathy

D. Ashok kumar[1], A. Sankari*[2]

[1] Department of Computer Science, Government Arts College, Trichirappalli-620 *022*
[2] Department of Computer Science, Part time Research Scholar, Bharathidhasan University, Trichirappalli-620 022

[1]*Corresponding Author: sankarimca86@gmail.com*

Available online at: www.ijcseonline.org

Abstract— Microaneurysms (MAs) is an earliest lesions in DR detection, its plays a challenging role in diabetic retinopathy (DR) diagnosis. It has been an active research in medical image processing and so many machine learning algorithms has been developed for MA detection. The First Stage of detection is consisting of clearer segmentation of optical disc area in retina using a new Minimum Intensity Maximum Solidity (MinIMas) algorithm on fundus dataset, then extract bright lesion and red lesion using Gaussian mixture models. Set of feature extracted that the second stage of the system, finally machine learning approach is a0pplied for lesion classification. In this paper, a hybrid of segmentation and unsupervised classification of sparse PCA (HM-SPCA) for MA detection is proposed so that enhanced output is obtained. This proposed algorithm achieves great analysis in lesion segmentation with minimum false alarm. Furthermore, effective features can be extracted due to sparse properties of PCA (Principle Component Analysis) which merge the elastic net penalty with PCA together. Thus, the projected DR detection system enhanced its performance by reducing false positives compared with existing algorithms in lesion classification, and hence this approach can be applied to improve the beneficence in earlier vision detection of patients for diabetic retinopathy.

Keywords—Diabetic Retinopathy, MA, Optic disk, Blood vessel, Classification, Sparse, PCA

I. INTRODUCTION

In recent years, Diabetic retinopathy (DR) becomes major cause for blindness, which is also known as eye diseases [1]. Regular screening is recommended to diabetic patients for early diabetic diagnosis, which can help them to prevent loss of sight. Furthermore, a huge amount of diabetic patients are undergone screening process, which leads to increase the workload for ophthalmologists. In order to overcome this problem, an automatic DR detection is necessary to improve the diagnosis speed and accuracy of detection [2]. Usually, retinal fundus images with low illumination and low contrast images are widely used for automatic DR detection [3]. In fundus images, two categorize of lesions can be detected, as name of red lesion (microaneurysms (MA) and hemorrhages), bright lesion (hard exudates and cotton wool spots). Among these lesion, MA plays a vital role in DR systems, which appear a small circular dark spots on the surface of retina [4] (refer Figure 1).The first phase towards automated detection of DR is the removal of background regions consisting of the optic disc (OD) and major portions of the blood vasculature in the fundus images and masking them out.

II. RELATED WORK

This phase is a challenging task in Diabetic retinopathy (DR) system, since automatic detection system falsely detects OD as exudate or blood vessels as red lesion if it's not masked out correctly. One previous work on recognition of DR utilizes the maximum variance to obtain the OD center and region-growing segmentation is used to detect bright lesions [5]. Moreover, optical disk center have been tracked using pyramidal decomposition and template-based matching is used for locating of disc by utilizing the Hausdorff distance measures on binary fundus edge images [5]. Once the background regions have been extracted from the foreground, then task is the detection of retinopathic lesions that can be classified in DR system [6].

In this paper two key contributions are invented towards DR detection system. In first phase, a novel MinIMas overlap algorithm is proposed to initialize the OD center in low contrast fundus image. Most of the prior approached have not been achieved successful arte more than 91% on real public dataset [7]. But our segmentation approach achieves 100% accuracy in dataset DRIVE [8] and 98.68% accuracy in STARE dataset [9] for OD detection. Additionally, most

existing algorithm [9] has not been robust to field of view (FOV) variation of input images.

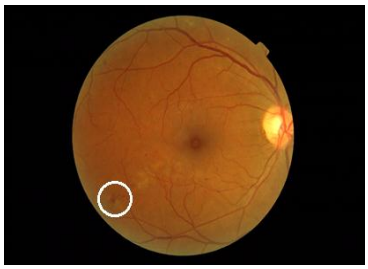


Figure 1: Fundus Image with MA

Also, some of the prior methods affect from image over-training since they follow the vessel-branch networks after extracting the blood vessels, looking for merging patterns in images that do not have a visibly bright OD, thus resulting in false detections [10]. Our segmentation algorithm is trained with varying FOV and illuminations images and thus, it does not affect from over-training. Also, the MinIMas overlap algorithm does not fail in the presence of exudates which similar to bright spot OD detection.

Several machine learning algorithms are proposed for MA classification in DR systems. In [11] proposed a hybrid approach for MA detection, which combined morphological top-hat transform with a Knn classifier. Furthermore [12] proposed a unified structure for MA detection in which combined the Gaussian mixture model (GMM) with a logistic regression is used. After that, [13] implemented a hybrid approach named as Dictionary Learning (DL) with Sparse Representation Classification (SRC) in MA detection system. In this method, firstly, Multi-scale Gaussian Correlation Coefficients filtering was applied on fundus image to locate all the possible features. After that two dictionaries were trained based on MA and non-MA candidate's features. Finally, original MA candidates were classified by SRC.

All mentioned prior approaches are having a common problem that non-Mas features vary in a wide range. In order to collect enough training for all non-Mas, the number of samples size has to be very large for training. MAs in all fundus images are having same circular stricture, but non-Mas are having different structure. In this scenario, collection of non-Ma training set is very difficult and large training set yields time consuming also will cause imbalance problem [14].

The paper is organized as follows: In section III, Fundus image pre-processing for OD segmentation are discussed. A Novel MinIMas overlap algorithm for Segmenting OD and blood vessel are discussed in section III.1. Types of Candidate Extraction are presented in Section III.2. The sparse based MA classification is proposed in section III.3. Experiments are carried out on fundus database and results are reported in Section IV and Section V concludes the paper.

III. METHODOLOGY

The complete structure of MA detection, implementation flow is proposed HM-SPCA method in Figure 2. It contains four stages of process named as, 1) Preprocessing 2) Segmentation 3) Feature Extraction 4) Classification. Firstly, preprocessing algorithm is applied to contrast enhancement and convert an image into more suitable form for next processing stage. Next, segmentation of optical disc area is founded by MinIMas algorithm and morphological based blood vessel extraction is applied on fundus image.

Furthermore, detection a foreground images consisting of subtracting the unwanted details of OD and blood vessels from background. Foreground image is consisting of red and bright lesion values. After that, all possible features are extracted from each candidates based on shape based, color and intensity based and Gaussian features. Finally, proposed HM-SPCA modal is used for feature training and classification.

A) Preprocessing Stage

As states in [15], Green channel image are more contrast to detect MA in fundus images. In this section image preprocessing works are discussed for noise removal and contrast enhancement.

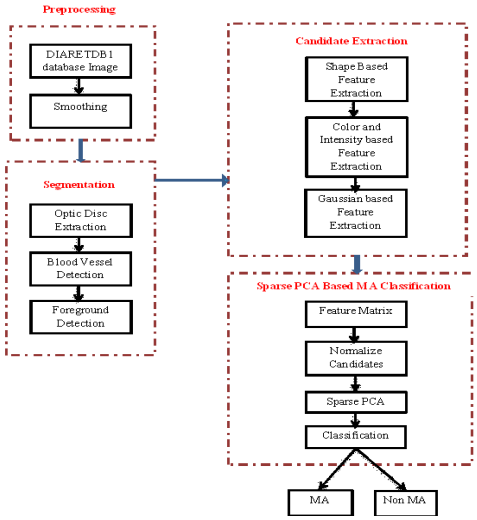
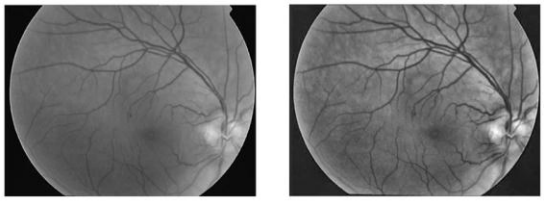


Figure 2 Proposed Method

In real world fundus images are affected by noise, intensity variation and low illumination. It leads to serious impact on lesion detection and MA classification in DR systems [16]. In order to address these issues, Green channel images are

extracted from fundus images, which having higher contrast with background [15]. Additionally, Green channel image is adapted to contrast enhancement based on contrast limited adaptive histogram equalization (CLAHE) [17] approach in order to make hidden features are more visible. Finally, smoothing filter is applied on enhanced images with width 5 and standard deviation 1 for reducing the effects of noise on images. Figure 3 describes about input fundus image and preprocessed image with contrast enhancement.



(a) (b)

Figure 3 a) Input Fundus image b) Preprocessed image

B) Segmentation

In first stage of our proposed method, the back ground of all fundus images are consisting of optical disk (OD and blood vascular arc is detected and masked out. After preprocessing all images are resized and all pixel ranges are rescaled in [0, 1].

Next, a novel minimum Intensity Maximum Solidity overlap algorithm is used to detect OD in fundus image. This process is challenging since at low illumination by non-mydriatic cameras result is significantly dark OD in an image. Additionally, due to varying fields of view (FOV), only a portion of the entire OD may exist in some images. The innovation of this algorithm is that it works for all such fundus images with varying illumination, such that images with any portions of the considerably bright OD get masked out, while images without the OD remain unmasked or slightly masked. Next, major portions of the blood vessels are morphologically extracted and masked out. The implementation stages are presented as follows.

C) OD Detection using MinIMas overlap algorithm

The OD is one of the brightest regions and should overlap one of the largest and reddest regions (blood vessel), in fundus images. Owing to it, blood vessels are masked out from OD. Additionally, for all images the bright lesions are lie around blood vessel; those regions will also most challenging intersection regions in an image. In order to mask out the OD region, domain knowledge is applied that the blood vessels are thickest region around OD and masked out into thinner vessels towards boundary. Since a scaled image has bright pixels with pixel value 1 and red pixels are having 0 values, consequently the summation of all pixel intensity within a circular mask applied to all the regions of intersection will encounter more pixels approaching 0 near the OD. Also, the OD is generally more solid in structure as compared to bright lesions. Thus, the OD would appear at the region of intersection with minimum pixel intensity within a circular mask and maximum compactness, where solidity is a degree of compactness.

Mathematically, two sub image lesion regions are extracted from I_greenimage, named as red lesion I_rand bright lesionI_b. Then, we define an overlap function, that uses a predefined constraint(ne), such that is pixels of are located in the neighborhood of pixels in I_b overlapping regions(R) are identified in I_b .Let $R \in R_1, R_2, .. R_i$ be the l overlapping regions discovered. Now, at individual R_i, we describe a circular mask whose center is at the centroid of R_i and radius is the maximum major axis length of all regions in R. Then, we estimate the sum of pixel intensities of all the pixels within individual such circular mask described by a function intensity. Also, we compute the solidity of individual region by,

$$Solidity(S) = \frac{Area(R_i)}{ConvexArea(R_i)} \leq 1 \quad (1)$$

Where convex area is states that the smallest convex polygon's area which contains the particular region of interest. The OD region will be a region with minimal intensity and maximal solidity [18]. Thus,

$$R_1 = \max_{intensity} (R_i)$$

$$R_2 = \max_{solidity} (R_i)$$

$$R_{OD} = R_1 \cup R_2$$

Such that, $R_i = overlap(I_b, I_r, ne)$ (2)

D. Blood Vessel Detection

Next, Major portion of blood vessels are masked out from fundus image, since OD is extracted. For each image, median filter is applied to achieve gradient smoothened background I_{bg} from green channel image I_{green} [19]. Shade corrected image I_{sc} can be estimated by subtraction of background image from green channel I_{green} image [19].

$$I_{sc} = I_{green} - I_{bg} \quad (3)$$

Only the negative pixel values of I_{sc} are employed while positive pixel values are ignored to estimate isolated shade corrected image (I_{iso}). After that, I_{sc} is undergone

Normalization, contrast adjustment with background masking with image '***g'*** and also thresholding is applied to obtain I_{iso} with only red intensity of the image. Finally, highest pixel areas are retained by region growing operation, followed by global thresholding is subjected to I_{iso}. This extracted region is highest pixel area of blood vessel region of an image I_{bv}.

E. Candidate Extraction (Red and Bright Lesion)

Next, foreground regions are identified by masking out of OD and major blood vessel portions from an input fundus image, in which several foreground regions can be identified as bright lesion and red lesion. However, thinner blood vessels that may remain unmasked might be sensed as a red lesion. Likewise, neighboring regions of the blood vessel may sense as candidate regions for bright lesions. Therefore, false positives need to be removed out from the foreground candidate regions by morphological operator.

a)Bright Lesion Detection

Image is subjected to contras enhancement to detect I_{green} candidate regions for bright lesion in an image. Then morphological erode operation is applied to enhanced image with linear structural element of size (50, 1), which is followed by image reconstruction. Next, this image is subtracted from I_{green}, followed by normalization and contrast enhancement to produce I_{br}. Next, I_{br} is normalized and globally thresholded to find candidate regions for bright lesions, also called as bright objects in image I_{bl}. Finally, the OD mask and the blood vessel are subtracted from the bright lesions, i.e., $I_{bl} = I_{bl} - (R_{oD} + g)$.

b) Red Lesion Detection

Finally, the red lesion objects are detected by subtracting blood vessel and optical disc area from shade corrected foreground image, which is also called as candidate regions of red lesions. i.e., $I_{ri} = I_{sc} - (I_{bv} + R_{oD})$.

c) Feature Extraction

In this section, features of all candidate regions are extracted based on shape-based, intensity and color based and Gaussian filtering based. Since MA region appears as red thicker region in circular shape with low intensity of fundus image. Therefore, we extract the following features for individual candidate to differentiate MA from non-MA candidate. Here, totally 18 candidate features are extracted for classification of lesion region effectively [11]. These features can be grouped into 3 categories. Table I describes shape based features. Color and intensity based features are described in Table II. Finally, Gaussian filtering based features is defined in Table III. In summary total 36 features have been extracted in for each candidate regions.

F. Sparse PCA Based MA Classification

In this section, the proposed sparse PCA is illustrated for MA classification. First, we will make description on PCA and sparse PCA. Next, MA statistics are given to control detection process. Finally, an evaluation method is utilized to validate the accuracy of classification in MA detection.

a)Principal Component Analysis

In recent days, PCA is a linear dimensionality reduction method in all statistical applications. Its basic norm is to increase the variance of predictions on new directions. Load vector will used to estimate a new direction by group of orthogonal vectors and projections variance on load vectors will determine the sequence of vectors. Assumed a set of training samples $X \in R^{NXm}$ with *m* variables and *N* samples, the objective function can be written as follows [20], which can be used for optimization problem.

$$\max_{v \neq 0} \frac{v^T X^T X v}{v^T v} \quad (4)$$

Where $v \in R^m$. Steady solutions of above equation can be determined by

$$\frac{1}{\sqrt{n-1}} X = U \Sigma V^T \quad (5)$$

Wherever $U \in R^{nXn}$ and $V \in R^{mXm}$ are all unitary matrices, $\Sigma \in R^{nXm}$ contains of non-negative singular value in descending order $\left(\sigma_1 \geq \sigma_2 \ldots \geq \sigma_{\min(m,n)} \geq 0\right)$ along the main diagonal, other elements in Σ are zeros. Load vectors is represent as orthogonal column vectors in *V*. Prediction of *X* on *i*th column has the variance equals to σ_i^2. Select the first 'a' columns in *V* to comprise a new load matrix $P \in R^{mXa}$ which is also called the principal component subspace (PCS).

$$T = XP \quad (6)$$

A sample vector *x* on the PCS:

$$t = P^T x \in PCS \quad (7)$$

Nearly the value 'a', it is assumed by Cumulative Percent Variance (CPV) which is calculated as follows:

$$CPV(a) = 100\left[\frac{\sum_{j=1}^{a} \lambda_j}{\sum_{j=1}^{m} \lambda_j}\right]\% \quad (8)$$

The CPV is the percent variance calculated by the first principle components (PCs); *m* is the total number of eigen values. It is very subjective what value for 'a' is suitable; this problem is discussed in [21].

Table I Shape Based Features

S.NO	Features	Formula	Variable declaration
1	Area	$a = \sum_{j=n} 1$	$\Omega = Numberofpixelincandidateregions$
2	Perimeter	$p = n_o \sqrt{2} + n_c$	$n_o = Numberofoddchaincodes$ $n_c = Numberofevenchaincodes$
3	Aspect Ratio	$r = \frac{l}{w}$	$l = MajorAxislenghtofeachcandidates$ $w = MinorAxislenghtofeachcandidates$
4	Circularity	$c = \frac{p^2}{4 * \Pi * a}$	$a = area$ $n_c = perimeter$
5	Compactness	$v = \sqrt{\frac{\sum_{j=2}^{n} (d_j - \bar{d})}{n}}$	$d_j = Distancefromcentertoboundarypixel$ $\bar{d} = Meanofalldistancefromcentriodtoboundary$ $n = Numberofedgepixel$

Table II Color and Intensity Based Features

1	Total Intensity_Green Plane	$I_{green} = \sum_{j=n} g_i$	$g_i = greenpixelvalueof j^{th} pixel$
2	Total Intensity_Shade Corrected image	$I_{sc} = \sum_{j=n} S_i$	$s_i = shadecorretedpixelvalueof j^{th} pixel$
3	Mean intensity_Green Plane	$m_{green} = \frac{i_{green}}{a}$	$a = area$ $i_{green} = Total\ intensityofgreenplane$
4	Mean intensity_Shade Corrected Image	$m_{sc} = \frac{i_{sc}}{a}$	$i_{sc} = Total\ intensityofshadecorrecte\ dimage$ $a = area$
5	Normalized Intensity_Green Plane	$NI_{green} = \frac{1}{\sigma}(i_{green} - \bar{x})$	$i_{green} = Total\ intensityofgreenplane$ $\sigma = Standard\ Deviationofbackground\ image\ I_{bg}$ $\bar{x} = Averagepixelvalueofbackground\ image\ I_{bg}$
6	Normalized Intensity_Shade corrected Image	$NI_{sc} = \frac{1}{\sigma}(i_{sc} - \bar{x})$	$i_{sc} = Total\ intensityofshadecorrecte\ dimages$ $\sigma = Standard\ Deviationofbackground\ image\ I_{sc}$ $\bar{x} = Averagepixelvalueofbackground\ image\ I_{sc}$
7	Normalized Mean Intensity_Green Plane	$NM_{green} = \frac{1}{\sigma}(m_{green} - \bar{x})$	$m_{green} = Mean\ intensityofGreenplane$ $\sigma = Standard\ Deviationofbackground\ image\ I_{bg}$ $\bar{x} = Averagepixelvalueofbackground\ image\ I_{bg}$
8	Normalized Mean Intensity_ Shade corrected Image	$NM_{sc} = \frac{1}{\sigma}(m_{sc} - \bar{x})$	$m_{sc} = Mean\ intensityofshadecorrecte\ dimage$
9	Mean contrast of Edge pixels	$C = \frac{\sum_{jca} g_j}{num_A} - \frac{\sum_{jca} g_j}{num_B}$	$A = 8neighbouringpixelofcurrentpixelwithhighintensity$ $B = Remaining8pixelofA$ $num_A = NumberofpixelinA$ $num_B = NumberofpixelinB$
10	The mean intensity of I_{bg}	$M_{bg} = \frac{I_{bg}}{num_{bg}}$	$num_{bg} = Numberofpixelinbackground\ image$
11	Standard Deviation of I_{bg}	$\sigma_{bg} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$	$N = Numberofpixels$ $x_i = pixelsati^{th} location$ $\mu = meanofanimage$

Table III Gaussian Features

1	Mean and standard deviation of Gaussian filter	i_{gf} σ_{gf}	Filter image of with different scale σ = 1,2,4,8
2	Correlation co-efficient of candidates	$cc_{\max}$ $cc_{\min}$ cc_{avg}	$cc_{\max} = correlationcoefficientsof \max imumcandidate$ $cc_{\min} = correlationcoefficientsof \min imumcandidate$ $cc_{avg} = correlationcoefficientsofaveragecandidate$

b)Sparse PCA

Many features are linear relevant with each other, but some of them are observed with irrelevant features. These variables are called as weak relevant. Meanwhile, the main aim of PCA is to find the maximum variance on certain loading vectors, some principal components will be unavoidably characterized by such kind of weak relevant variables which will decrease the detection process.

Additional L1 norm constraint will increase the objective function (4) by LASSO based PCA:

$$\sum_{j=1} \left\lfloor V_j \right\rfloor \le sparsity \qquad (9)$$

Here v_j represents j[th] element of the load vector $j \in \{1,2..m\}$, number of non-zero elements in load vector can be represented by sparsity. BY using this constraint, PCA will be represented by less novel variables. Numerous algorithms have been established for solving this issue. Amongst, one of the most effective and representative methods is proposed by [22]. Mostly, PCA can first be altered exactly in terms of a regression problem. Then, the L1 norm constraint can be announced, altering this regression problem to an elastic-net regression. A short evaluation will be given as follows.
Assume the fact that each PC is a linear combination of the 'm' variables, thus its sparse loadings can be achieved by regressing the PC on the 'm' variables. For each 'i', let $z_i = XV_i$ represents the i[th] principal component and λ be a positive constant. The regression approximation is given by:

$$\hat{\mu}_{regg} = \arg_{\mu} \min \left\| z_i - X\mu \right\|^2 + \lambda \left\| \mu \right\|^2 \qquad (10)$$

Here, $\hat{\mu}_{regg}$ van be obtained by $\hat{v} = \dfrac{\hat{\mu}_{regg}}{\left\| \hat{\mu}_{regg} \right\|^2}$ and $\hat{v} = V_i$ is represents i[th] loading.

Now, let's discuss the contribution of L2 norm in the objective function and the selection of λ. When there are numerous correlated variables in a linear regression model, their constants can be poorly estimated and display high variance. A broadly large positive constant on one variable can be disregarded by a similarly large negative constant on its correlated variables. By imposing a size constraint on the coefficients, as the L2 norm in ridge regression, this problem is relieved with a relatively small value of λ. Also, when m>N, if $\lambda = 0$, normal multiple regression has not produce novel solution that is exactly equal to V_i, it is the same when N>m, and 'X' is not a full rank matrix. With the L2 norm, we can constantly provide the unique solution in all conditions.

Then, we will enhance the L1 penalty to Eq. (10) and find the objective function using below expression:

$$\hat{\mu} = \arg_{\mu} \min \left\| T_i - X\mu \right\|^2 + \lambda \left\| \mu \right\|^2 + \tau \left\| \mu \right\| \iota \qquad (11)$$

Here, $\left\| \mu \right\|_i = \sum_{j=1}^{m} \left| \mu_j \right|$ is the L1 norm of μ. And $\bar{V}_i = \dfrac{\mu}{\left\| \mu \right\|}$ an approximation of the loading vectors. τ is a tradeoff constraint controls the sparsity of loading vectors, larger τ relates to fewer non-zero elements in loading vectors, In training, we can select a proper τ by specifying the sparsity which will be conversed later. Resolving this objective function, we can attain the sparse PCA and find the sparse loading vectors V_i. As exposed in Eq. (11), penalized least squares problem can be resolved by

effective LARS-EN algorithm [23] using the elastic net penalty.

Sparse PCA has numerous advantages, the inspiration introducing sparse PCA for MA detection are shortened as follows [24]:

1) L1 norm constraint is presented to achieve the variance-sparsity trade-off and weak relevant variables in feature matrix can be nominated out by sparsity. Also, sparse loadings make the principal components easy to understand by a human.

2) L2 norm constraint is presented to provide the novel (Unique) solution in all conditions. Also, multicollinearity can be improved with this constraint.

IV. CLASSIFICATION

After the bright objects and red objects have been detected, Sparse PCA classifiers are trained to reject false positives, so that the accuracy of the lesion detection increases. Next step of the algorithm, we firstly train the sparse PCA classifiers on the bright and red objects of the training data set comprising of 28 images in the DIARETDB0 data set, and test the performance of classification on the objects of the test set including of 100 images from the same data set.

a)Sparse Classification

Subspace orthogonal matching pursuit (SOMP) method is used to increase the speed up the classification process, which will efficiently classify MA from Non- MA candidates.

Consider N images with MX N feature matrix, our classification steps are follows:

- Training matrix is constructed by extracted features for all N images. i.e., $T_N = F_1, F_2, ..F_N$ Each F contains 1× 36 features for both bright lesion and red lesions.
- For each test image y, $y = F_y, x_i, i = 1,2,3..N$ is extracted and coefficient vector is obtained $x_1, x_2, ...x_N$ by orthogonal matching pursuit algorithm.
- The recognition coefficient x_i of each candidate is related to the testing sample to calculate the residual:

$$r_i = \left\| y - F_y x_i \right\|^2, i = 1,2..N$$

- Finally, recognition result is identified by, $Identity(y) = \arg\min(ri), i = 1,2,..N$

This SOMP matching method is greatly decrease the size of the training sample, which makes it easy to calculate sparse coefficients. Further, reduces the number of the loop count and improve the recognition rate with minimum error rate for large size of testing sample.

The method SOMP is used to greatly decrease the size of the calculated sample, which makes it easy to calculate sparse coefficients and error results in the comparison, and reduces the number of the loop count accordingly, improving the precision level than the calculation of samples with larger size. More important, this method can significantly improve the recognition result.

V. RESULTS AND DISCUSSION

A. Dataset

Our proposed DR system has been trained and tested by both normal and DR affected patients by using following publically available datasets.

- DIARETDB1 data set [6] contains of 89 images with 50°FOV. These 89 images have been divided to two groups: 28 training images and 61 test images, by the authors such that the training and test data sets were independent of one another. All the 89 images are manually marked for lesions as ground truth (RHE;RCWS;RMA;RHA).

- A set of 20 images from DIARETDB0 was used to train the Sparse PCA classifier in the first candidate object extraction step as described in previous section. Also, DIARETDB1 was used to train and test the MA detection system. All of the images in both sets were acquired digitally.

B.Performance Comparison

In this section, performance of our DR system is analyzed with existing method in term of MA detection. In the image segmentation stage, the finding of the OD and blood vessel region plays a critical role to ensure fewer instances of false positive while detecting bright lesions. Our novel MinIMas overlap algorithm has exposed to have a higher accuracy of 99% for optic disc and blood vessel segmentation on public data sets [6].

Performance of our proposed DR detection system is estimated by following metrics.

$$sensitivity(SEN) = \frac{TP}{TP + FN}$$

$$specificity(SPEC) = \frac{TN}{TN + FP}$$

$$Accuracy\,(Acc) = \frac{TP + TN}{TP + TN + FP + FN}$$

Here,
TP – True Positive
FP – False Positive
TN – True Negative
FN – False Negative

Finally, Our DR system can classify the fundus images as normal with free from retinopathy features and as abnormal with red and black lesion features.

Comparative analysis of proposed DR system in classifying lesion is compared with existing approaches is shown in Table IV. From which, we can say our proposed method having higher sensitivity and specificity in bright and red lesion detection with existing works on public datasets. This performance comparison graph is shown in Figure 4 and Figure 5.

Table IV- Comparative Analysis of Lesion Detection

Lesion Type	Method	SEN (%)	SPE (%)
RED	Bhalero et.al.[24]	82.6	80.2
RED	Kauppi et.al.[6]	77.78	76.47
RED	DREAM[25]	80	85
RED	Proposed Method	85	87
Bright	Sophorak et.al.[26] (47 images)	43.48	99.31
Bright	Walter et.al.[27] (47 images)	66	98.64
Bright	Welfer et.al.[28] (47 images)	70.48	98.84
Bright	DREAM [25]	74.2	98
Bright	Proposed Method	79	99

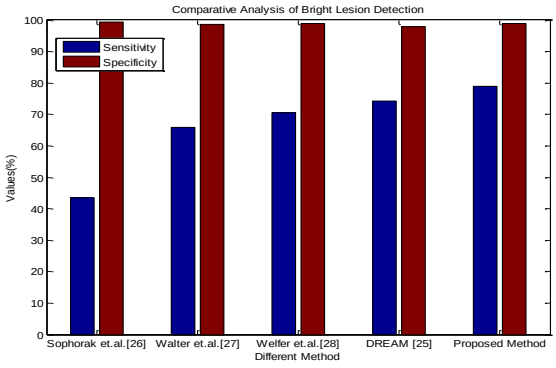


Figure 4 Comparison of Red Lesion Detection

From Figure 4 and Figure 5, we detect that existing works [25] on lesion classification using the DIARETDB1 data set also verified a preference for high SPEC for bright lesion classification and high SEN for red lesion classification, which is in accordance with our proposed classifier evaluation criteria.

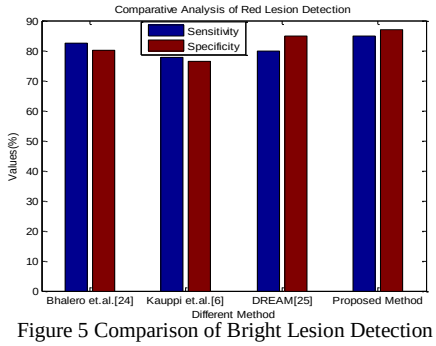


Figure 5 Comparison of Bright Lesion Detection

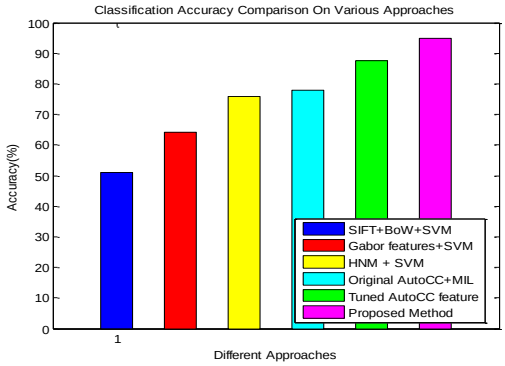


Figure 6 Comparison of MA accuracy

To further understand where the misclassified are placed and how the classification is performing with respect to individual classes, we have also provided Accuracy graph in Figure 6 for the different feature extraction approaches and our proposed HM-SPCA hybrid approach. When using the proposed feature with 40-dimensional space, Tuned Auto CC features performs the classification almost equally well for all the classes [29]. But our proposed hybrid approach has better accuracy that tuned method.

Our proposed work have achieved sensitivity, specificity and accuracy of 98%,98%,95% respectively as compared to Mohamed Oma. [30] who achieved classifier of 81.55%, 81.80%, and 79.27% respectively using SVM and 90.63%, 88.70%, 88.93%, respectively with the GMM classifier and RBF neural network achieved 98.68%, 94.81% , 96.73%. Our proposed system produces higher performance with the same dataset DIARETDB0, which make it suitable for an automated medical system for MA classification in fundus image. Figure 7 shows the performance evaluation of proposed method with

Mohamed Oma [30] system in terms of sensitivity, specificity and accuracy.

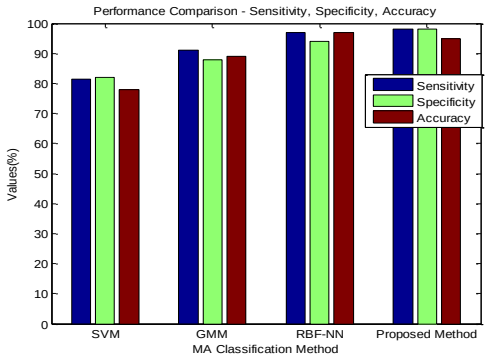


Figure 7 Comparison of MA accuracy

VI. CONCLUSION and Future Scope

Data for non-MAs vary in a large range, how to collect non-MA training set is quite subject, large training set is not only time consuming but also will cause the class imbalance problem. In this paper, hybrid of segmentation and unsupervised classification of sparse PCA (HM-SPCA) was proposed to detect MA. Sparse PCA is applied to detect the latent structure of MA data, once a model has been developed that reflect the MAs, any departure from standard MAs are detected by monitoring the data. Meanwhile there is no need to consider the non-MA class samples, the class imbalance problem can be avoided. When Normal (MA) and abnormal (Non-MA) status is unknown or difficult to find, in that cases un-supervised classification is more effectively suitable. Meanwhile, sparse PCA automatically selected effective features from an image. Finally, the experiment is carried out on public dataset, which shows promising result compared with the state-of-the-art methods. The future work contains the following issues: firstly, more effective features such as texture feature should be included for distinguishing the MAs from FPs. Then, applying our proposed HM-SPCA framework to hard exudates and hemorrhages detection is another motivating topic for future study.

REFERENCES

[1] B. Antal, A. Hajdu, "*An ensemble-based system for microaneurysm detection and diabetic retinopathy grading*," IEEE transactions on bio-medical engineering, vol. 59, no. 6, pp. 1720-6, 2012.

[2] D. B. Mukamel, G. H. Bresnick, J. C. Dickinson, D. R. Cole, " *A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy*," Ophthalmology, vol. 107, no. 1, pp. 19-24, 2000.

[3] M. Abràmoff, M. Garvin, M. Sonka, "*Retinal imaging and image analysis*," IEEE Reviews in Biomedical Engineering, vol. 3, pp. 169–208, 2010.

[4] B. Antal, A. Hajdu, "*Improving microaneurysm detection using an optimally selected subset of candidate extractors and preprocessing methods*," Pattern Recognition, vol. 45, no. 1, pp. 264-270, 2012.

[5] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H.Williamson,"*Automated localisation of the optic disk, fovea, and retinal blood vessels from digital colour fundus images*," Br. J. phthalmol,, vol. 83, no. 8, pp. 902–910, 1999.

[6] T. Kauppi, V. Kalesnykiene, J.-K. Kmrinen, L. Lensu, I. Sorr, A. Raninen, R. Voutilainen, H. Uusitalo, H. Klviinen, and J. Pietil, "*Diaretdb1 diabetic retinopathy database and evaluation protocol,"* Proc. of the 11th Conf. on Medical Image Understanding and Analysis (MIUA2007), pp. 61–65, 2007.

[7] A.-H. Abdel-Razik Youssif, A. Ghalwash, and A. Abdel- Rahman Ghoneim, "*Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter,"* IEEE Transactions on Medical Imaging, vol. 27, no. 1, pp. 11 –18, jan. 2008.

[8] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "*Ridge based vessel segmentation in color images of the retina*," IEEE Transactions on Medical Imaging, vol. 23, pp. 501–509, 2004.

[9] A. Hoover, and M. Goldbaum, "*Locating the optic nerve in retinal image using the fuzzy convergence of bood vessels,"* IEEE Transactions on Medical Imaging, vol. 22, August 2003.

[10] F. ter Haar, "*Automatic localization of the optic disc in digital colour images of the human retina*," M.S. Thesis in Computer Science, UtrechtUniversity, Utrecht, 2005.

[11] M. Niemeijer, G. B. Van, J. Staal, M. S. Suttorp-Schulten, M. D. Abràmoff, "*Automatic detection of red lesions in digital color fundus photographs,"* IEEE Transactions on Medical Imaging, vol. 24, no. 5, pp. 584-92, 2005.

[12] C. I. Sánchez, *"Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images*," *Proceedings of SPIE* - The International Society for Optical Engineering, pp. 7260: 72601M-72601M-8, 2009.

[13] B. Zhang, F. Karray, Q. Li, L. Zhang, "*Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction*," *Information Sciences*, vol. 200, no. (1), pp.78-90, 2012.

[14] G. G. Gardner, D. Keating, T. H. Williamson, A. T. Elliott, "*Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool*," British Journal of Ophthalmology, vol. 80, no. 11, pp. 940-4, 1996.

[15] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, J.C. Klein, "*Automatic detection of microaneurysms in color fundus images*," Medical Image Analysis, vol. 11, no. 6, pp. 555-566, 2007.

[16] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, C. Roux, "*Optimal wavelet transform for the detection of microaneurysms in retina photographs*," IEEE Transactions on Medical Imaging, vol. 27, no. 9, pp. 1230-41, 2008.

[17] K. Zuiderveld, "*Contrast Limited Adaptive Histogram Equalization*," Graphics Gems, pp. 474-485, 1994.

[18] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "*Screening fundus images for diabetic retinopathy*," in 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, systems and Computers (ASILOMAR), 2012, pp. 1641–1645.

[19] A. Frame, P. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "*A comparison of computer based classication methods applied to the detection of microaneurysms in ophthalmic uorescein angiograms*," Comput. Biol. Med., vol. 28, pp. 225–238, 1998.

[20] B. C. Moore, "*Principal component analysis in linear systems: controllability, observability, and model reduction*," Automatic Control IEEE Transactions on, vol. 26, no. 1, pp. 17-32, 1981.

[21] S. Valle, A. Weihua Li, S. J. Qin, "*Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods†*," Industrial & Engineering Chemistry Research, vol. 38, no. 11, pp. 653-658, 1999.

[22] I. T. Jolliffe, N. T. Trendafilov, M. Uddin, "*A modified principal component technique based on the lasso,"* Journal of Computational & Graphical Statistics, vol. 12, no. 3, pp. 531-547, 2003.

[23] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, *"Least angle regression,"* Mathematic, vol. 32, no. 2, pp. 2004, 2004.

[24] A. Bhalerao, A. Patanaik, S. Anand, and P. Saravanan, "*Robust detection of microaneurysms for sight threatening retinopathy screening*," in Sixth Indian Conference on Computer Vision, Graphics Image Processing, 2008. ICVGIP '08, Dec. 2008, pp. 520 –527.

[25] Sohini Roychowdhury, Dara D. Koozekanani and Keshab K., *'DREAM: Diabetic Retinopathy Analysis using Machine Learning'*, IEEE Journal of Biomedical and Health Informatics 2013.

[26] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "*Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods*," Computerized Medical Imaging and Graphics, vol. 32, no. 8, pp. 720 – 727, 2008.

[27] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, "*A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina*," IEEE Transactions on Medical Imaging, vol. 21, no. 10, pp. 1236 –1243, Oct. 2002.

[28] D. Welfer, J. Scharcanski, and D. R. Marinho, "*A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images*," Computerized Medical Imaging and Graphics, vol. 34, no. 3, pp. 228 – 235, 2010.

[29] Ragav Venkatesan, Parag Chandakkar, Baoxin,*''Classification of Diabetic Retinopathy Images Using Multi-Class Multiple-Instance Learning Based on Color Correlogram features*",IEEE Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE 2012.

[30] Mohamed Omar,Fouad Khelifi and Muhammad Atif Tahir,*''Detection and Classification of Retinal Fundus Images Exudates using Region based Multiscale LBP Texture Approach*",IEEE, International Conference on Control, Decision and Information Technologies (CoDIT), 2016.

Authors Profile

Dr. D. Ashok Kumar working as Associate Professor & Head in the Department of Computer Science, Government Arts College, Tiruchirappalli - 620 022.

A.Sankari currently doing Part time Research Scholar in the Department of Computer Science, Bharathidhasan University, Tiruchirappalli-620 022.