# Clinically Relevant Classification and Retrieval of Diabetic Retinopathy Images

by

Parag Shridhar Chandakkar

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved August 2012 by the
Graduate Supervisory Committee:

Baoxin Li, Chair
Pavan Turaga
David Frakes

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

Diabetic retinopathy (DR) is a common cause of blindness occurring due to prolonged presence of diabetes. The risk of developing DR or having the disease progress is increasing over time. Despite advances in diabetes care over the years, DR remains a vision-threatening complication and one of the leading causes of blindness among American adults. Recent studies have shown that diagnosis based on digital retinal imaging has potential benefits over traditional face-to-face evaluation. Yet there is a dearth of computer-based systems that can match the level of performance achieved by ophthalmologists.

This thesis takes a fresh perspective in developing a computer-based system aimed at improving diagnosis of DR images. These images are categorized into three classes according to their severity level. The proposed approach explores effective methods to classify new images and retrieve clinically-relevant images from a database with prior diagnosis information associated with them. Retrieval provides a novel way to utilize the vast knowledge in the archives of previously-diagnosed DR images and thereby improve a clinician's performance while classification can safely reduce the burden on DR screening programs and possibly achieve higher detection accuracy than human experts. To solve the three-class retrieval and classification problem, the approach uses a multi-class multiple-instance medical image retrieval framework that makes use of spectrally tuned color correlogram and steerable Gaussian filter response features. The results show better retrieval and classification performances than prior-art methods and are also observed to be of clinical and visual relevance.

*To my Family and Friends*

ACKNOWLEDGEMENTS

First and foremost I would like to express sincere gratitude towards my advisor Dr. Baoxin Li for his continuous motivation, support and patience. Without his valuable guidance and immense knowledge, this thesis would not have been possible.

I also take this opportunity to thank the rest of my committee members, Dr. Pavan Turaga and Dr. David Frakes for being a part of thesis committee and examining my thesis report. I am thankful to all the administrative staff of ASU, especially Darleen Mandt, who has always helped me and has promptly processed my requests.

I thank Ragav Venkatesan, for his valuable contribution as a teammate on this project. I also want to thank all my lab-mates viz. Dr. Peng Zhang, for his timely help in solving all my doubts (and in creating new ones!) which has immensely helped me understand this problem better. Qiang Zhang, for helping me build a strong theoretical foundation and solving my numerous doubts. Hima Bindu, for providing me with the much needed morale boost when my approaches did not seem to work. This acknowledgement is incomplete without Charan Prakash and Jidnyasa Babar, without whom my thesis draft could never have been complete.

Again I would like to mention my friends back in India, and undergraduate Professor, Ashish Vanmali who introduced me to the beautiful field of Image Processing.

Lastly I would like to thank my parents, my brother Yogesh and sister-in-law Shraddha, and the rest of my family for their undying love and faith in me.

TABLE OF CONTENTS

Page

CHAPTER PAGE

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

# 1. INTRODUCTION

This chapter provides background regarding the DR problem and motivations for this thesis.

## 1.1 Diabetic Retinopathy: A Background

Diabetic retinopathy is an eye related problem that people may face as a complication of diabetes. To study DR in detail, it is very important to understand the simple model of the eye as shown in Fig. 1.

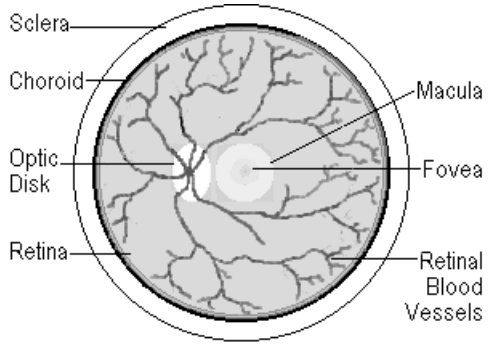



Fig. 1. Cross Sectional Model of the Eye

1. Retina: The retina is the film of the eye. It converts light rays into electrical signals and transmits them to the brain through the optic nerve. Peripheral vision comes from the sides of the retina. The center region of the retina is called the macula and it contributes to color vision. Most of the vision related problems occur in retina.
2. Choroid: It is a layer of blood vessels, which lies between the retina and sclera (the wall of the eye). Its function is to supply blood to the retina. Abnormal and fragile blood vessels grow in between the retina and choroid damaging the macula.
3. Optic Disk: The spot on the retina where the optic nerve leaves the eye is

called as the optic disk. A blind spot is created due to the lack of sensory cells. Blind spot of one eye is covered by the other eye and vice-a-versa. The brain interpolates the missing information.

A notable effect of DR is that a large amount of glucose coursing through the circulatory system. This is called glycosylation reaction and it occurs between sugar and proteins in vessel walls. Organs like the eyes and the kidney have microvasculature (the portion of the circulatory system composed of smallest vessels) and are more susceptible to glycosylation.

The effect of DR can be seen distinctly in two stages:

1. Non-proliferative Diabetic Retinopathy (NPDR)
2. Proliferative Diabetic Retinopathy (PDR)

Retinopathy associated with diabetes, progressively characterized by Microaneurysms, intra-retinal punctate hemorrhages, yellow, waxy exudates, cotton-wool patches, and macular edema are the preliminary stage of DR usually labeled as Microaneurysms (MA). This stage is more commonly known as non-proliferate diabetic retinopathy (NPDR) or background retinopathy. In this thesis, NPDR is also interchangeably referred to as MA, for the sake of convenience. Microaneurysms are small spots of swelling in the retina's tiny blood vessels as shown in Fig. 3.

NPDR can be further classified into moderate and severe NPDR. In moderate NPDR, only some blood vessels that drive blood to the retina get blocked. This is observed along with MA. As moderate NPDR progresses into severe NPDR, a

large number of blood vessels get blocked. As a result, several areas of retina are deprived of blood. In addition to microaneurysms; yellow, waxy exudates, cotton-wool patches can be observed. In Fig. 3, yellow and white arrows show exudates and microaneurysms respectively. Fig. 4 shows cotton wool spots indicated by blue arrow.

PDR is the advanced stage of the disease. The important characteristic of PDR is hemorrhage that is severely red in appearance and is shown in Fig. 4 by red arrow. Due to a large number of blood vessels being blocked, there is a need to grow new blood vessels. This process of formation of new blood vessels is called neovascularization (NV). In the remainder of this document PDR and NV will be interchangeably used. Some of the new blood vessels formed due to NV become abnormal and fragile and cause blood leakage. Accumulation of such leaked blood in the retina causes severe vision loss and eventually blindness.

In a normal diagnosis there is a clear absence of retinal landmarks such as microaneurysms, exudates, hemorrhages, and cotton wool spots etc. Fig. 2 shows fundus camera images of a normal eye.

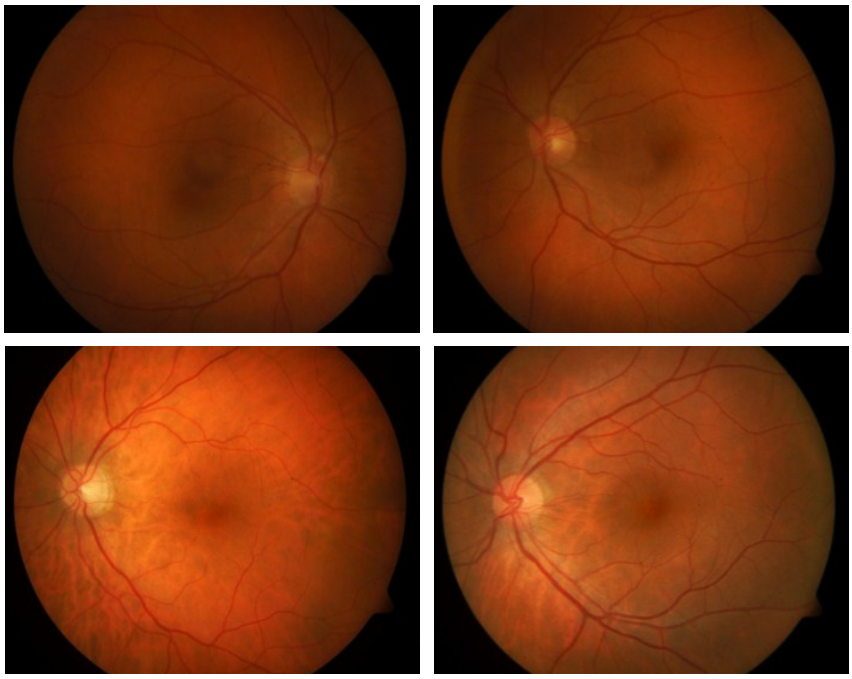


Fig. 2. Fundus Camera Images of a Normal Eye(s)

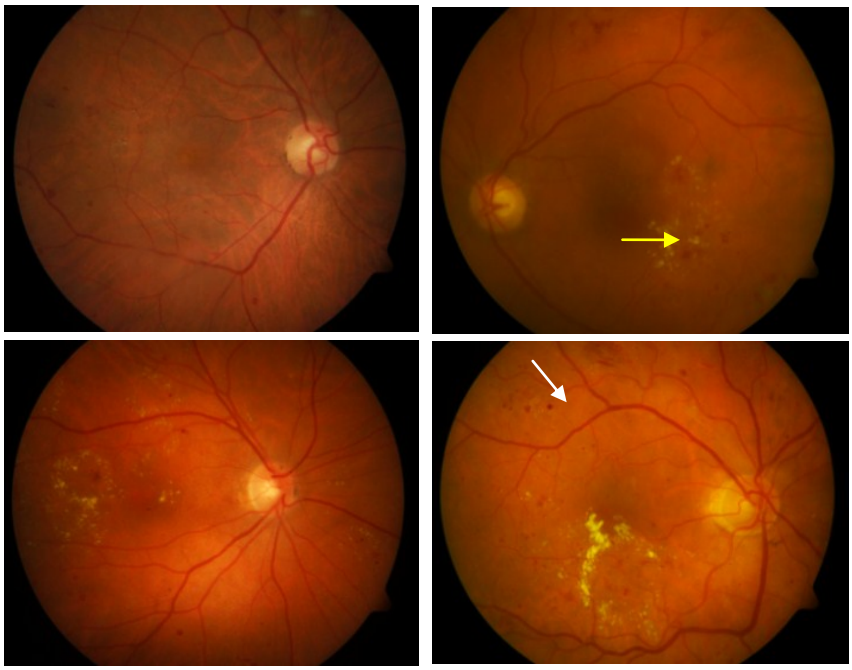


Fig. 3. Fundus Camera Images of an Eye with NPDR

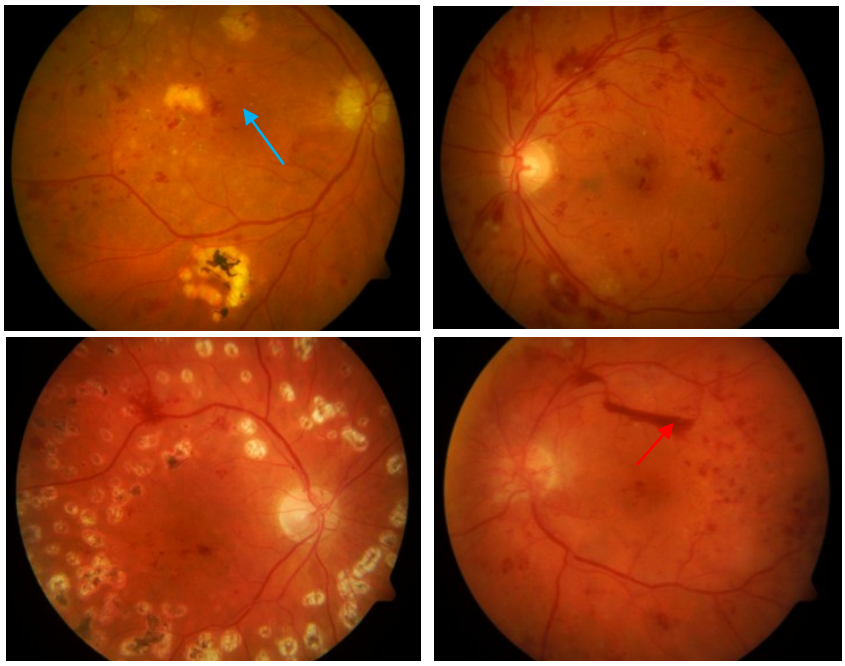


Fig. 4. Fundus Camera Images of an Eye with PDR

**1.2 Motivation**

Diabetic retinopathy (DR) is a common cause of blindness in working populations [2]. A World Health Organization collaborative study had projected that 221 million people around the globe might already have DR by 2010 [3]. A more recent study in August 2010 [4] showed that 28.5% and 4.4% of US adults who have diabetes are also affected with MA and NV respectively. Predictions of the future suggest that DR will become a public health problem [5]. Despite advances in diabetes care, visual impairment still remains a potentially devastating complication. Studies have shown that timely eye examinations and treatment can significantly reduce the risk of severe vision loss [6]. A study conducted by the Diabetes Control and Complications Trial (DCCT) among 1441 patients who were examined for six and a half years, revealed that the use of an

insulin pump lowered the risk of DR by 76% and it also slowed down the progression of DR by 54%. The frequency of occurrence of severe or moderate DR was lowered by 47% [7].

Different ophthalmologists have different expertise in assessing DR. Their performance in evaluating severity of the disease depends on their experience and training. This remains to be the case even if their diagnosis is based on digital retinal imaging, which is becoming a preferred alternative to conventional face-to-face evaluation. This process is laborious and there are many factors like reviewer fatigue or human error that contribute to inaccuracies. This is a motivation to explore efficient and accurate methods of automatic detection and evaluation of DR.

Increasing the accuracy of assessing the DR severity level could greatly reduce vision disabilities. Improved consistency and speed, over human experts are potential benefits of automated DR diagnosis. Previous studies have shown that sensitivity of an automated algorithm is higher than manual graders albeit false detections [8]. Therefore, if these methods were to be improved, it will culminate in reduced human workload and expenditure [9].

Yet there is no automated computer-based system for evaluating DR by incorporating all the experience and the cognitive knowledge that ophthalmologist uses to detect and determine the DR severity level. This thesis is dedicated to exploring a new and effective methodology for classifying and retrieving clinically-relevant images, thereby decreasing blindness due to DR.

### 1.3 Specific Aims

1. To develop an effective approach that classifies the given DR image into one of three classes according to its severity level.
2. To develop a content-based retrieval system that references DR images to assist clinical diagnosis. The system should have a user-friendly interface and intuitive annotation schemes.

### 1.4 Related Work

This section will survey related work. Segmentation of retinal landmarks and localizing the optic disk are important processes in retinal image analysis and therefore are actively researched topics. The optic disk is usually the brightest area on the fundus camera image. This assumption is exploited by several methods that localize optic disk as a cluster of high intensity pixels [10, 11]. These methods fail when there is a presence of exudates that are also small clusters of high intensity pixels. To overcome this problem, use of principal component analysis (PCA) to project a new image into the 'disc space' was proposed in [12]. The optic disk was found by calculating minimum distance between the original and the projected image.

Another way to identify the optic disk is by using the variance of pixel intensities between optic disk and adjacent blood vessels [13]. Though this method achieved very high success rate it often fails for images with a large number of white lesions due to the large variance of pixel intensities between the white lesions and the blood vessels. Probabilistic Hough transform was employed

by [14]. This method fails when optic disk is not perfectly circular.

An important observation from a retinal image is that, all the blood vessels converge on the optic disk. This fact was exploited by [15, 16]. Another method is to describe the shape of optic disk by a global model. This method achieves 98.89% success rate [17].

Motivated to identify MA, approaches are developed using morphological processing in threshold fundus images [18, 19, 20, 21]. Since MA and hemorrhages are distinct from the other portions of the fundus image, a distance-based approach can be used to differentiate among these retinal landmarks. Mahalanobis classifier is used to achieve sensitivity of 69% and 83% in detecting MA and hemorrhages respectively [22].

Neural networks along with region-growing segmentation is also a popular method to detect MA and hemorrhages [23, 24]. Similar techniques were also used to detect exudates [23, 25, 26]. The reported sensitivity for exudate detection ranges from 93%-99% whereas the reported specificity ranges from 80%-94%. A detailed analysis about retinal images and a more exhaustive survey of the literature in retinal landmark detection can be found in [27]. Though the prior mentioned approaches try to identify DR symptoms, there is no method in literature that assesses DR severity level. Such an assessment offers the necessary clinical help to ophthalmologists.

Several content-based image retrieval (CBIR) systems for non-DR applications have been developed in the past. A system retrieving Positron

Emission Tomography images based on physiological kinetic features has been developed previously [28]. A semantic model for CBIR, which captures spatial and temporal semantics of neural image databases, was developed in [29]. Use of texture and histograms was proposed to retrieve CT scan images [30]. A method to retrieve medical images having tumors was proposed in [31]. The approach defines a distance function that retrieves similar shapes. It also penalizes shapes that are different. CBIR of brain neuroscience images using operational semantics was performed in [32]. ASSERT is a system that uses computer vision algorithms to extract visual abnormalities in high-resolution computer tomography (HRCT) lung images [33].

These initial CBIR works in the medical field are promising since all the papers report either improved performance of the system or benefits to a clinician. But most of the methods are restricted to certain domains (CT, tomography, X-ray) and thus cannot be directly applied to solve DR problem.

Structured Analysis of Retina (STARE) project is aimed at automatic diagnosis and comparisons of retinal fundus images. It searches for images that are similar in content [34, 35]. A prototype for CBIR of ophthalmological images was later built in 1996 [36]. This prototype proposed a similarity metric based on basic image features called primitives. A more recent CBIR prototype was built for automated diagnosis of retinopathy [37].

Though these projects have made a significant progress in automated DR diagnosis, there is a lack of a unified solution that has been unanimously accepted

by ophthalmological community. Though recent works published in [37, 38] are encouraging, there is no existing system that considers clinical relevance and there lies the contribution of this thesis.

### 1.5 Contributions

The contribution of this thesis is two-fold.

1. A novel approach is proposed to perform multi-class classification of DR images.
2. A computer system with user-friendly interface is built to perform DR image retrieval.

The classification is based on spectrally tuned color correlogram and Citation-KNN. For performing CBIR, spectrally tuned color correlogram and steerable Gaussian filters are used along with a distance based approach. A distance framework, "Rank-KNN" is designed. It is capable of performing effective multi-class multiple-instance retrieval. Rank-KNN shows considerable increase over the other state-of-art retrieval frameworks.

## 2. SPECTRALLY TUNED COLOR CORRELOGRAM FEATURES

The color spectrum of DR images is unique. Color distribution is the distinguishing characteristic among normal and affected images. Thus color features are the best feature descriptors for DR images. In the following section, the state-of-art color features are reviewed.

### 2.1 State-of-Art Color Features

#### 2.1.1. Color Histograms

The color histogram is the count of the occurrences of each color in an image [39]. The colors in the image are quantized for the sake of computational efficiency. Histograms are invariant to translation and rotation. Though color histogram is very simple to implement and computationally inexpensive. It has many disadvantages:

1. Sensitive to pixel intensity variations and color quantization errors.
2. High dimensionality as the number of bins needs to be sufficiently large.
3. No consideration for the color similarity among adjacent color pixels.
4. No incorporation of spatial information.
5. It is a course representation of an image and different images can have same histograms.

To overcome some of the disadvantages mentioned above, a histogram refinement technique imposes further constraints on color histogram based matching [40].

### 2.1.2. Color Coherent Vectors

Color coherence vector (CCV) is a more advanced form of histogram refinement in which the pixels within a particular color bin are first classified as coherent or incoherent. Coherent pixel is a part of large group of pixels that fall in the same color bin while an incoherent pixel can be seen as an isolated pixel. Pixel groups are determined by finding connected components and if the size of that pixel group exceeds a specific value then the pixel lying in that group is coherent. If the number of coherent and incoherent pixels in $j^{th}$ bin are $\alpha_j, \beta_j$ respectively, then CCV of an entire image consisting of $n$ color bins can be written as, $\langle(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_n, \beta_n)\rangle$. CCVs have some distinct advantages over color histograms:

1. CCVs account for spatial coherence of colors in an image since it checks for connected regions in an image.
2. Fine distinctions can be made between different CCVs because of additional constraints which is not possible in case of color histograms.

In spite of this, CCVs fail to catch the global distribution of colors.

### 2.1.3. Fuzzy Color Histograms

Color histogram does not consider the color dissimilarity in different bins or color similarity in same bin. Therefore to overcome these disadvantages, [41] proposed a method called fuzzy color histogram (FCH) which considers color similarity of the color of every pixel associated to all the histogram bins through membership values. The FCH of an image I defined as,

$FCH(I) = [fch_1, fch_2, \ldots, fch_k]$ where $fch_i = \frac{1}{N}\sum_{j=1}^{N} \mu_{ij}$ , where N is the total number of pixels in the image and $\mu_{ij}$ is the membership weight associated to $j^{th}$ pixel to the $i^{th}$ color bin and it is given by $\mu_{ij} = \frac{1}{1+\frac{d_{ij}}{\gamma}}$ , where $d_{ij}$ is the Euclidean distance between the actual pixel color and the color bin associated with it. $\gamma$ is the average distance between the colors in the quantized color space. This method has some advantages over simple color histograms. Due to membership weights, the colors which lie in different bins will be penalized and those which lie in the same bin will be boosted. Thus the method gets an ability to encode color similarity. This method is shown to be robust to quantization noise and contrast changes. But it has some obvious disadvantages, for example, high dimensionality and more computation. The other disadvantage is that it is important to choose appropriate membership weights to get relevant results. Also, FCH does not encode spatial information and the global distribution of colors in an image.

### 2.1.4. Color Correlograms

Color correlogram (CC) is a well-studied feature for CBIR of natural images [42, 1]. CCs are considered superior to color histograms, CCVs etc. A CC of an image as proposed in [42] is a table indexed by color pairs, where the $k^{\text{th}}$ entry for the color pair $(i, j)$ specifies the probability of finding a pixel of color $j$ at a distance $k$ from a pixel of color $i$. Let $I$ be a square image of side $N$. Typically, to reduce the dimensionality of the features, the image is quantized into $m$ color

bins, say $c_1, c_2, \ldots c_m$. Let $C(p)$ denote the color of a pixel $p = (x, y) \in I$. An $L_\infty$-norm is then caluclated to measure the distance between two pixels, say $p_1 = (x_1, y_1),\ p_2 = (x_2, y_2)$, with $|p_1 - p_2| = \max(|x_1 - x_2|, |y_1 - y_2|)$. The histogram of the image $I$ (denoted by $h_i$) with respect to each bin is now calculated. This is equivalent to calculating the number of pairs of pixels, such that $|p_1 - p_2| = k$ and dividing it by the total number of pixels belonging to that particular bin in the entire image. The search neighborhood is limited to $3X3$, therefore $k$ is set to 1. If the distribution of only those pixels is analysed which lie in the same bin $(i = j)$, then the features are called as color auto-correlogram (AutoCC). This achieves dimensionality reduction since the number of bins in an image is also the dimensionality of the feature space as the histogram of correlations of the bins becomes the feature vector. These two constraints reduce the computational load but compromise on image representation.

**2.2 Spectrally Tuned Color Correlogram**

It can be inferred that the performance of the color correlogram features depends on the chosen quantization scheme. In 2007, [1] proposed a quantization scheme modeled after human vision that takes into account the color spectrum of natural images as shown in Fig. 5. The spectrum of DR images is also shown in . A distinct difference could be observed between the spectrums of natural and DR images. It can be observed that DR image spectrum is saturated in red channel and the distinguishing characteristics mainly lie in the green and blue channel. Therefore, histogram equalization is performed in the red channel as part

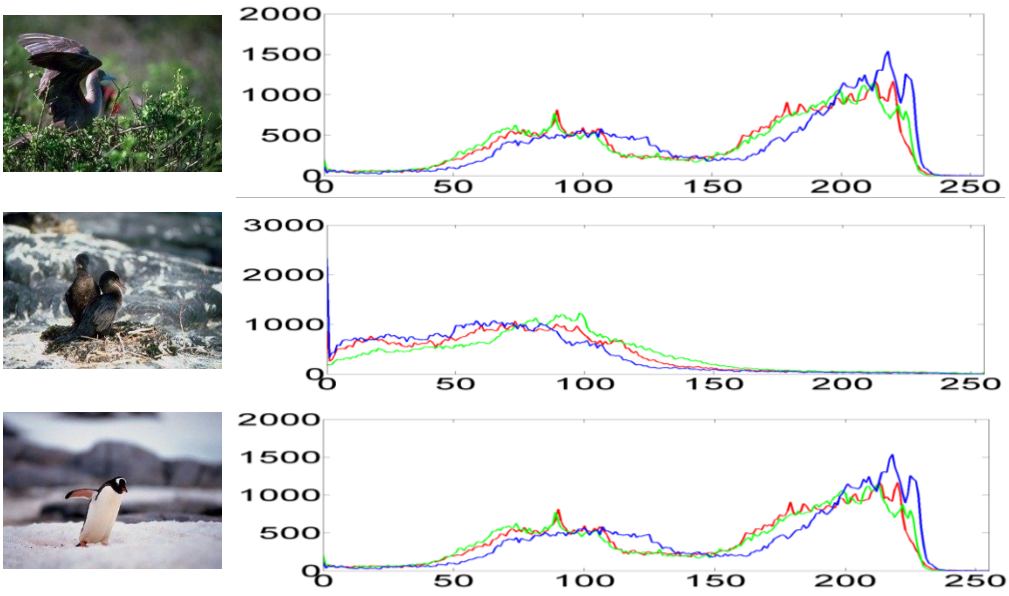



Fig. 5. Histograms of Some Natural Images. The Variation in the Histogram is Clearly Noticed within the Three Sample Images

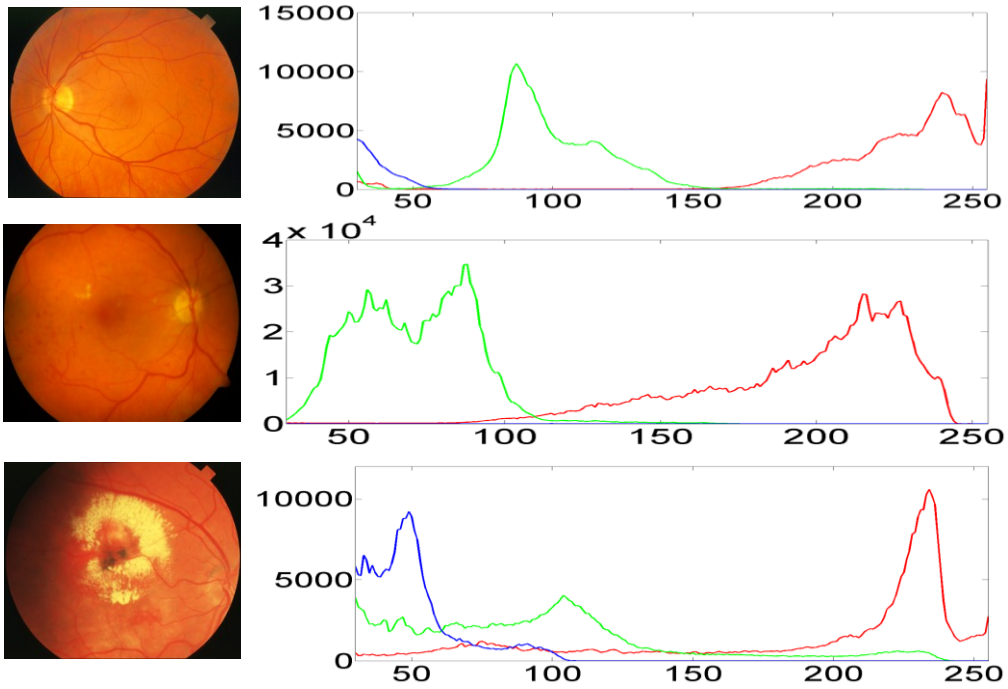



Fig. 6. Histograms of Typical Fundus Images. The Top Image is Diagnosed as a Normal Image, the Middle Image as MA and the Bottom Image NV. It can be Noticed how All the Red Channels are Saturated

of pre-processing. To adapt to unique shade pattern in DR images, a new quantization scheme is proposed which is later explained in detail. This results in a feature set which is spectrally tuned towards DR images. Thus CC features give a better representation of DR images. Following subsections elaborate on the design of quantization scheme and the benefits of spectrally-selective tuning.

### 2.2.1. Quantizer Design

To design an effective quantization scheme for DR images, all the images in the training set are considered and the red channels of all of them are equalized. After preprocessing, the training images are converted into 16-bin non-uniformly quantized one-channel images. The choice of the 16-bins was empirically determined. The procedure of getting 16 bins is described below.

All the unique shades in the training set are extracted. Here a shade implies a $\langle R, G, B \rangle$ triplet. After the extraction process, shades are arranged in a $NX3$ matrix. K-means Clustering is then performed on the above mentioned matrix to divide the color space into 16 clusters. The centroids of 16 clusters are the codewords of quantization bins. The generated codebook is the new quantization scheme which is designed for DR images. CC features are then calculated for every pair of bins $(b_1, b_2)$ where $(b_1, b_2) \in [1,2, \dots ,16]$. A 256-D feature vector is generated which describes the distribution of each bin with respect to all other bins. These features can be called as spectrally-tuned color correlogram features. Fig. 7 and 8 show a visualization of this clustering approach for the quantization scheme proposed in [1] and the new approach respectively.

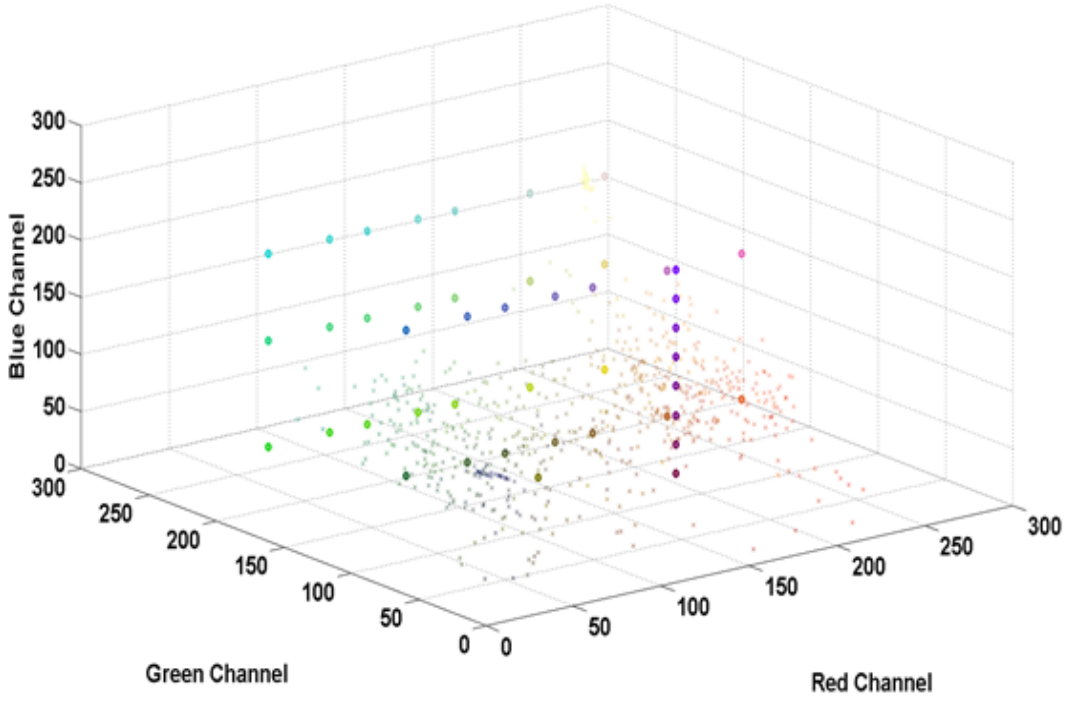



Fig. 7. Quantization of AutoCC Approach Proposed in [1]

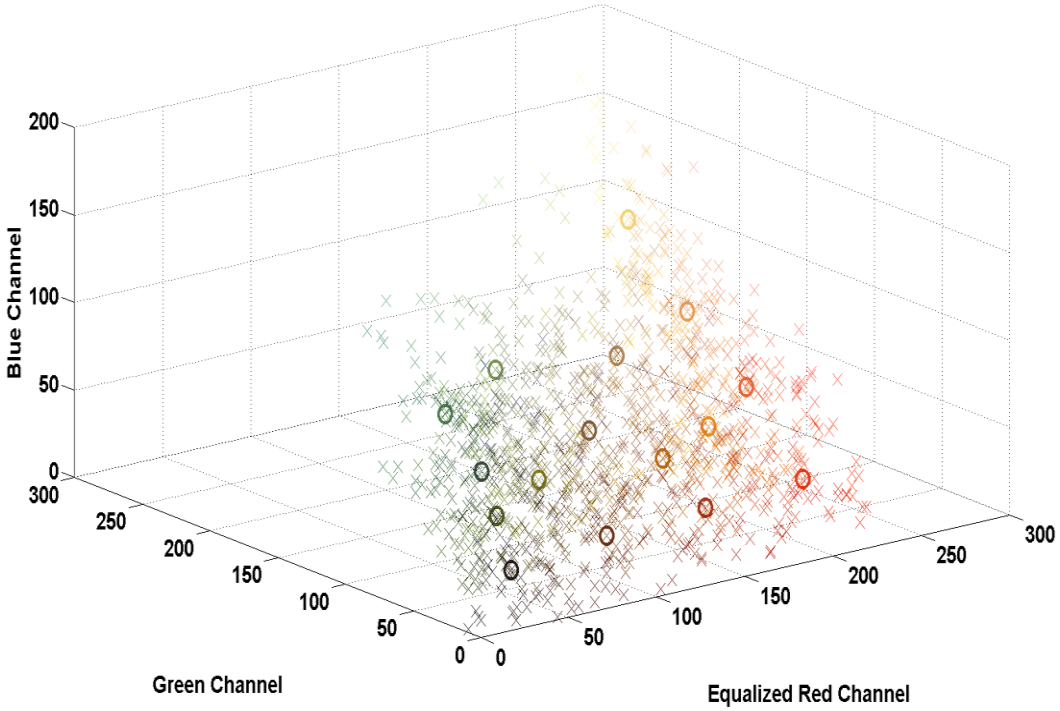



Fig. 8. Visualization of Quantization of the Proposed AutoCC Approach

The approach illustrated in Fig. 7 is designed for natural images and has different number of color shades associated with each centroid while the proposed approach has a uniform density. In Fig. 7, the centroids are uniformly distributed in the color space but densities of associated shades are non-uniform. While in the proposed approach, though the centroids are distributed non-uniformly, the associated shades per centroid are uniform. This utilizes bandwidth provided by 16 color bins to full extent in the proposed 256-D feature space. Fig. 9 and 10 illustrate the effectiveness of the clustering appraoch used for developing the proposed quantization scheme through a histogram of quantized shades in the database from the approach proposed in [1] and the proposed approach respectively.

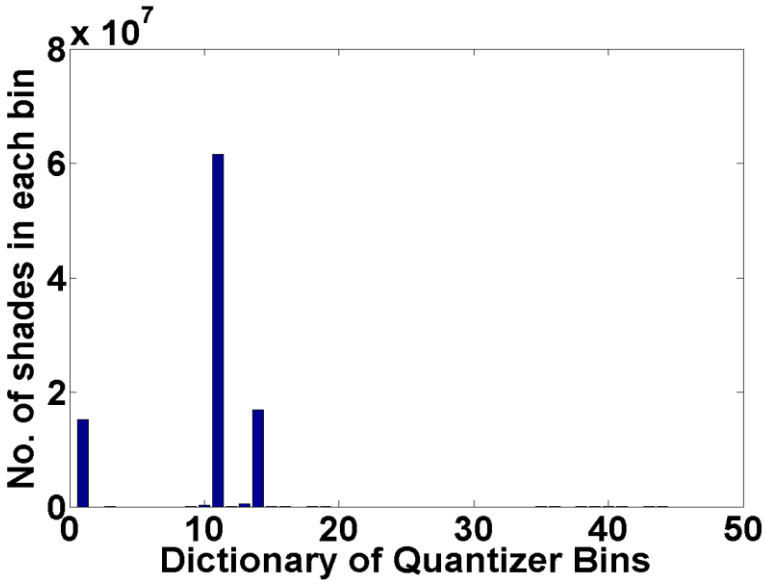



Fig. 9. Visualization of Quantization of AutoCC Approach Proposed in [1]

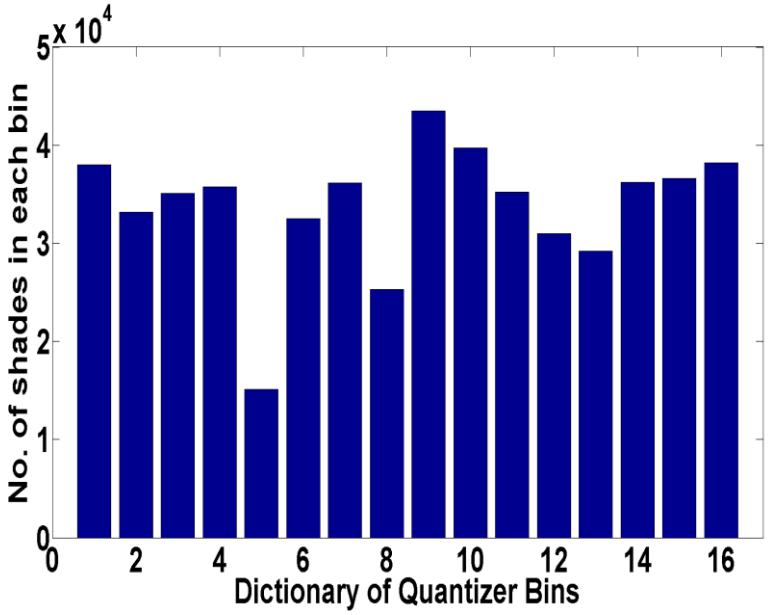



Fig. 10. Visualization of Quantization of the Proposed CC Approach

In Fig. 9, in spite of the uniform placement of bins, most of them do not have any color shades associated with them. But the proposed quantization scheme shown in Fig. 10 ensures even distribution of number of shades over all the bins. Thus in the proposed set of spectrally-tuned CC features, the semantic information is spread evenly along the 16 bins which is a desirable property for a feature space and will be useful for a later stage of learning.

The quantization scheme presented in [1] works for natural image but for a DR image which has a distinct spectrum, it almost always fails. Fig. 11 demonstrates this by showing a natural and a DR image quantized using the original quantization scheme and the same DR image quantized using the proposed approach. It can be clearly seen that the proposed quantization scheme better represents a DR image.

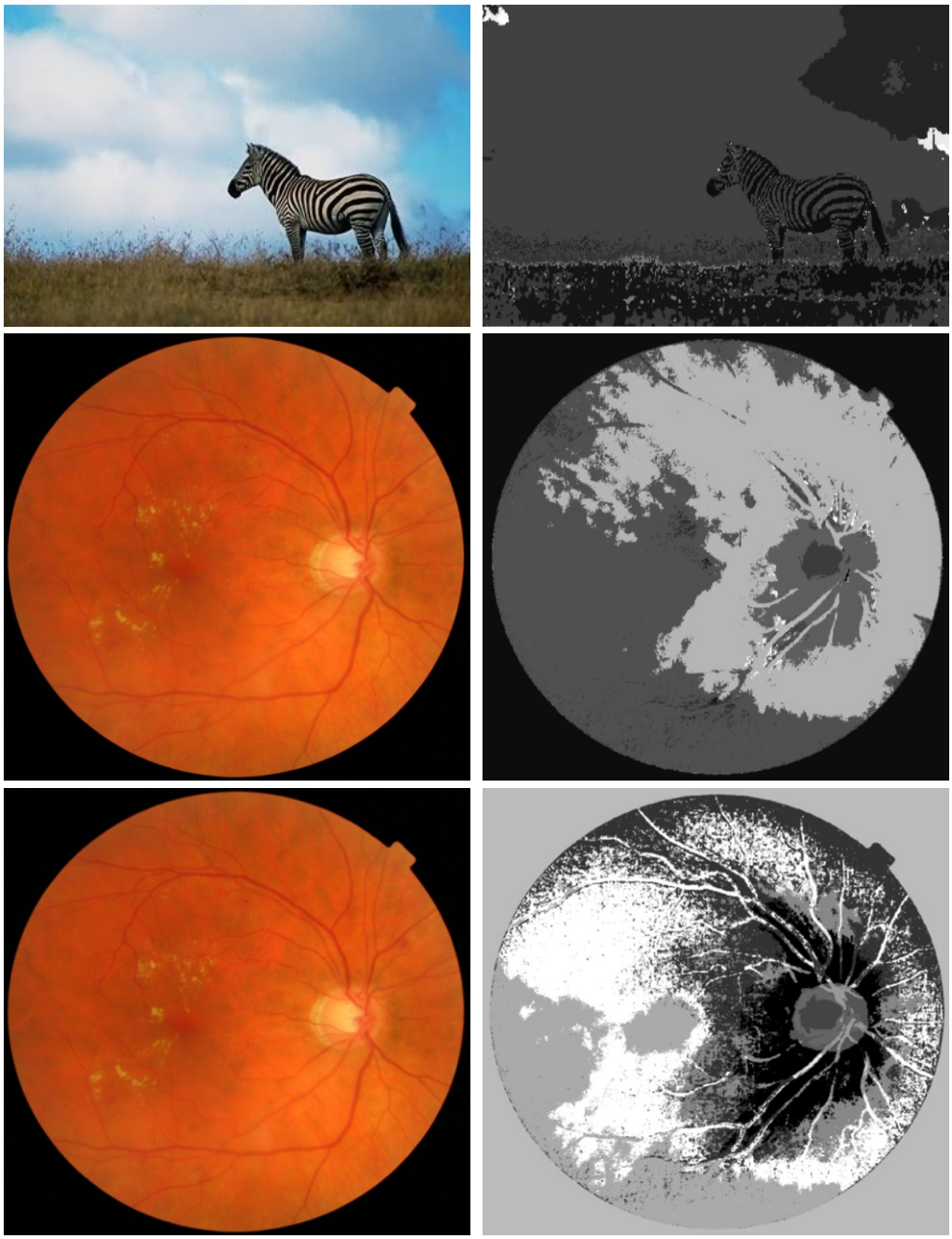


Fig. 11. First Two Rows Show Quantization using the Original Approach [1] and the Last Row Shows the Same using the Proposed Approach.

### 2.2.2. Color Correlogram Feature Extraction

After the image is quantized into 16 colors by using the aforementioned approach, the CC features are extracted on the quantized image. The algorithm for the procedure of extracting CC features is described below. A flow chart representation of the same is also shown.

***Algorithm for Computing CC features of an RGB Image:***

***Step 1.*** Covert the RGB image into a quantized image made of 16 colors by using the proposed approach described previously.

***Step 2.*** Consider a window of 3X3 for a pixel (since $k = 1$). Compare the value of center pixel with the value of all the 8 pixels in the neighborhood. Count the number of pixels for every possible pair of color bins, one bin out of that pair always being the color bin in which the center pixel lies.

***Step 3.*** Maintain separate counts for every possible color bin pair. Thus there will be 256 counts for 256 color bin pairs. Add the count obtained in step 2 to its respective color bin pair. This process is repeated for all the pixels in the image.

***Step 4.*** The vector thus formed represents the local spatial distribution of the pixels. Divide the count of every color bin pair by the count of number of pixels which lie in the first color bin in that pair. This gives the color correlogram of the image. This results in a 256 dimensional feature vector for an image.

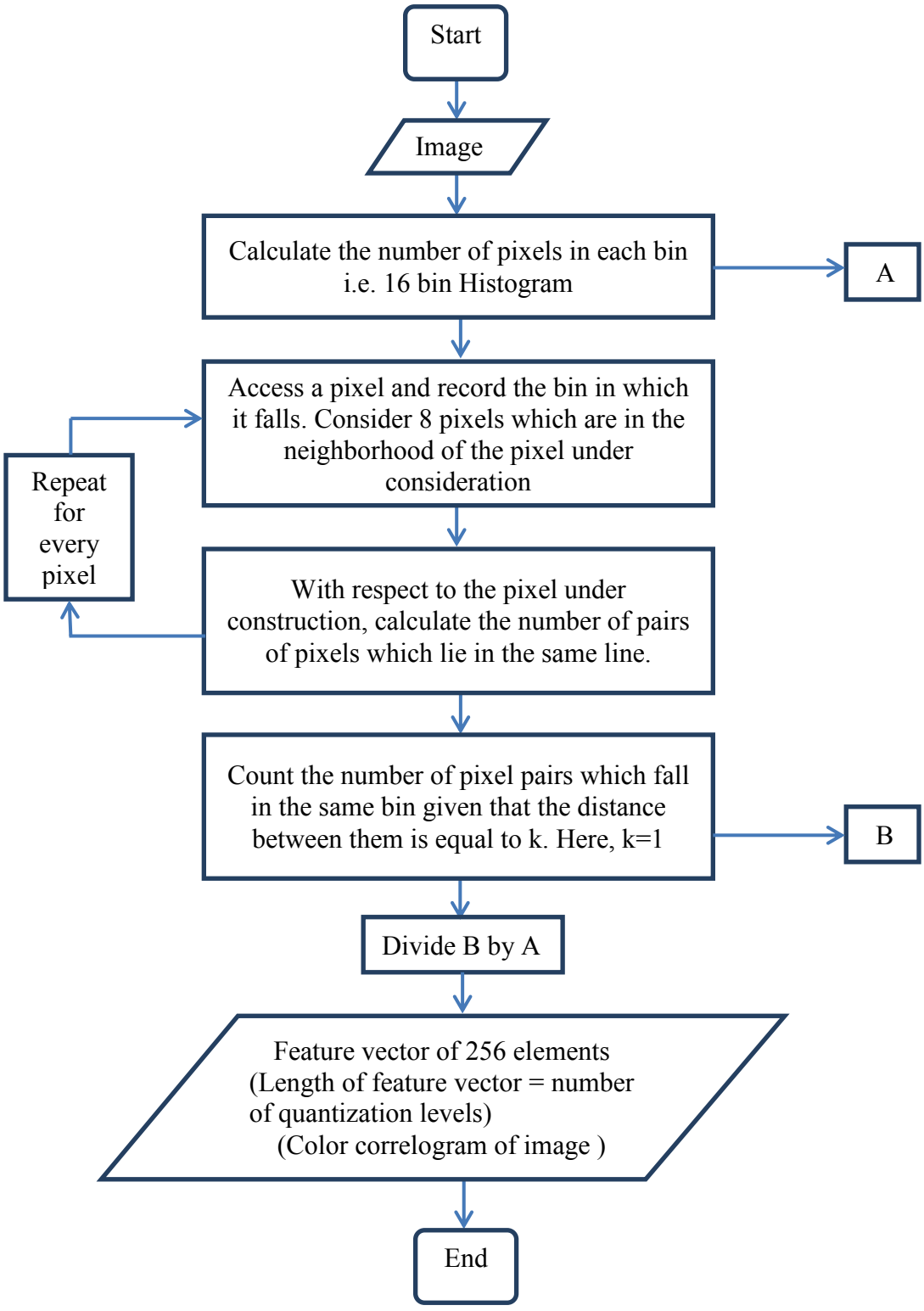



Fig. 12. Flow Chart for Computing Spectrally-tuned CC Features

# 3. CLASSIFICATION OF DIABETIC RETINOPATHY IMAGES

Prevention of blindness due to DR is possible by an accurate and early diagnosis and followed by correct treatment. It is therefore important that each person gets his/her eyes examined at least once per year and take a treatment even if slightly threatening symptoms are detected. Various screening strategies are employed for this purpose, for example, direct examination by ophthalmoscopy, and recently digital imaging. In major hospitals, primary screening is performed by experts and a further examination is conducted if abnormalities are detected. Then an ophthalmologist decides on the treatment pattern. But diabetes affects over 2% of the population [43] and the aforementioned screening process is indeed very costly in monetary as well as personnel terms. Thus automated classification or grading of DR images could result in resource savings and more consistent results. Still there is a lack of unified DR grading system. Therefore in this section, an automated algorithm for DR image classification is proposed which takes a completely distinct approach of machine learning combined with image processing techniques to obtain better classification performance over three categories.

## 3.1 State-of-Art Classification Frameworks

In literature, there are many classification frameworks which are proven to be robust and have usually produced consistent results over all types of data. In the following section, the superiority of proposed classification framework is shown against well-established classification frameworks.

### 3.1.1. Color Histograms

Color histogram is a simple, fast and low-level method. It shows the global distribution of the color in an image. It has shown satisfactory results in the past in the area of image indexing and retrieval [44]. The simplicity and speed of feature extraction makes this method suitable for tasks similar to image retrieval.

A color is represented by a three dimensional vector where each dimension corresponds to a color component. In order to construct a color histogram, it is necessary to perform quantization on the image. Therefore, choices of color space and quantization steps are the important factors to be computed. HSV is one of the most widely used color space in literature. It is considered effective and suitable since it separates hue and saturation from the luminance component and it is less sensitive to illumination changes. However, HSV does not produce great increase in the results because after performing quantizing image into certain number of bins, no information about the color space is utilized by the classifier.

Color histogram with support vector machine (SVM) as the classifier is used natural image classification in [45]. This method achieves 16.3% error rate on Corel image dataset. This shows that color histogram is a simple and effective technique for image classification.

### 3.1.2. Bag of Words (BoW) Approach

A Bag-of-words (BoW) approach was intorduced in [46]. Three main steps are involved in converting an image into bag of words. They are,

1. Feature detection
2. Feature description
3. Feature quantization

In the first stage, differene-of-Gaussian (DoG) points are detected as the distinctive image features. Approximately 100-500 image regions that are stable and rotationally invariant over number of scales are extracted using scale-invariant feature transform (SIFT) [47]. The 128-dimensional SIFT descriptors of the image pathches are further encoded into a visual word through a learned codebook. Image classification can then be performed on such a feature space by employing classifiers such as support vector machines (SVM). SIFT features coupled with a bag-of-words (BoW) approach has been shown to be very effective.

### 3.1.3. Gabor Texture Features

Out of many texture features, Gabor features are seen to be similar to human visual system, and they have been found to be suitable for texture segmentation and classification [48]. Also, they have been successfully used for browsing and retrieval of images [49]. Two-dimensional Gabor filters are appropriate for image texture analysis because of many reasons:

1. They have tunable orientation frequencies and center frequencies.
2. The demodulated Gabor frequency envelops usually contain low spatial frequencies and by comparing the frequency envelopes, the boundaries between textured regions differing significantly can be detected by the

change in spatial frequency content.

3. Discontinuities arising from the surface can be detected by observing large variations in the phase envelopes.

This has been explained in greater detail in [48]. This shows that Gabor features combined with a suitable classifier can be considered as a good approach.

### 3.1.4. Histogram of Color Moments

Color histogram only expresses the number of pixels of each color and hence gives the global distribution of color in an image. But it fails to illustrate the spatial distribution of the color. Therefore, histogram of neighborhood color moments (HNM) was proposed in [50]. Firstly, image is quantized in HSV color space. To consider the spatial relationship between the pixels, $3 \times 3$ neighborhood of each pixel in the quantized image is considered and then the first three low-order central moments are calculated. They can be calculated as,

$$\mu_{ij}^{3} = \frac{1}{9} \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} q_{mn} \tag{1}$$

$$\sigma_{ij}{}^{3} = \sqrt{\frac{1}{9} \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} (q_{mn} - \mu_{ij}{}^{3})^{2}} \tag{2}$$

$$S_{ij}^{3} = \sqrt[3]{\frac{1}{9} \sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} (q_{mn} - \mu_{ij}{}^{3})^{3}} \tag{3},$$

here $(i, j)$ are pixel co-ordinates and $q_{mn}$ is the pixel value at position $(m, n)$. It can be inferred that all three moments lie in the range of $[0,255]$ Thus,

there are 3 histograms of neighborhood low-order central moments as features. After normalizing, these features can be used along with a classifier to perform image classification.

### 3.2 Proposed Classification Framework

All the classification frameworks discussed previously cannot achieve good results when applied to DR problem. In the following section, the reasons of the same will be discussed and a classification framework is proposed using the set of spectrally tuned features mentioned in Spectrally Tuned Color Correlogram.

Color histograms features describe only the global distribution of color and not the spatial correlation of colors. This implies that multiple images can have same color histogram and therefore this feature does not distinguish well enough between different categories of images. Also, the predefined quantization steps may be inappropriate for representing DR images as they are predominantly red in color unlike any other natural image.

"SIFT+BoW+SVM" is generally one of the many existing powerful classification approaches, but it fares poorly in the problem of DR image classification. Analysis shows that SIFT mainly captures the optic disk, blood vessels and other edges of the images. It indeed captures some key points for exudates and cotton wool spots because of their distinctive yellowish shades. On the other hand, it hardly captures the subtle red and dark red shades of microaneurysms and hemorrhages which are abundantly present in NV images. This can be seen in Fig. 13, 14 and 15.

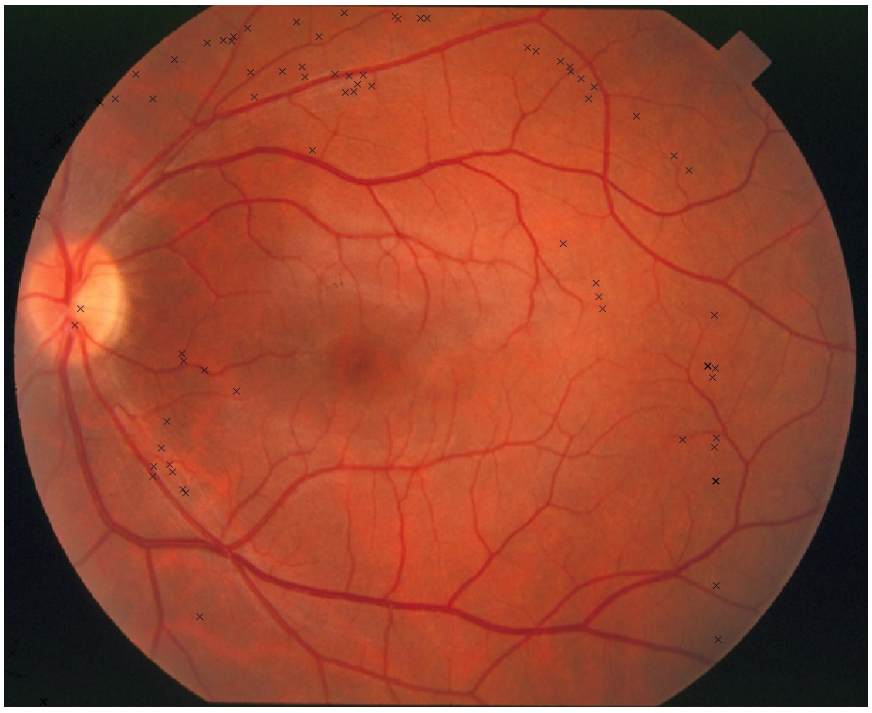


Fig. 13. Key Points in a Normal Image

Fig. 14. Key points in an MA Image

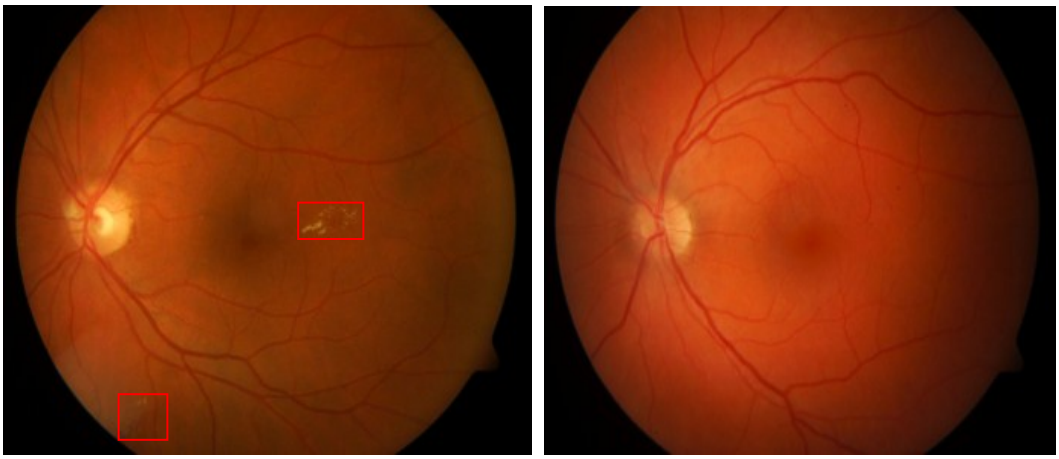
Fig. 15. Key Points in an NV Image

Also, certain unique pattern of objects might get missed by the BoW approach because it appeared for the first time in the test set and was never before seen in the training set. This leads to confusion and it reduces the classification accuracy.

Gabor filters provide are known to provide good textural analysis of an image. As said before, the Gabor filters are known to be similar to human visual system. As the texture of normal and affected images is different in small localized regions, so using textural analysis on a global level might not give better results. In case of MA versus NV, this difference becomes even smaller and hence classifying them becomes difficult. For Gabor feature extraction, first an image is divided into blocks (usually $128 \times 128$) and then a Gabor filter is applied on each image block. The statistics of Gabor filter response are taken as features.

Histogram of neighborhood mean moments provides global distribution of color and spatial distribution as well to some extent. But it fails to model the spatial correlation of the color in the image. Its second drawback is that the quantization steps are predefined and are not suitable for a particular class of images. In the literature, usually HSV color space is used and each color component is divided into 16 bins giving $16^3 = 4096$ bins in total.

The proposed classification framework uses the spectrally tuned color correlogram features as it offers many advantages over other methods as discussed in Color Correlograms. Now, it is necessary to design a learning framework in order to perform effective three-category classification. Multiple class multiple instance learning (MCMIL) approach is suitable for this problem and it is expected that this approach will provide better results than training the classifier with a single image. MCMIL is preferred over the conventional learning approach because of the following reasons. The symptoms which affect the class label of an image lie only in small localized regions of an image. In the MIL framework, only a small set of instances in a bag determines the class of the entire bag. This can be observed in Fig. 16(a). The instances marked in red are the portions of the image that characterize the image as an affected image, while all other instances resemble instances in a normal image. If that instance is called as the characterizing instance then the problem now becomes to identify the characterizing instance(s) in an image.



(a) (b)

Fig. 16. Visualizing the Necessity of MIL Framework (a) MA Image, Instances Marked in Red are the MA Instances (b) Normal Image

A bag has multiple instances in the MIL formulation and thus it corresponds to multiple points in the instance space. In accordance with this, the image is divided into 64 blocks on a $8X8$ grid. Instances that contain majority of black background regions are dropped by thresholding.

### 3.2.1. Review of Multiple Instance Learning Algorithms

The term multiple instance learning was first introduced during an attempt to solve the problem of prediction of drug molecule activity level. First method which tackles this problem was learning axis-parallel concepts in 1997 [51]. After that, many methods have been proposed and their performance has been evaluated on the drug molecule activity prediction problem, for example, diverse density [52], expectation maximization diverse density (EMDD) [53] and citation-KNN [54]. Along with the drug molecule activity problem, these methods have also been applied to content-based image retrieval [55], text categorization [56].

### 1. Learning Axis Parallel Rectangles

Learning axis-parallel concepts [51] is the first algorithm which was aimed at solving multiple instance problems. This method proposes to find an axis-parallel hyper-rectangle (APR) in the n-dimensional feature space to represent the target. This APR should contain at least one instance from each positive bag and at the same time the rectangle should exclude all the instances from negative bags. To find such a hyper-rectangle, three algorithms were proposed in [51]:

1. In the “standard” algorithm, the aim is to find the smallest APR that contains all the instances from positive bags.
2. In the "outside-in" algorithm, first a smallest APR is constructed that contains all the instances in positive bags and then that rectangle is shrunk to exclude false positives.
3. An "inside-out" algorithm starts from an initial point and then keeps on growing a rectangle from it until it finds the smallest APR that contains at least one instance from every positive bag and no instances from negative bags.

The performances of three algorithms are evaluated on one artificial and two real data sets for drug molecule activity prediction problem, and the "inside-out" algorithm is proved to be the most effective one. It correctly predicts the activity of 93.4% drug molecules in the first set and 89.2% in the second set, where the second is judged to be relatively difficult by human experts. All the multiple instance algorithms will be evaluated on these same datasets.

## 2. Diverse Density

Diverse Density (DD) was proposed in [52] and it was called as a general framework for solving MIL problems. The aim of this approach is to find a concept point in the feature space such that the distance of that concept point and at least one instance from every positive bag is less and the distances from the instances in all the negative bags are maximized. The optimal concept point can be defined as the one with the maximum diverse density, which measures the number of instances in the positive bags near the point, and how many negative instances are farther away from that point.

Derivation of DD from probabilistic point of view is presented below. Suppose positive bags are denoted by $B_i^+$ and the $j^{th}$ instance in that bag is represented by $B_{ij}^+$. Each instance has a feature vector and let the $k^{th}$ feature of instance $B_{ij}^+$ be represented by $B_{ijk}^+$. Similarly $B_i^-$, $B_{ij}^-$, $B_{ijk}^-$ represent the above quantities in terms of negative bags. The true concept point $t$ is defined by maximizing the DD which is defined as $DD(t) = P(t|B_1^+, \dots, B_n^+, B_1^-, \dots, B_m^-)$ over the entire feature space. Assuming uniform prior over the concept point location and by using Bayes rule, it can be shown to be equivalent to maximizing the following likelihood:

$$\arg\max_t \prod_i P(B_i^+|t) \prod_i P(B_i^-|t) \quad (3).$$

By using Bayes rule once again with the uniform prior assumption, it gets converted to,

$$\arg\max_{t} \prod_{i} P(t|B_i^+) \prod_{i} P(t|B_i^-) \quad (4).$$

Now, the label of the bag is a combination of labels of its instances. This can be modeled as a "logical-OR" operation carried over all of its instances. Since all the instances in a bag have to have the label 0 for that to be a negative bag, so it can be written as,

$$P(t|B_i^-) = \prod_{j} \left(1 - P\left(B_{ij}^-\right)\right) \quad (5)$$

$$P(t|B_i^+) = 1 - \prod_{j} \left(1 - P\left(B_{ij}^+\right)\right) \quad (6).$$

In the last step, $P(t|B_i^+)$ or $P(t|B_i^-)$ is estimated by a Gaussian-like distribution.

$$P(t|B_i^+) = \exp\left(-\left\|B_{ij}^+ - t\right\|^2\right) = \exp\left(-\sum_{k} w_k \left(B_{ijk}^+ - t_k\right)^2\right) \quad (7),$$

where $w_k$ is a non-negative scaling factor that weights the features In DD algorithm, gradient ascent method is used to search the concept point in the feature space which has the maximum DD. The instances in every positive bag are taken as initial points and the search is repeated.

DD has been effectively applied to solve various MIL problems. Along with the drug activity problem [52], it has been applied to the problem of natural scene classification [57] and image retrieval [58, 55].

### 3. Expectation-Maximization Diverse Density (EM-DD)

In the MIL problem, "most positive" instance in the bag determines the label of the bag. In other words, the instance which is has the highest probability of

being the positive instance in that bag. But the problem comes from the lack of knowledge of the instance which is the most likely one to be the positive instance. In [53], a method was proposed based on DD which uses a set of hidden variables to identify the instance which determines the label of the bag. The hidden variables are estimated using an approach called expectation maximization. This results in an EMDD algorithm which combines the expectation maximization approach with DD algorithm.

EMDD starts with an initial guess of the target point $t$. The initial target point can be achieved using the original DD algorithm described before. Then, the algorithm iteratively carries out two steps:

1. The current hypothesis of target point $t$ is used to select the most probable positive instance from each bag using a given generative model.
2. Now, the new target point $t'$ is estimated by maximizing a DD defined on the instances which have been selected in step 1 by using gradient search. Then, the old target point $t$ is replaced by the new target point $t'$.

The above two steps are repeated until the algorithm converges. EM-DD has been applied to the drug activity problem and it outperforms algorithms like DD, Citation-KNN [54] and APR [51] by a significant margin.

**4. Citation-KNN**

The widely used k-nearest neighbor approach can be used for MIL problems if it is possible to define the distance between the bags. [54] proposes the use of minimum Hausdorff distance to calculate distance between bags by using the

distance between their respective instances. It is defined as the shortest distance between two instances from each bag. Mathematically, it can be written as,

$$Dist(A,B) = \underset{\substack{1 \le i \le m \\ 1 \le j \le n}}{Min}\left(Dist(a_i, b_j)\right) = \underset{a \in A}{Min}\ \underset{b \in B}{Min} \|a - b\| \tag{8}$$

here, $(A, B)$ denotes bags and $(a, b)$ denotes instances in bags A and B respectively. $Dist(\cdot,\cdot)$ can be any distance, for example, euclidean, correlation, cosine, hamming etc. Bag A has m instances and bag B has n instances. Using this bag-level distance metric, it is possible to predict the label of an unlabeled bag using KNN algorithm.

In MIL problem, many times *k*-nearest neighbors do not give information about the true label of the bag. The reason of this can be the false positive instances in each bag. The KNN algorithm can easily get confused because of its majority voting scheme. To overcome this problem, use of references and citers was proposed which gave competitive results. The nearest neighbors of bag B are called as the references and the bags which consider bag B as their neighbor are called as citers of B. This notion of references and citers came from library and information sciences [59]. If a research paper cites another previous research paper (known as its reference), the paper is said to be related to the reference. Analogously, if a paper is cited by an article (known as its citer), the paper is also seen to be related to its citer. Therefore, for a given paper, reference and citer serve as related documents. The following example taken from [54] clearly explains the concept of references and citers. Suppose there are 6 bags of

instances $\{b_1, b_2, b_3, b_4, b_5, b_6\}$. Their 2 nearest neighbors as well as 2 nearest citers are $\{b_3, b_2\}$and $\{b_2, b_3, b_5\}$ respectively. The bold letters indicate the citers of $b_1$ in the Table 1. N indicates the nearest rank number.

Table 1. Nearest Neighbors of 6 Bags $\{b_1, b_2, b_3, b_4, b_5, b_6\}$.

| | $N = 1$ | $N = 2$ | $N = 3$ | $N = 4$ | $N = 5$ |
|---|---|---|---|---|---|
| $b_1$ | $b_3$ | $b_2$ | $b_5$ | $b_4$ | $b_6$ |
| $b_2$ | $\boldsymbol{b_1}$ | $b_4$ | $b_5$ | $b_3$ | $b_6$ |
| $b_3$ | $b_5$ | $\boldsymbol{b_1}$ | $b_2$ | $b_6$ | $b_4$ |
| $b_4$ | $b_6$ | $b_2$ | $b_1$ | $b_3$ | $b_5$ |
| $b_5$ | $\boldsymbol{b_1}$ | $b_2$ | $b_3$ | $b_6$ | $b_4$ |
| $b_6$ | $b_4$ | $b_3$ | $b_1$ | $b_2$ | $b_5$ |

Now it is important to combine the information from R-nearest references and C-nearest citers of an unseen test bag $b$ to infer its class. Assume that out of R-nearest references, there are $R_p$ and $R_n$ positive and negative bags respectively for the test bag $b$ under the consideration; and out of C-nearest citers, there are $C_p$ and $C_n$ positive and negative bags respectively. Let $p = R_p + C_p$ and $n = R_n + C_n$. If $p > n$, then the class of bag $b$ is predicted as positive. Otherwise, then it is set to negative, even in case of a tie. The idea of introducing citers has proved to be more robust that simple KNN which is only based on references. Another alternative introduced in this paper is the Bayesian method. It computes the posterior probabilities of the label of an unseen bag based on the

labels of its neighbors. This method has been tested on the drug molecule activity prediction problem and it showed comparative performance to DD and APR algorithm. It should be noted that citation-KNN is unable to predict the labels of the instances unlike the DD method.

## 5. Support Vector Machine (SVM) for MIL

In 2002, two approaches were proposed in [56] to modify the SVMs in order to tackle the MIL problem.

1. mi-SVM: It is used for instance level classification.
2. MI-SVM: It is used for bag level classification.

mi-SVM treats the instance labels $y_i$ as latent variables which are subjected to certain constraints defined by their bag labels $Y_I$. The ultimate goal is to maximize the instance margin jointly over the unknown instance labels. Also, a linear or kernelized discriminant function is to be maximized, given below:

mi-SVM
$$\min_{\{y_i\}} \min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i \tag{9}$$

$$s.t.\ \forall i: y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, \xi_i \geq 0, \sum_{i\in I} \frac{y_i + 1}{2} \geq 1, \forall I\ s.t.\ Y_I = 1,$$

$$and\ y_i = -1, \forall I\ s.t.\ Y_I = -1$$

where the second line of the constraint enforces the relation between the instance labels and bag labels by making use of the fact that all instance labels have to be negative in order to have a negative label for that bag. In comparison, MI-SVM tries to maximize the margin between two bags, which can be defined as the

margin of "most positive" or "least negative" instance in case of positive or negative bags respectively. It is given as,

MI-SVM
$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{I}\xi_I \tag{10}$$

$$s.t.\ \forall I: Y_I \max_{i \in I}(\langle w, x_i\rangle + b) \geq 1 - \xi_I, \xi_I \geq 0$$

It should be noted that, in mi-SVM the margin of each instance is taken into account. Therefore, one can set the instance labels as long as the constraint between the instance labels and their bag labels is followed. In MI-SVM, the margin of the bag is determined and therefore only one instance per bag matters. The mi-SVM technique is appropriate for the applications where the user also cares about instance labels. MI-SVM technique can be used where only bag labels have to be predicted correctly.

### 3.2.2 Classification using Proposed Approach

As mentioned earlier, every image is cropped to ensure that it maximally fits a rectangular bounding box on the circular retinal region of the fundus image. Then AutoCC features are calculated as discussed above. The image is then divided into 64 non-overlapping blocks of equal size. This supports a MIL approach where each block is an instance and the image can be visualized as a bag. Once the features are extracted, they are used to train a MCMIL approach. In this work, the MCMIL toolbox is used which is publicly available at http://www.cs.cmu.edu/~juny/MILL/. Out of all the MIL algorithms, it is found that "citation-KNN" gives the best detection accuracy.

### 3.3 Experimental Results and Analysis

The dataset used to evaluate the proposed set of features consists of 425 images. It has been assembled manually from well-known and publicly available databases including DIARETDB0 [60], DIARETDB1 [61], STARE [62] and Messidor (kindly provided by the Messidor program partners, see http://messidor.crihan.fr). There are 160 normal images, 181 MA images and 84 NV images.

The efficacy of proposed features is evaluated against some prior arts classification frameworks in the medical imaging field like Gabor features [49] and HNM [50]. The powerful SVM classifier is applied on these feature spaces to form robust classification schemes. Furthermore, a state-of-the-art classification scheme in computer vision, SIFT+BoW+SVM, has also been considered since it is been effective in many computer vision tasks. It is worth noting that whenever SVM is used, the implementation is based on the LibSVM toolbox [63]. Finally, it is necessary to provide an analysis that purely evaluates the benefits of the proposed spectrally-turned AutoCC feature. Therefore, the performance of the original AutoCC features in [1] along with the MIL approach is reported in this section.

For better assessment of the performance of these various classification frameworks on the aforementioned dataset, 5-fold cross-validation was used for multiple runs, and then the mean accuracy of classification was computed over all iterations. The results are summarized in Table 2. The proposed approach which

makes use of the proposed feature along with MIL formulation, achieves 87.6% accuracy along with 2.28% standard deviation as shown in Table 2. Thus it outperforms all the competing approaches by a large margin. This suggests that the proposed method is apt for DR image classification. The aforementioned method is originally proposed in a recently published article [64].

Table 2. Mean Accuracy of Various Methods.

| **Approach** | **Mean Accuracy** |
|---|---|
| SIFT + BoW + SVM | 68.22±2.74% |
| Gabor Features + SVM | 70.11±3.76% |
| HNM + SVM | 73.38±2.23% |
| Spectrally-tuned AutoCC + SVM | 79.83±2.36% |
| Original AutoCC+MIL | 78.01±3.94% |
| Proposed Algorithm | 87.61±2.28% |

It is necessary to analyze the feature space in lower dimensions so that the separability of features for different categories can be visualized. This is also called dimensionality reduction. For this purpose, principal component analysis (PCA) is used. Fig. 17, 18 and 19 show such separability in 3 dimensions.

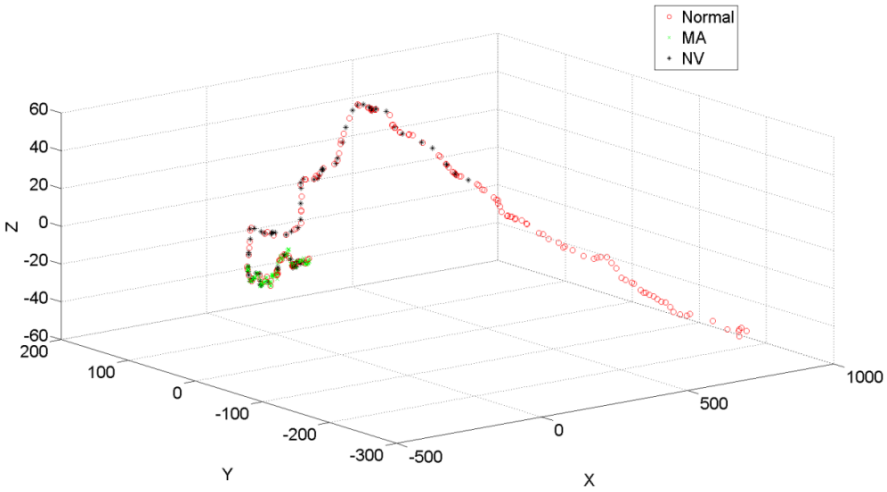



Fig. 17. Separability of SIFT Features

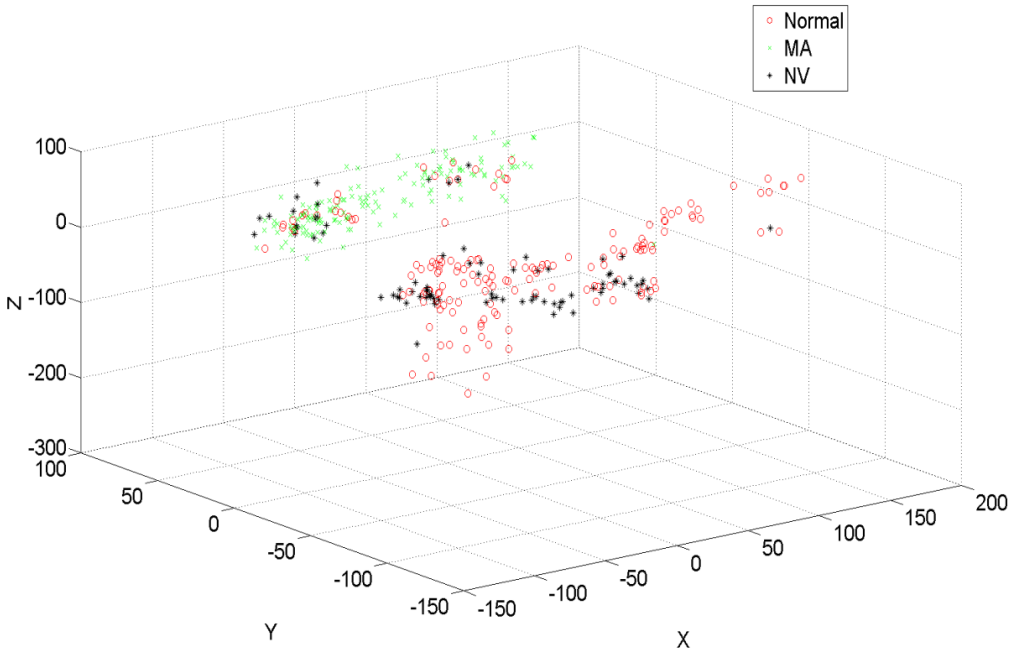



Fig. 18. Separability of Gabor Features

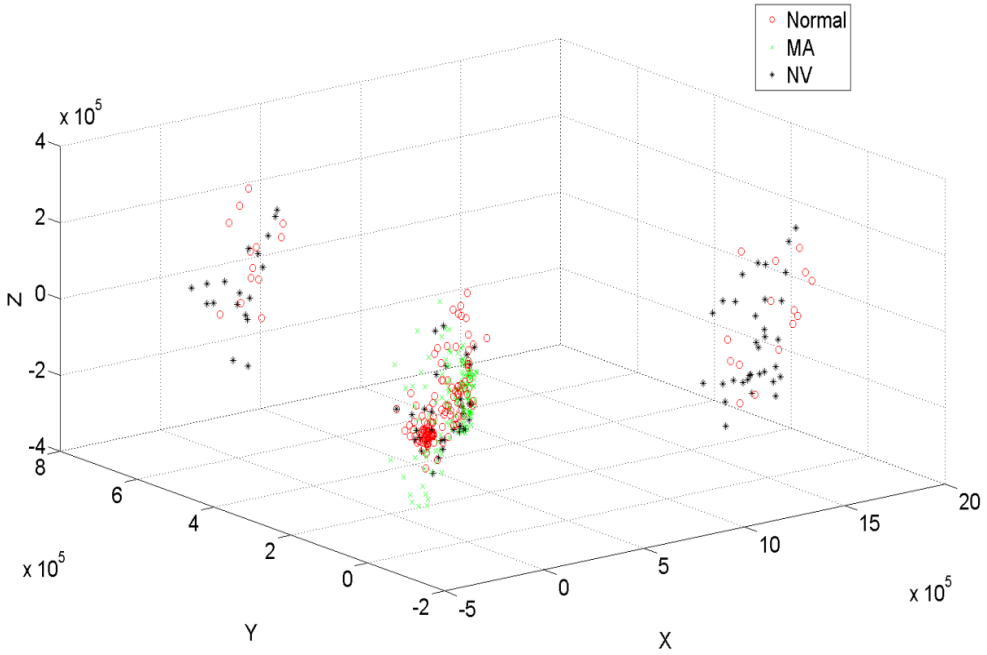



Fig. 19. Separability of HNM Features using PCA

It is necessary to further analyze the distribution of misclassified samples and the classification performance with respect to individual classes. Therefore, the confusion matrices are provided in Table 3 and 4 for the "original AutoCC+MIL" approach and the "spectrally-tuned AutoCC+MIL" approach respectively.

Table 3. Confusion Matrix for Original AutoCC+MIL

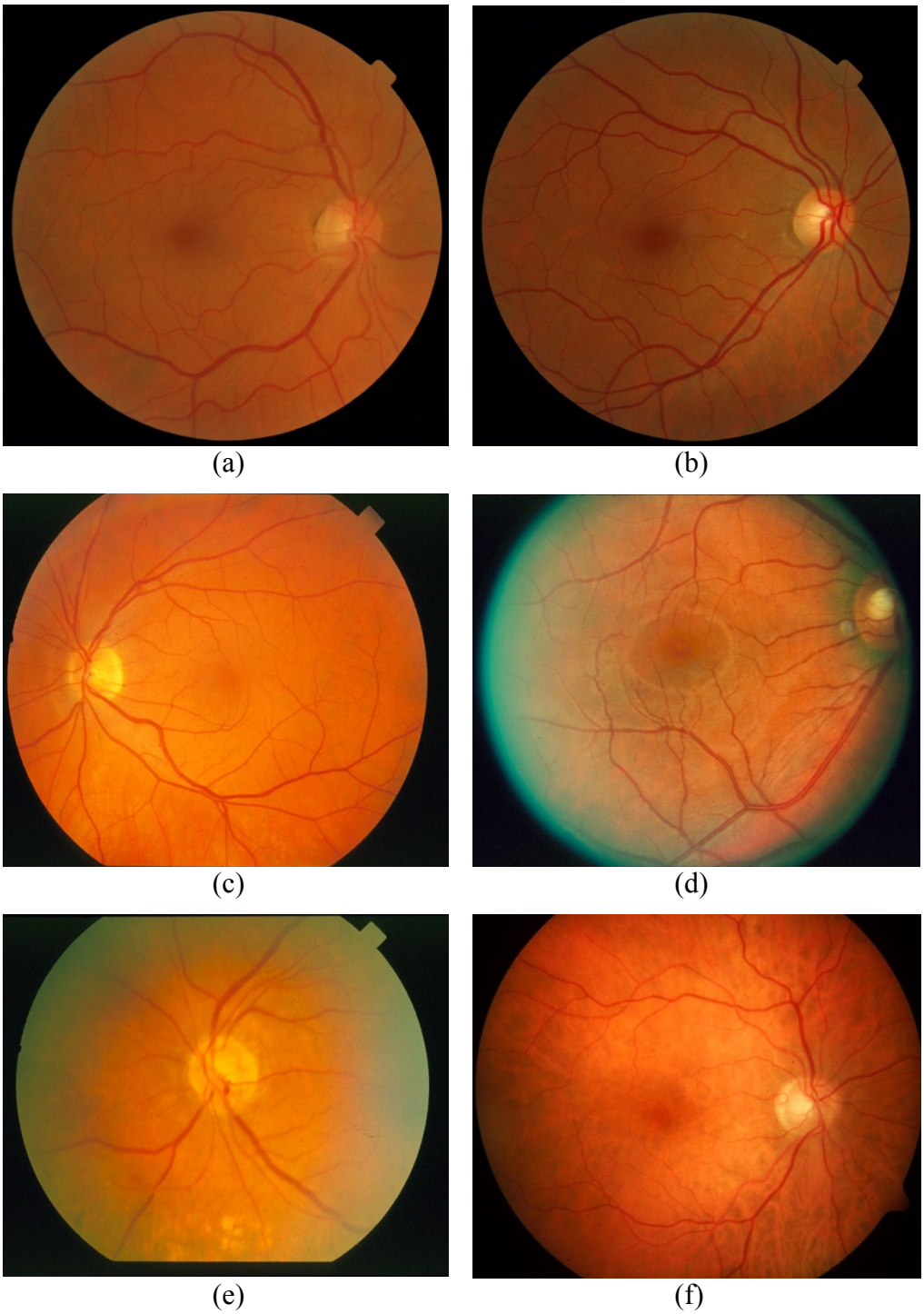
| | Normal | MA | NV |
|---|---|---|---|
| Normal | 79.34 | 14.11 | 6.54 |
| MA | 5.31 | 92.19 | 2.5 |
| NV | 2.46 | 52.61 | 44.93 |

Table 4. Confusion Matrix for the Proposed Approach

| | Normal | MA | NV |
|---|---|---|---|
| Normal | 88.12 | 8.11 | 3.77 |
| MA | 3.05 | 86.05 | 10.9 |
| NV | 1.41 | 8.59 | 90 |

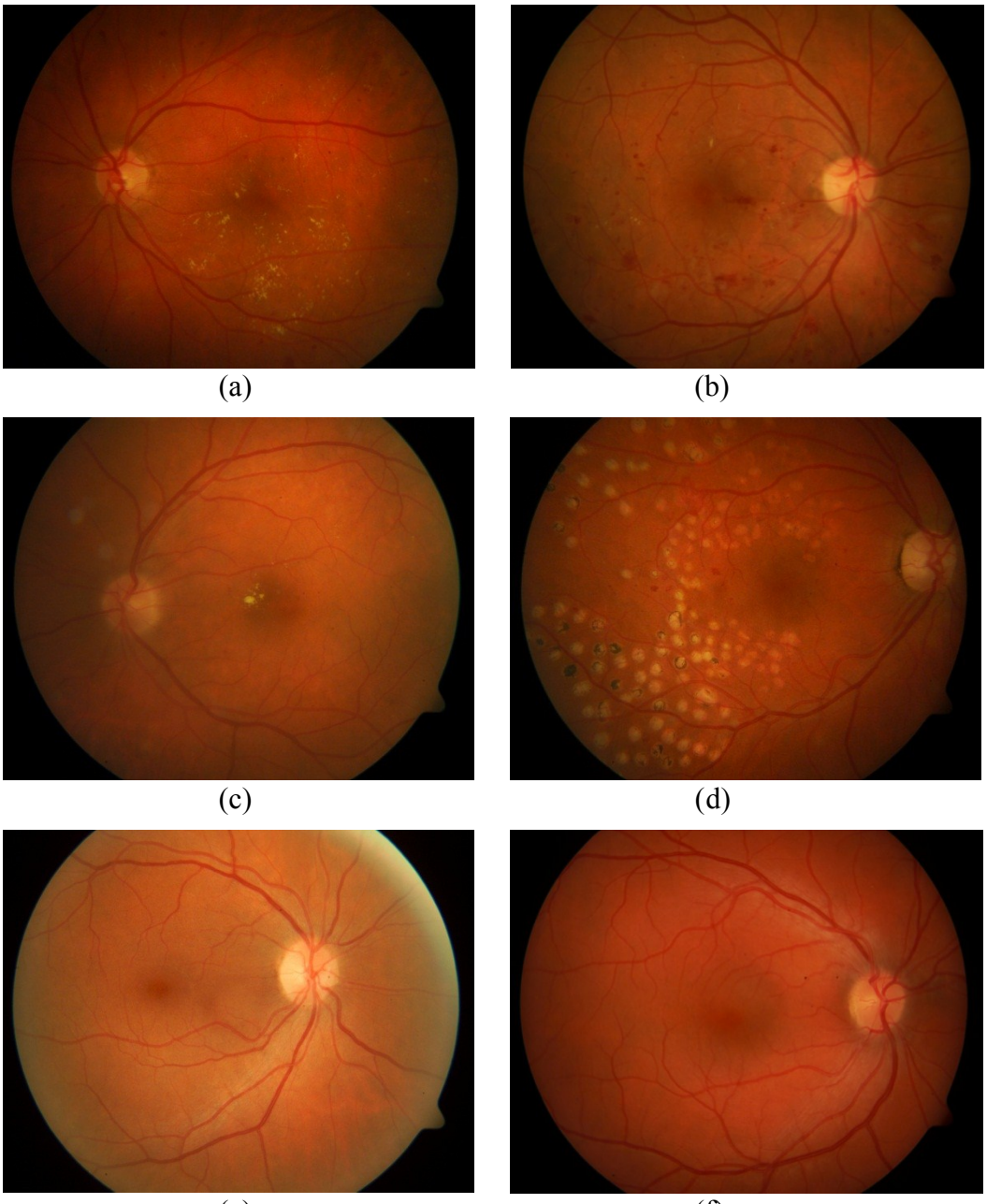
By observing the confusion matrices, it can be inferred that the proposed framework performs the classification almost equally well for all the classes with only a 64-D feature space. In contrast, the original AutoCC feature has difficulty distinguishing between MA and NV. This illustrates the importance and efficiency of the spectrally tuned quantizer. It should be noted that in this comparison the classifier is the same (i.e., MIL) and the parameters (number of references and citers) are also well-tuned to get the best performance. Thus the gain is purely due to the improved feature design. Examples of correctly and incorrectly classified images are shown in Fig. 20, 21 and 22.

### 3.4 Classification of DR Images:



(a) (b) (c) (d) (e) (f)

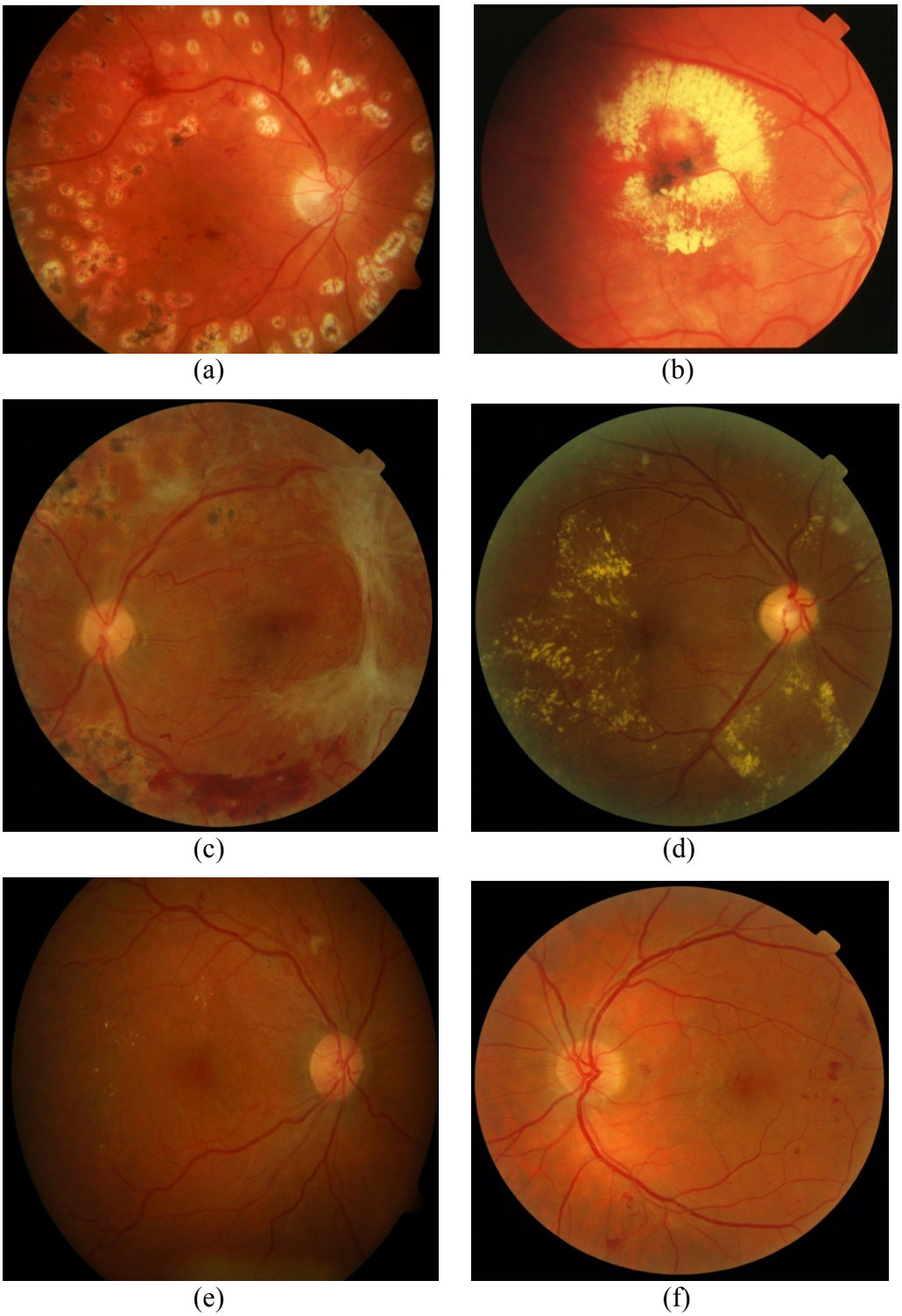
Fig. 20. Classification of Normal Images (a), (b), (c)- Correctly Classified, (d), (e), (f)- Incorrectly Classified



(a) (b) (c) (d) (e) (f)

Fig. 21. Classification of MA Images

(a), (b), (c)- Correctly Classified, (d), (e), (f)- Incorrectly Classified



(a) (b) (c) (d) (e) (f)

Fig. 22. Classification of NV Images (a), (b), (c)- Correctly Classified, (d), (e), (f)- Incorrectly Classified

# 4. RETRIEVAL OF DIABETIC RETINOPATHY IMAGES

The use of digital retinal imaging is increasing day by day. Also, because of everyday use of computers, it has become convenient to keep archived data stored in them rather than maintaining a written record of each patient. Therefore it is necessary to develop automated algorithms which can give instant access to archived images. If developed successfully, this will be a novel way of exploring vast knowledge hidden in archived images. Thus retrieval of medical images can point the doctor in right direction to start the treatment.

In the later section, an approach is developed to retrieve DR images which are categorized into classes, namely, normal, MA and NV. It is shown that the proposed retrieval framework outperforms other state-of-art retrieval frameworks.

## 4.1 State-of-art Retrieval Frameworks

Some of the state-of-art image retrieval systems built using similarity metrics are Gabor features [49] and semantics of histogram of neighborhood moments [50]. Another well-studied feature for CBIR puposes related to general images is the color correlograms [42, 1]. They describe the global correlation of local spatial correlation of colors. This is a strong representation of textures with considerably small dimensionality. These features (especially HNM) are preferred in medical image retrieval but might not be appropriate for DR images because of their peculiar color spectrum. The reasons of the same have been detailed in Classification using Proposed Approach. Owing to all those reasons, an approach has to be developed which retrieves clinically relevant images with greater

accuracy. By clinical relevance, it is meant that images of the same class will be retrieved and also with similar lesion types. This section describes the details of an automated algorithm for DR image retrieval. It takes a novel approach of retrieving images using multiple instances with a proposed *rank-Citation KNN* method to obtain better retrieved images.

## 4.2 Proposed Retrieval Framework

The proposed retrieval framework operates on a feature space of CC, statistics of steerable Gaussian filter (SGF) response along with fast radial symmetric transform. Clinically relevant images are retrieved by using above features and a multiple instance KNN called *rankKNN*.

### 4.2.1. The Features

**1. Color Correlogram Features:**

As it was described in Spectrally Tuned Color Correlogram, the CC features provide the global correlation of local spatial correlation of pixels of quantized image. When the spatial distribution of only those pixels which lie in the same bin is considered, the CC features become AutoCC features. AutoCC features combined with some other features are used in many retireval algorithms. Li in [1] proposed a method that uses AutoCC along with other texture features to perform natural image retrieval.

In this section, the use of modified CC features is proposed. Firstly, all the unique shades (RGB triplets) in the training set are first extracted and then

arranged in a $NX3$ matrix. K-means clustering is then performed on the matrix to get 16 clusters. The 16 centroids can be visualized as the dictionary of the quantizer. CC is then calculated for every pair of bins $(b_1, b_2)$ where $(b_1, b_2) \in [1,2,\dots,16]$ as described in Color Correlogram Feature Extraction. This generates a 256 dimensional vector describing the distribution of each bin with respect to all other bins.

**2. Steerable Gauss Filters (SGF):**

In many image processing and vision tasks, there is often a need to apply the same filter with different orientations under adaptive control and calculate response of the same. Oriented filters are applied in various computer vision and image processing tasks, such as edge detection, image enhancement, image compression and texture analysis. Freeman and Adelson proposed a method in [65] to synthesize filters of arbitrary orientations from linear combinations of basis filters. This is called as "steering" a filter to any desired orientation. It is possible to determine the filter output as a function of orientation. The term "steerable filter" describes a class of filters of arbitrary rotations synthesized as linear combinations of a set of basis filters.

As an introduction to steerable filters, a simple two dimensional case is discussed here. Consider a 2-D, circularly symmetric Gaussian function G represented in terms of Cartesian coordinates $x$ and $y$:

$$G(x, y) = e^{-(x^2+y^2)} \quad (11).$$

The $n^{th}$ derivative of a Gaussian in the $x$ direction is represented by $G_n$. The

rotation operator $(\cdot)^{\theta}$ for $f(x,y)$ is represented as $f^{\theta}(x,y)$, a version of $f(x,y)$ rotated by an angle $\theta$ about the origin. The derivative of $G(x,y)$ in the $x$ direction can be given as,

$$G_1^{0^\circ} = \frac{\partial}{\partial x} e^{-(x^2+y^2)} = -2xe^{-(x^2+y^2)} \tag{12}$$

The above function when rotated by $90^\circ$ gives,

$$G_1^{90^\circ} = \frac{\partial}{\partial y} e^{-(x^2+y^2)} = -2ye^{-(x^2+y^2)} \tag{13}.$$

It is possible to construct a $G_1$ filter at any arbitrary rotation by taking a linear combination of $G_1^{0^\circ}$ and $G_1^{90^\circ}$:

$$G_1^{\theta^\circ} = \cos(\theta)\, G_1^{0^\circ} + \sin(\theta)\, G_1^{90^\circ} \tag{14}$$

Since $G_1^{0^\circ}$ and $G_1^{90^\circ}$ span the entire set of $G_1^{\theta^\circ}$ filters, they can be called as the *basis filters* for $G_1^{\theta^\circ}$. Also, the $\cos(\theta)$ and $\sin(\theta)$ terms are called the corresponding *interpolation functions* for the basis filters. Since convolution is a linear operation, it is possible to synthesize an image at any desired rotation by taking linear combinations of images filtered with $G_1^{0^\circ}$ and $G_1^{90^\circ}$. If $*$ represents convolution operator then,

$$R_1^{0^\circ} = G_1^{0^\circ} * I \tag{15}$$

$$R_1^{90^\circ} = G_1^{90^\circ} * I \tag{16}$$

Then,

$$R_1^{\theta^\circ} = \cos(\theta)\, R_1^{0^\circ} + \sin(\theta)\, R_1^{90^\circ} \tag{17}.$$

The visualization of derivatives of Gaussian filters exemplifies steerability.

The basis functions as well as certain linear combinations of these two filters are shown in Fig. 23.

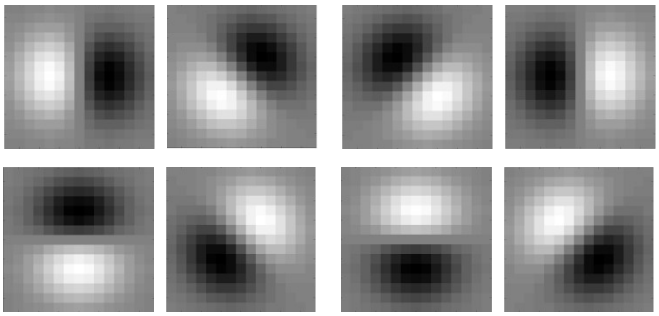


Fig. 23. Basis Filters (first column) and their Linear Combinations

### 3. Fast Radial Symmetry Transform (FRST)

Many context-free points of interest operators have been proposed in the past. These operators concentrate on local radial symmetry as a measure of interest. This agrees well with the psychophysical findings as reported by many researchers in the past [66, 67, 68]. There are many points of interest operators but one of the well-known operators is *generalized symmetry transform* [69] and has been applied for detecting facial features. But the major drawback of this method is its low computational efficiency- of the order of $O(KN^2)$, where $K$ is the number of pixels in an image and $N$ is the width of the neighborhood. There were many attempts to design points of interest operator with a high computational efficiency. Keeping this in mind, Loy and Zelinsky proposed fast radial symmetry transform [70] which is of the order of $O(KN)$.

FRST is calculated at one or a set of radii $n \in N$, where $N$ is set of radii. The value of the transform at radius $n$ informs about the contribution of the gradients

which are at a distance of $n$ from each point to the radial symmetry. This approach determines the contribution of every pixel to the symmetry of the pixels surrounding it. At each radius $n$, an orientation projection image $O_n$ and magnitude projection image $M_n$ are generated. Now, a positively affected pixel is defined as the pixel to which the gradient vector $g(p)$ is pointing. Similarly, a negatively affected pixel is defined as the pixel from which the gradient vector is pointing away. They are calculated as,

$$p_{+ve}(p) = p + round\left(\frac{g(p)}{\|g(p)\|} * n\right) \tag{18}$$

$$p_{-ve}(p) = p - round\left(\frac{g(p)}{\|g(p)\|} * n\right) \tag{19}$$

where, $round(\cdot)$ performs the rounding operation.

$O_n$ and $M_n$ are initialized to zero. For each pair of affected pixels ($+ve$ and $-ve$), the pixel corresponding to $p_{+ve}$ in $O_n$ and $M_n$ is increased by 1 and $\|g(p)\|$ respectively, while for $p_{-ve}$, it is decremented by 1. This can be mathematically written as,

$$O_n\big(p_{+ve}(p)\big) = O_n\big(p_{+ve}(p)\big) + 1 \tag{20}$$

$$O_n\big(p_{-ve}(p)\big) = O_n\big(p_{-ve}(p)\big) - 1 \tag{21}$$

$$M_n\big(p_{+ve}(p)\big) = M_n\big(p_{+ve}(p)\big) + \|g(p)\| \tag{22}$$

$$M_n\big(p_{-ve}(p)\big) = M_n\big(p_{-ve}(p)\big) - \|g(p)\| \tag{23}.$$

Now, the radial symmetry contribution at radius $n$ is defined a process of convolution,

$$S_n = F_n * A_n \tag{24}$$

where,

$$F_n(p) = \frac{M_n(p)}{k_n}\left(\frac{\tilde{O}_n(p)}{k_n}\right)^{\alpha} \tag{25}$$

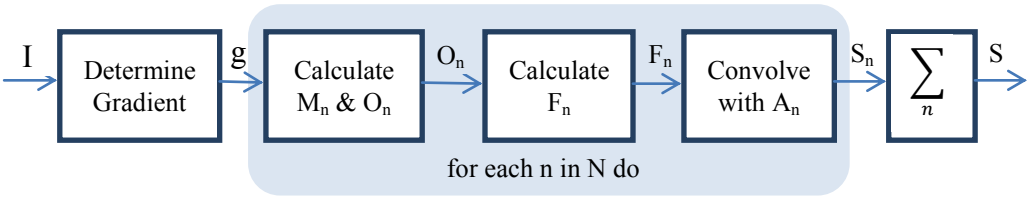
and

$$\tilde{O}_n(p) = \begin{cases} O_n(p) \; if \; O_n(p) < k_n \\ k_n \quad otherwise \end{cases} \tag{26}.$$

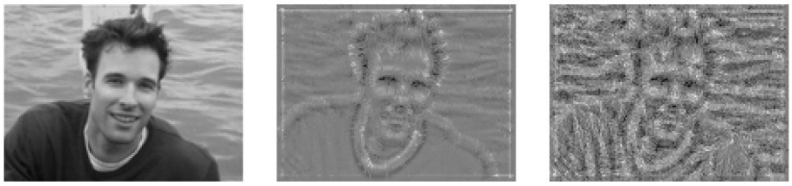
$A_n$ is a two dimensional Gaussian, $\alpha$ is called as the radial strictness parameter and $k_n$ is scaling factor for normalization of $O_n$ and $M_n$. The final transform is defined as the mean of the symmetry contributions over the set of radii,

$$S = \frac{1}{|N|}\sum_{n \epsilon N} S_n \tag{27}.$$

This transform is able to process a frame of $240 \times 320$ in 13.2 ms on a Pentium III processor. The following figure explains the steps involved in computing the transform and gives an example.




(a)



$I$ $M_3$ $O_3$

(b)

Fig. 24. Overview of FRST (a) Steps Involved in Computation of FRST (b) Visualization for $n = 3$ and $\alpha = 2$ [70]

### 4.2.2. Application of SGF and FRST to DR Images

SGF features are applied to DR images since it is possible to design them in such a way that they adaptively filter out the structures in a desired orientation as well as it can be used for contour detection. Also, the SGF filter response can be used to classify and identify lines, edges and contours, all having the potential of becoming well-distinguishing features in case of DR images. The details of SGF are explained in the section 4. Fig. 23 shows the basis filters as well as all the linear combinations of the filters used in this application. The 8 filters are separated by 45°. By using these filters as kernels, it is possible to calculate the directional derivatives in each direction in an image. Fig. 25 shows the filter response of a typical DR image to basis oriented at 225 degrees.

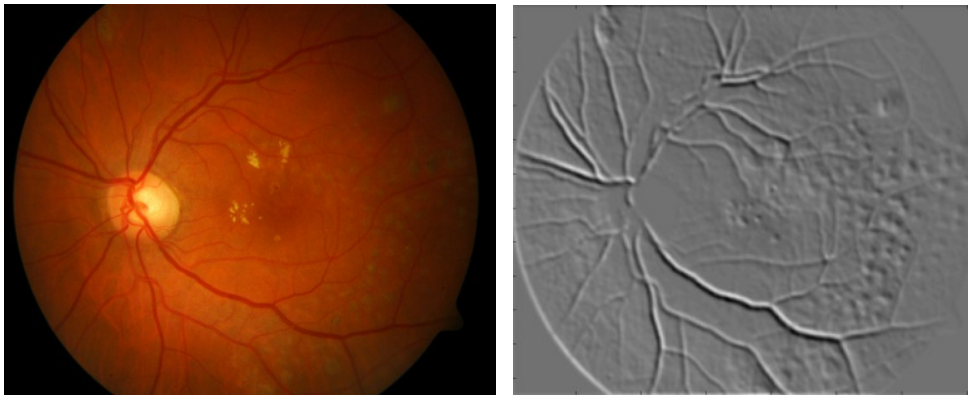


Fig. 25. SGF Filter Response to an MA Image. Input MA Image (left) and filter Response (right) of a Typical DR Image.

MA image is given as an input. It can be seen in the right image that the SGF response highlights the location of nerves and the MA spots. The highlighted spots can be statistically viewed as variation in the values of SGF response. Therefore the standard deviation, skewness and kurtosis of the SGF response are

taken as features. This generates a 3-D vector for each angle. For all the 8 filters shown in Fig. 23, we have a 24 dimensional feature vector for each image.

In Fig. 25, it is observed the MA symptoms are rightly appreciated but the nerves are unnecessarily highlighted too. Due to this a lot of noise on the feature space is produced as the highlighted regions affect the moments of the SGF responses. Thus the response of SGF cannot be used as a feature alone. To overcome this problem, it is needed to specifically select regions of operation around potential points of interest. To detect such points of interests, the use of FRST is proposed [71]. The transform can be computed for a set of radii for accurately detecting points of interest. In this context to negate the presence of the nerves and accurately detect the lesions, we use a set of radii- $\{1,3,5\}$. The interest point detection on a Normal, MA and NV is shown in Fig. 26.

Though majority of the detected points are lesions, there also exist false detections too. The value of FRST at these points is very large on positive and negative side. This is because FRST works on the local maxima detection and the interest points are detected by performing non-maxima thresholding on the image generated by FRST. The detected points are subsequently superimposed on the original image for visualization purposes. Thus the points detected by FRST cannot be used as a feature by itself.

After detecting the interest points, a neighborhood of $15 \times 15$ is chosen around each of those ineterst points and the aforementioned statistics are extracted

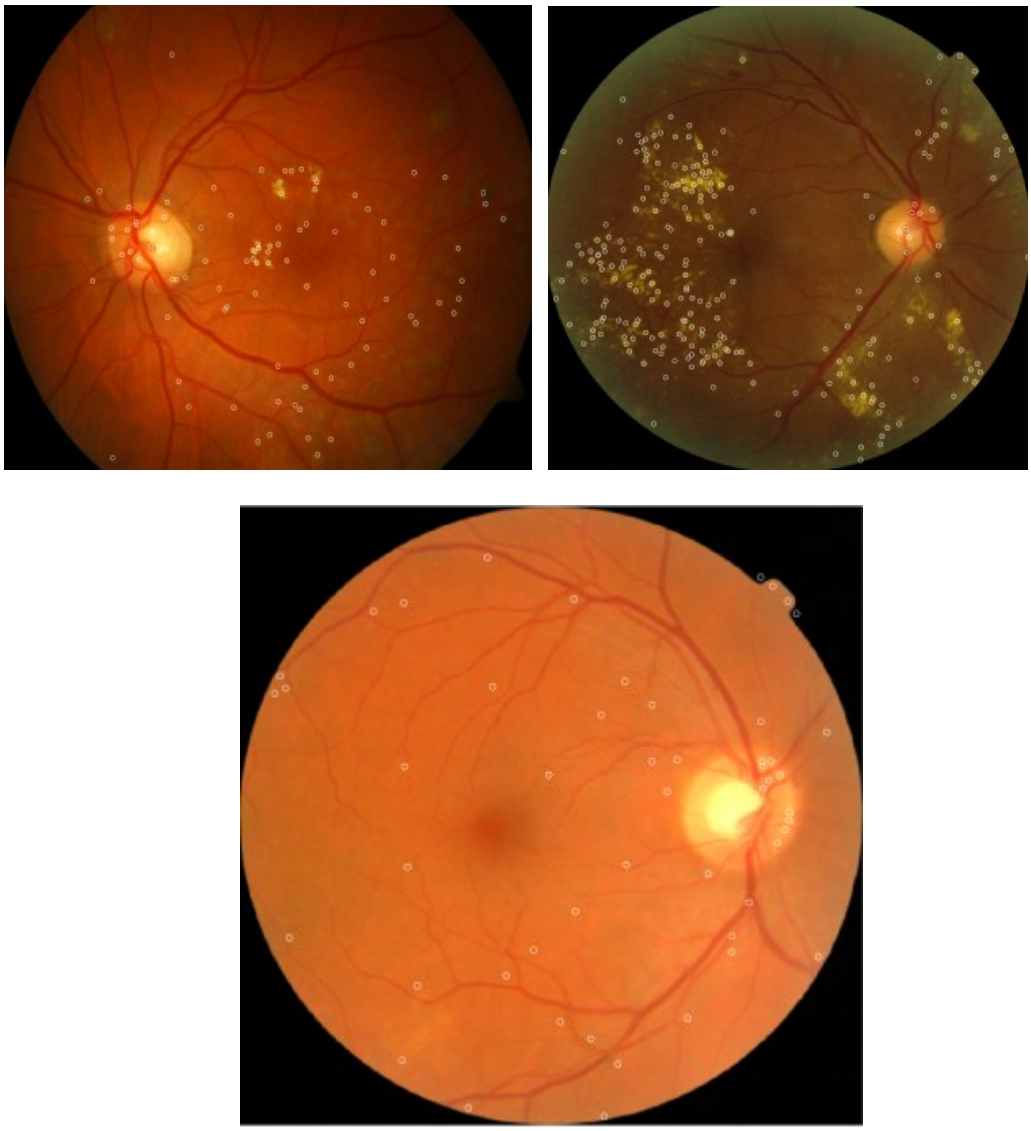


Fig. 26. Interest Points Detection using Fast Radial Transform. The Top Rows are Affected Eyes, while the Bottom is a Normal Eye. (Zoom in for Better Viewing)

for every $15 \times 15$ window. Assume that there are $\lambda$ number of points of interest detected on a given image, so $\lambda$ feature vectors will be extracted for the image, the dimensionality of each feature vector being 24-D. When there are multiple points lying in a $15 \times 15$ window, the feature space is reduced to a single feature vector of 24 dimensions by taking mean over each dimension individually for

every detected point in that window. This provides consistency in the dimensionality of the feature space. In the multiple instance framework that is discussed later, this action is suitable as each instance is a part of the image which consists of a very few number of interest points almost always belonging to the same lesion of interest. In a situation where no interest points are detected for a given window, the feature vector is all zeros. This makes the feature space consistent for the proposed McMIL algorithm.

### 4.2.3. The MCMIL Framework for DR Image Retrieval

As mentioned earlier, in the MIL formulation, a bag has multiple instances and therefore it corresponds to multiple points in the instance space. According to this formulation, the image is divided into 64 blocks on an 8X8 grid. Unwanted instances that contain majority of black background regions are dropped by thresholding. The features discussed in the section 4.2.1 are extracted for each instance. In case of the SGF response features, the FRST is performed on a global level. This identifies points of interest on a global level, whereas the statistics of SGF response features are extracted for the detected points in each instance. This gives us a 280 (256-CC + 24-statistics of SGF features) dimensional feature vector for each instance and the number of instances per image (thereby bag) might vary since the number of instances depend on the thresholding scheme.

This problem is solved by developing a modified approach called *rankKNN* originally based on Citation-KNN [54] which falls into the category of lazy learning. Citation-KNN uses a modified Hausdorff distance. Hausdorff distance

provides a measure of dissimilarity between two subsets of metric space. Hausdorff distance is very sensitive to outliers and to overcome this problem, use of a minimal Hausdorff distance was proposed in [54]. It is defined by,

$$H(A,B) = \min_{b \in B} \min_{a \in A} \| a - b \| \quad (28),$$

where $H(A,B)$ is the Hausdorff distance between any two bags $A$ and $B$ while $a$ and $b$ are instances in bags $A$ and $B$ respectively. $H(A,B)$ gives the least distance between two bags. In case of DR images, the minimal Hausdorff distance between an MA and a Normal image will be almost equal to the distance between two Normal images, for example, in Fig. 16(a) only one instance of an MA image (marked by red border) differs from the instances in Normal image shown in Fig. 16(b). Thus the use of the minimal Hausdorff distance does not suffice. Therefore it is necessary to develop a new way of measuring the distance between two bags.

***Computation of Distance between two bags**:*

Suppose there are two bags $A$ and $B$ with $m$ and $n$ instances each. The minimum distance between the $i^{th}$ instance of $A$ and all instances of $B$ is given by,

$$D_{(i)}(A,B) = \min_{b \in B} d(a_i, b_j) \;\; \forall\, j \in \{1,2,..n\} \quad (29),$$

here $d(\cdot)$ represents Euclidean distance between two instances. $D$ is an $m$-dimensional vector containing the minimum distances between each instance of $A$

and all instances of $B$.

Let $Q$ be the bag of instances for a query image for which clinically-relevant retrieval has to be performed. The minimum distances between every instance in $Q$ and instances of all the images in the database are calculated as described in (29. After sorting individually for every instance, this gives a similarity list (SL) for every instance in $Q$. The mean of ranks in the SL of instances belonging to that particular bag is calculated. This is repeated for every bag and it is called a bag-level aggregated similarity rank (ASR). This is termed as the m-Ranking procedure and the rank list thus formed the *m-Rank*. The procedure of *m-Ranking* is explained in the following algorithm.

***Algorithm for calculation of m-Rank**:*

Step 1. Calculate $\mathrm{D}$ between $\mathrm{Q}$ (assume number of instances in $\mathrm{Q} = m$) and every image in the database and form the vector $\mathrm{D}' = [\mathrm{D}_1, \mathrm{D}_2, \dots, \mathrm{D}_\mathrm{T}]$. Here, $\mathrm{D}'$ is a $\mathrm{m} \times \mathrm{T}$ matrix, where $\mathrm{T} + 1$ is total number of images in database.

Step 2. $\mathrm{D}'$ is now sorted along columns and the sorted indices gives SL where $\mathrm{SL}(\mathrm{u}, \mathrm{v}) = \mathrm{w}$ represents the $\mathrm{v}^{\mathrm{th}}$ best match for the $\mathrm{u}^{\mathrm{th}}$ instance of $\mathrm{Q}$ and the best match is an instance belonging to image $\mathrm{w}$.

Step 3. For a new bag $\mathrm{N}$ in the database, collect the column indices of all occurances of $\mathrm{N}$ in SL and arrange them in an array *col*.

Step 4. Calculate ASR for $\mathrm{N}$ as $\mathrm{ASR(N)} \triangleq \mathrm{avg(col)}$.

Step 5. Repeat Steps 3 and 4 for all bags in the database to arrive at the complete ASR.

Step 6. Sort ASR, whose indices will give m-Rank list.

The minimum distance between two bags need not necessarily give optimum retreival for DR images as only few instances contribute towards the label. To avoid this problem, *m-Rank* is made more robust by incorporating *citer-Rank* to it. This is accomplished by obtaining a final *meanRank* which is the average of *m-Rank* and *citer-Rank* of $Q$.

The top *k* retreived images are based on the *meanRank* list thus created. In the algorithm, *k* doesn't affect the performance of the algorithm as the calculation of *meanRank* is independent of *k*.

### 4.3 Experimental Results and Analysis

The experimental setup is exactly same as mentioned in 3.3. The proposed system is evaluated against some state-of-art computer vision retrieval systems using Gabor textural features [49], semantic of neighborhood color moment histogram features (HNM) [50] along with the original AutoCC features that are proposed in [1]. All of these methods are based on a distance based *k-nearest neighbor* search systems. The distances used are those that are recommended in the specific articles. These systems are popular and prior art in medical image retrieval literature. All the 425 images in the database are individually queried and the top $k$ $(k = 5)$ images are retrieved using the proposed approach. Evaluation

metrics are adopted from [72, 73]. The evaluation metrics are briefly described below:

1. Hit: If one of the images retrieved has the same label as the queried image then it is declared as a hit.
2. $\geq k$-Hit Rate: Percentage of images out of all test images that have $\geq k$-Hit(s). This measures the probability of retrieving at least $k$ images correctly out of all retrieved images.
3. Mean Reciprocal Rank (MRR): The rank of the first relevant image in all the retrieved images is noted and then all the ranks are averaged over all the query images to get MRR. This metric provides useful information about the ability of the system to return an image at the top of the rankings.
4. Success at rank $k$ ($S@k$): It is defined as the probability of retrieving a relevant image among top $k$ images. So, $S@2$ gives the probability of finding a relevant image among top 2 retrieved images.

The following tables present the results of $\geq k$ Hit-Rate, MRR and $S@k$.

Table 5. Mean Accuracies and $\geq k$ Hit-Rate in Percentages

| | Mean Accuracy | $\geq 2$ Hit-Rate | $\geq 3$ Hit-Rate | $\geq 4$ Hit-Rate | $\geq 5$ Hit-Rate |
|---|---|---|---|---|---|
| AutoCC | 60.85 | 79.06 | 64.47 | 42.35 | 23.29 |
| Gabor | 68.61 | 83.76 | 75.76 | 59.52 | 31.76 |
| HNM | 68.04 | 84.47 | 74.58 | 55.76 | 32.00 |
| Proposed | 75.48 | 88.70 | 82.58 | 68.00 | 43.29 |

Table 6. Success at $k^{th}$ Rank in Percentages

| | $S@2$ | $S@3$ | $S@4$ | $S@5$ |
|---|---|---|---|---|
| AutoCC | 63.88 | 62.19 | 61.88 | 60.85 |
| Gabor | 68.47 | 68.31 | 68.11 | 68.61 |
| HNM | 71.05 | 70.27 | 68.82 | 68.04 |
| Proposed | 77.17 | 75.60 | 75.76 | 75.48 |

Table 7. Mean Accuracy at $k^{th}$ Rank in Percentages

| | Rank 1 | Rank 2 | Rank 3 | Rank 4 | $Rank\ 5$ |
|---|---|---|---|---|---|
| AutoCC | 64.24 | 63.88 | 62.19 | 61.88 | 60.85 |
| Gabor | 69.41 | 67.52 | 68.00 | 67.52 | 70.58 |
| HNM | 73.64 | 68.47 | 68.70 | 64.94 | 64.94 |
| Proposed | 77.41 | 76.94 | 72.47 | 76.23 | 74.35 |

The mean accuracies reported in Table 5 suggest that the proposed approach clearly outperforms all the competing methods considered. From Table 5, it can also be observed that the $\geq k$ Hit-rate of this approach is much higher than the other methods. This implies that the proposed method retrieves more relevant images in the top five retrieved images than the other methods. This provides significant assistance to an ophthalmologist in clinical diagnosis where it is necessary to retrieve maximum number of clinically relevant images.

HNM fails to retrieve images with clinically similar lesions in particular to that of hemorrhages. Therefore, this reduces the clinical relevance of the method and thus it may offer little less assistance to the ophthalmologist. Fig. 27-Fig. 32 show 5 retrieved images for 6 query images (2 from each category) using proposed approach. The retrieved images are not only similar in appearance to the

given image, but they also all carry similar lesions, hence this approach achieves desired clinical relevance.

The successes at each rank are reported in Table 6. They are not only higher than the other methods but also illustrate consistency. This implies that the probability of finding clinical relevance in a retrieved image at any rank is fairly uniform and that the ratio of relevant retrieved images to irrelevant retrieved images is high. This makes the retrieved images more trustworthy for an ophthalmologist as an image retrieved at any rank is considerably trustworthy.

Another observation that was drawn from the retrieved images is that, though for MA images, NV images are retrieved occasionally and vice-versa, this is still clinically relevant in the sense that most NV images contain symptom of MA and those MA images which are about to turn NV contain symptoms of NV. This provides more information to an ophthalmologist though the labeling was incorrect. This effectively increases the satisfaction in retrieval.

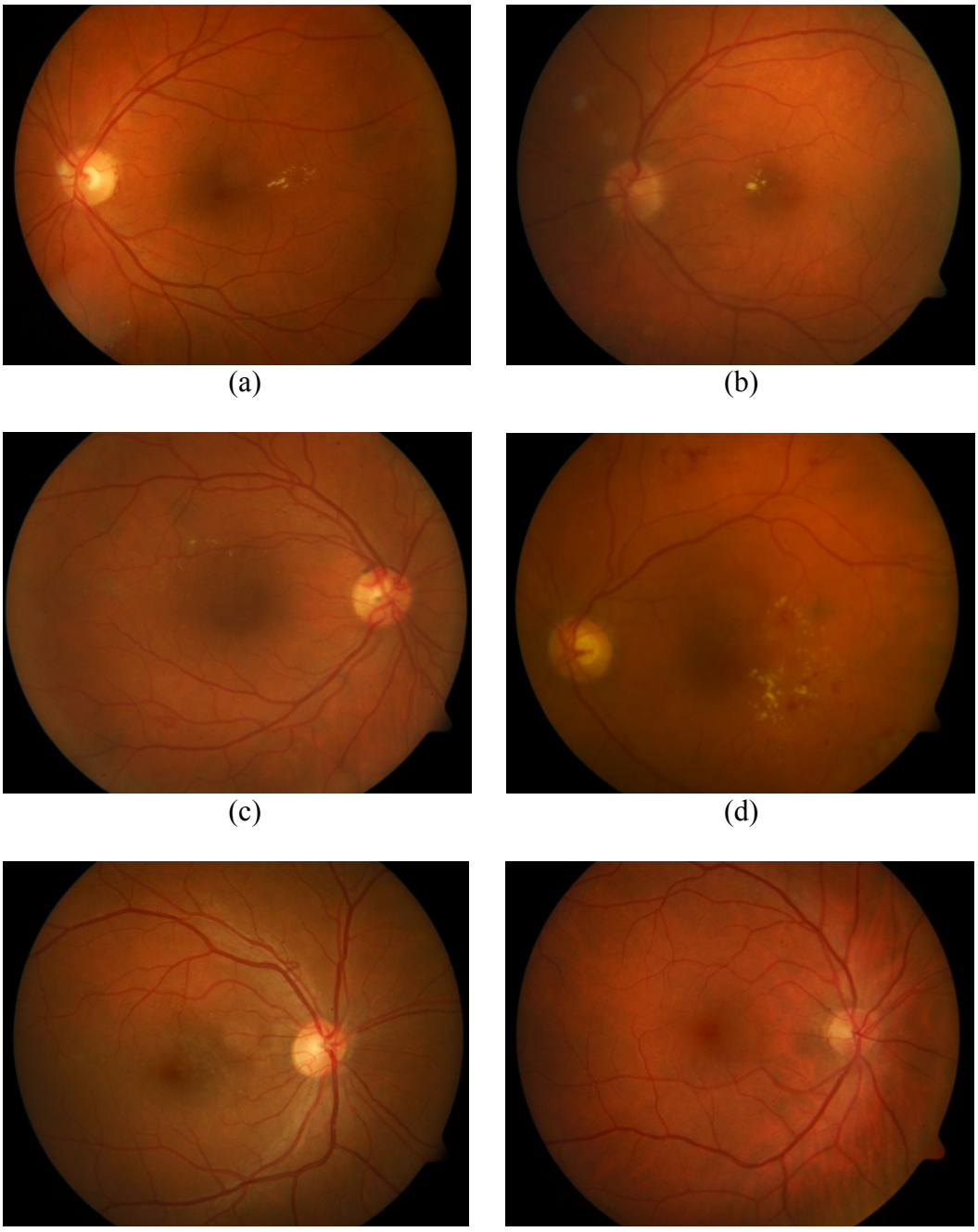
A mean confusion matrix is created for the proposed approach and shown in Table 8. The following example explains how the confusion matrix is constructed. Suppose 100 normal images are queried and therefore there will be 500 retrieved images. The 1$^{st}$ row of the confusion matrix indicates that 71.5% of the 500 images were normal, 15.87% images were MA and 12.62% images were NV. Similarly the second and third row can be explained. Thus the confusion matrix shows the distribution of top 5 retrieved images for queried images of each category. It can be seen from the confusion matrix that the errors increase while

retrieving for normal and NV images. Some errors in NV are owing to the fact that certain NV images in the database are treated for MAs. The retrieval accuracy for MA in particular is higher and this is particularly advantageous for a clinician as NV and Normal are usually extreme cases that can be easily diagnosed. Following pages show some examples of DR image retrieval.

Table 8. Mean Confusion Matrix for the Proposed Approach

| | Normal | MA | NV |
|---|---|---|---|
| Normal | 71.50 | 15.87 | 12.62 |
| MA | 9.72 | 83.42 | 6.85 |
| NV | 17.61 | 17.14 | 65.23 |

### 4.4 Retrieval of DR Images



(a) (b)

(c) (d)

(e) (f)

Fig. 27. Top 5 Retrieved Images for a Query MA Image. (a) is a Query MA Image and (b)-(f) are Retrieved MA Images

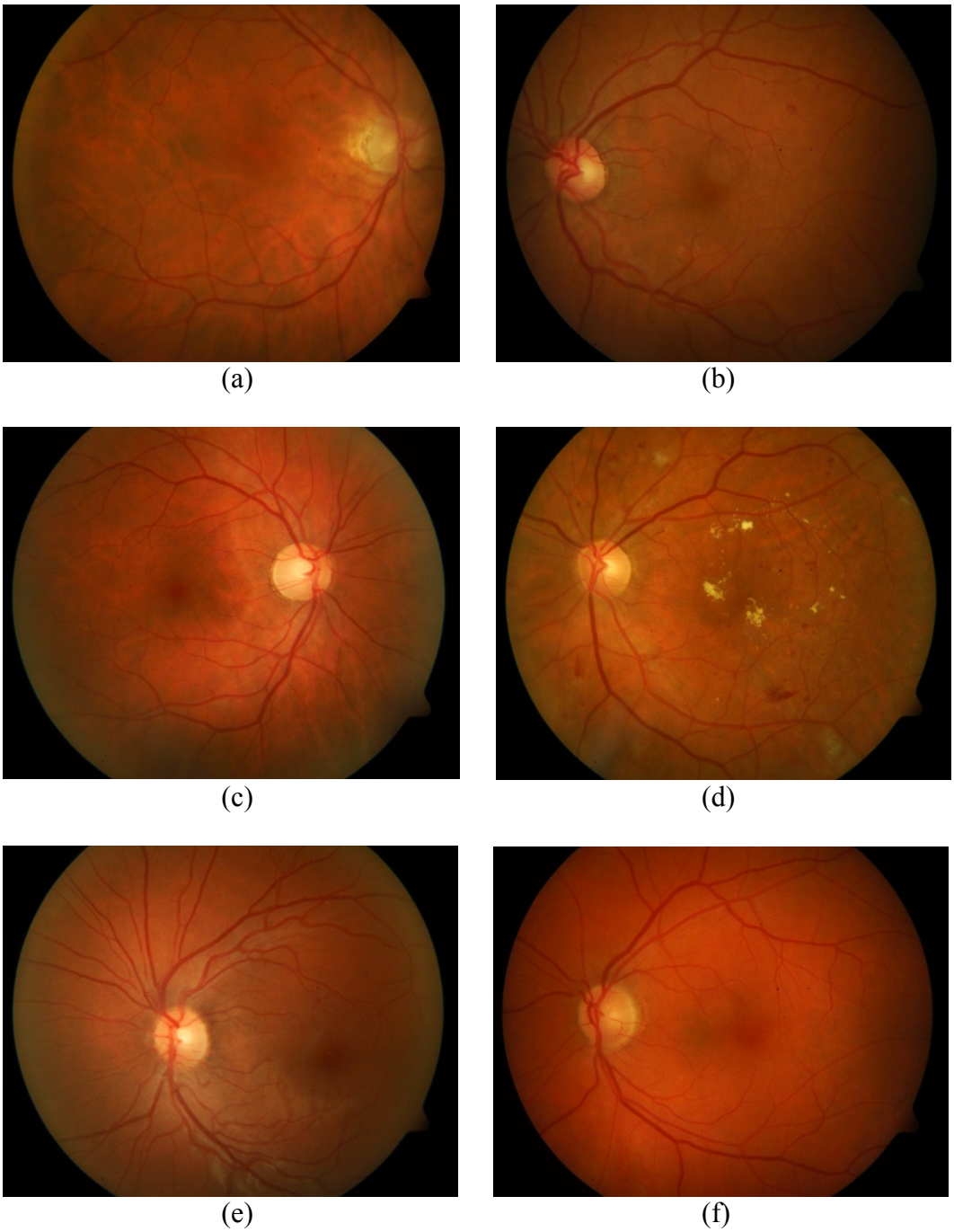
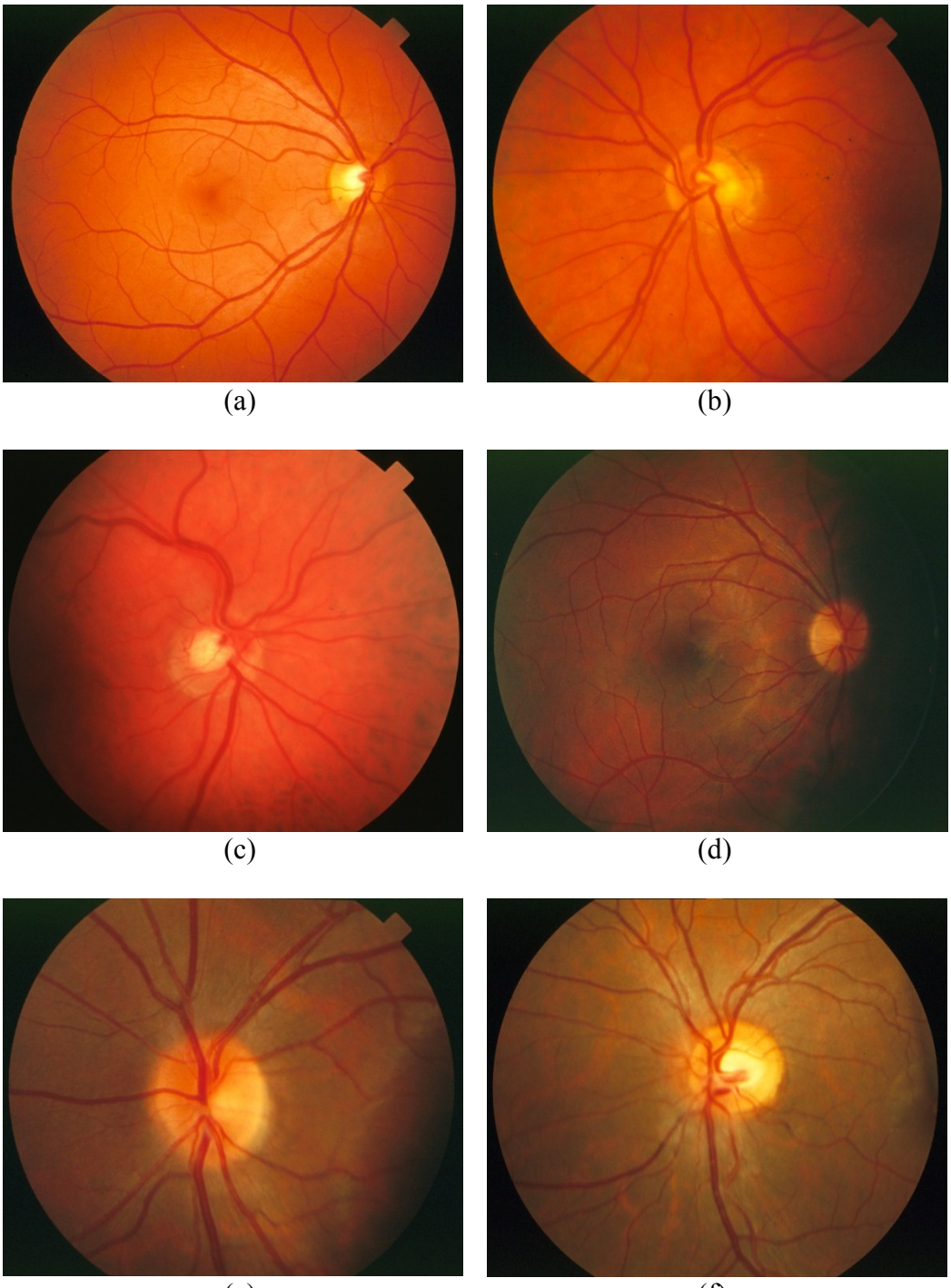


Fig. 28. Top 5 Retrieved Images for a Query MA Image. (a) is a Query MA Image and (b), (c), (d) and (f) are Retrieved MA Images. (e) is a Normal Image



(a) (b) (c) (d) (e) (f)

Fig. 29. Top 5 Retrieved Images for a Query Normal Image. (a) is a Query Normal Image and (b)-(f) are Retrieved Normal Images

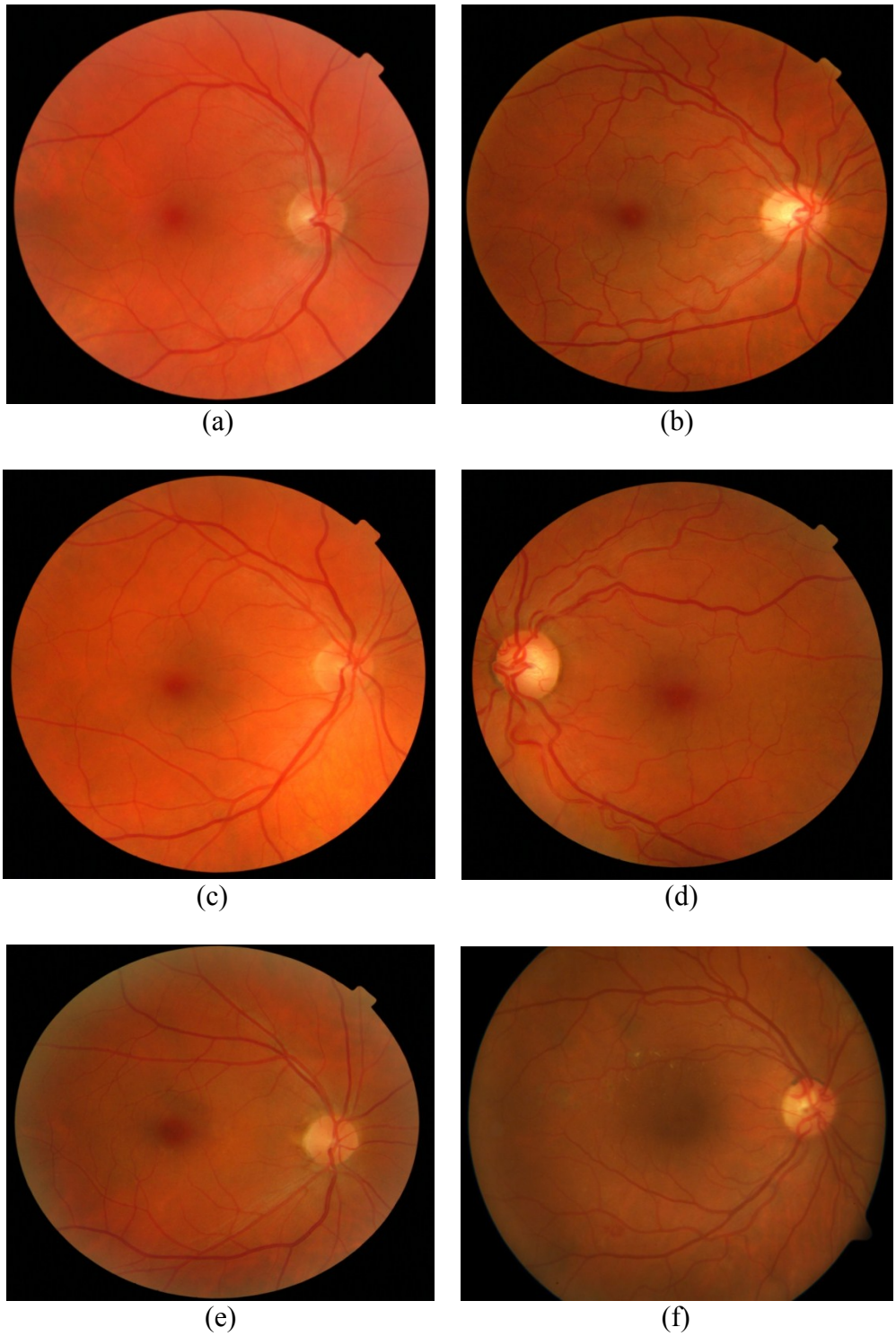
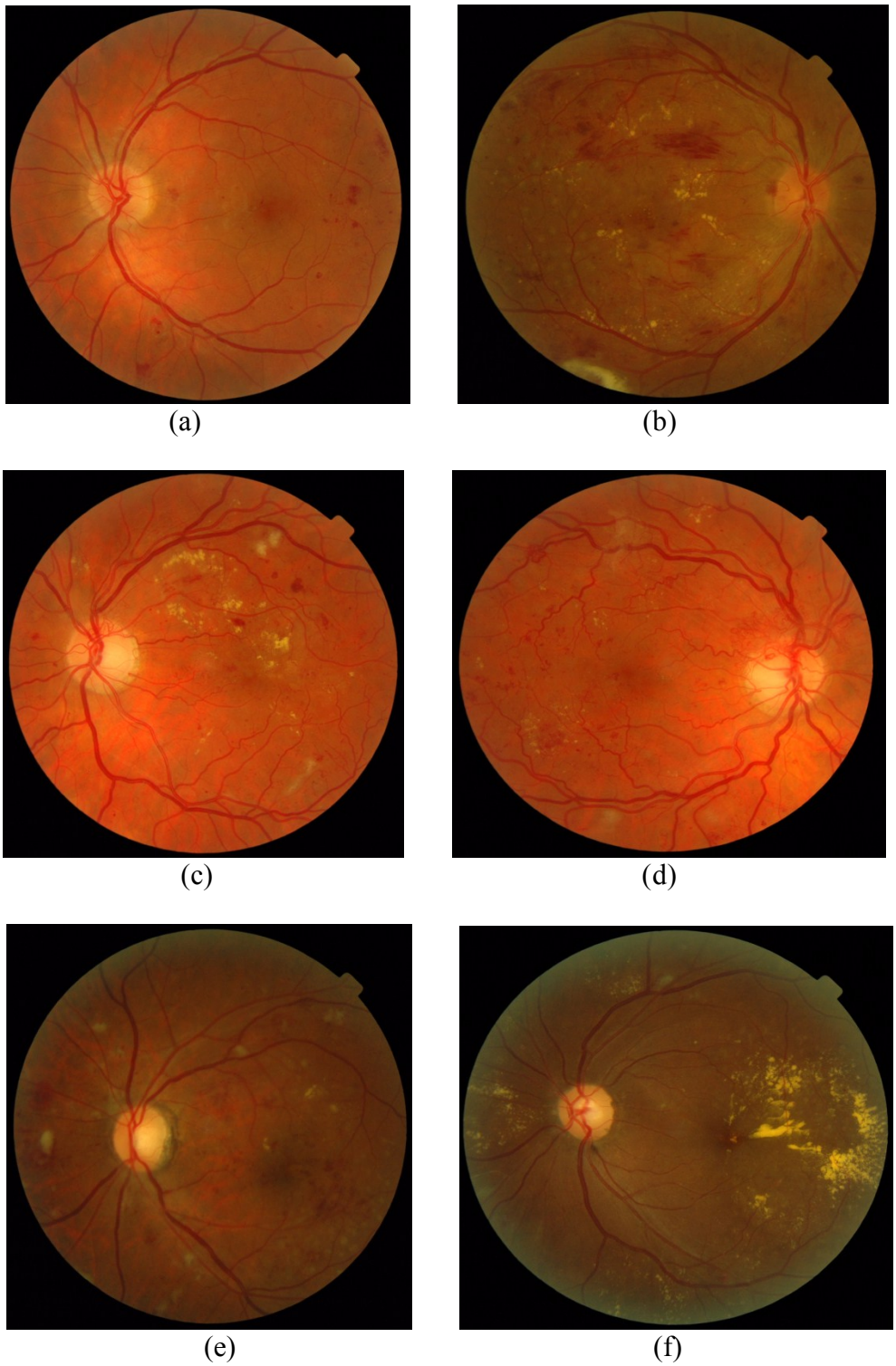
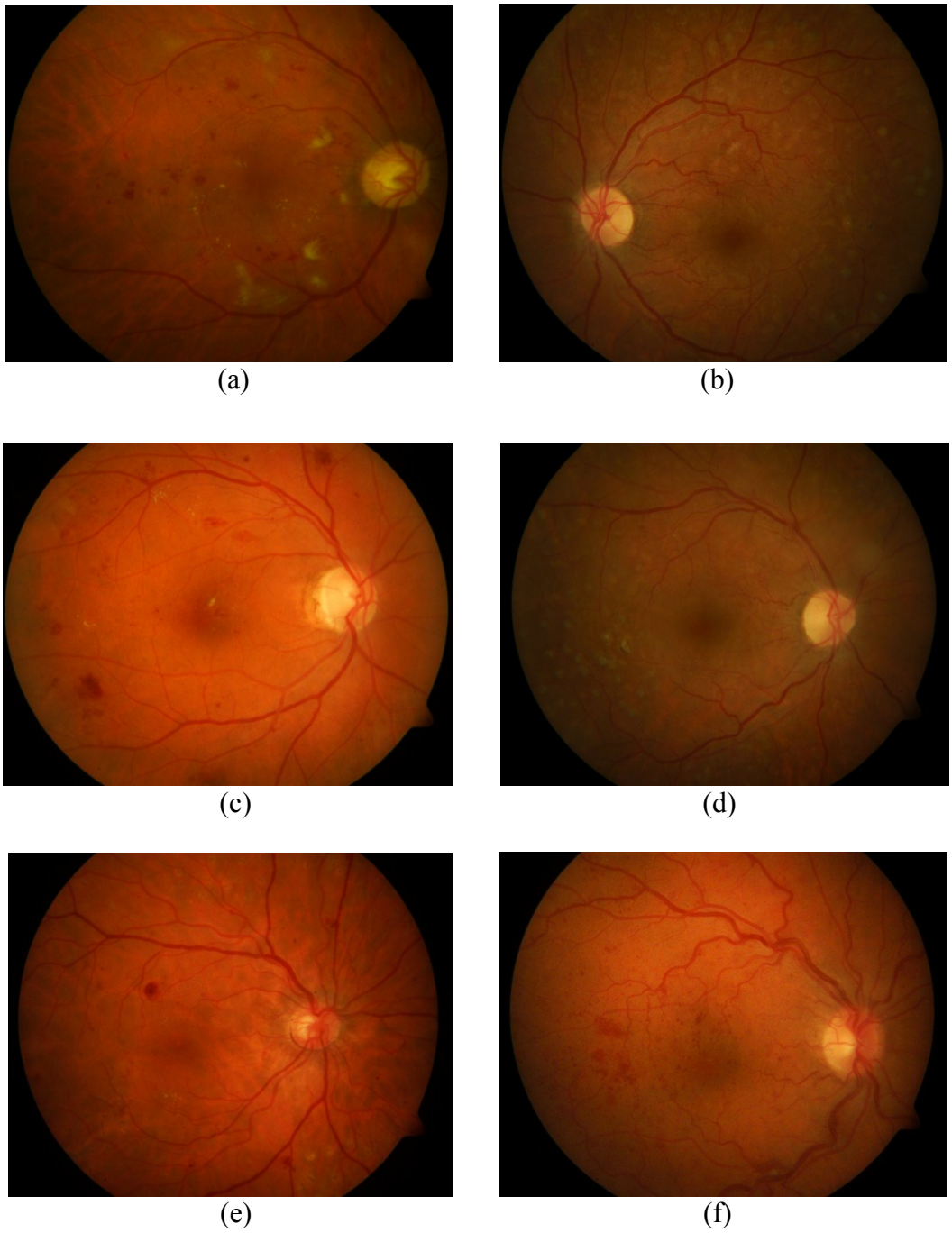


Fig. 30. Top 5 Retrieved Images for a Query Normal Image. (a) is a Query Normal Image and (b)-(e) are Retrieved Normal Images. (f) is an MA Image.



(a) (b) (c) (d) (e) (f)

Fig. 31. Top 5 retrieved images for a query NV image. (a) is a query NV image and (b)-(f) are retrieved NV images



(a) (b) (c) (d) (e) (f)

Fig. 32. Top 5 Retrieved Images for a Query NV Image. (a) is a Query NV Image and (b), (d) and (e) are Retrieved NV Images. (c) and (f) are MA Images

## 4.5 DR Retrieval System

It is necessary to develop an online system for on-the-field testing of the proposed algorithm. Thus in the final phase of this work a system is proposed which also has a graphical user interface making it easier for an ophthalmologist to operate with. Once an image is captured with a fundus camera, the image is loaded onto the software. Upon command to retrieve images, the software retrieves top 5 clinically relevant images using the proposed algorithm from an archived database along with patient details. This gives additional assistance to ophthalmologist that helps him/her in diagnosis. The first version of such a system is shown in Fig. 33. In the future, this system can be made more sophisticated and both classification and retrieval of DR images can be included.

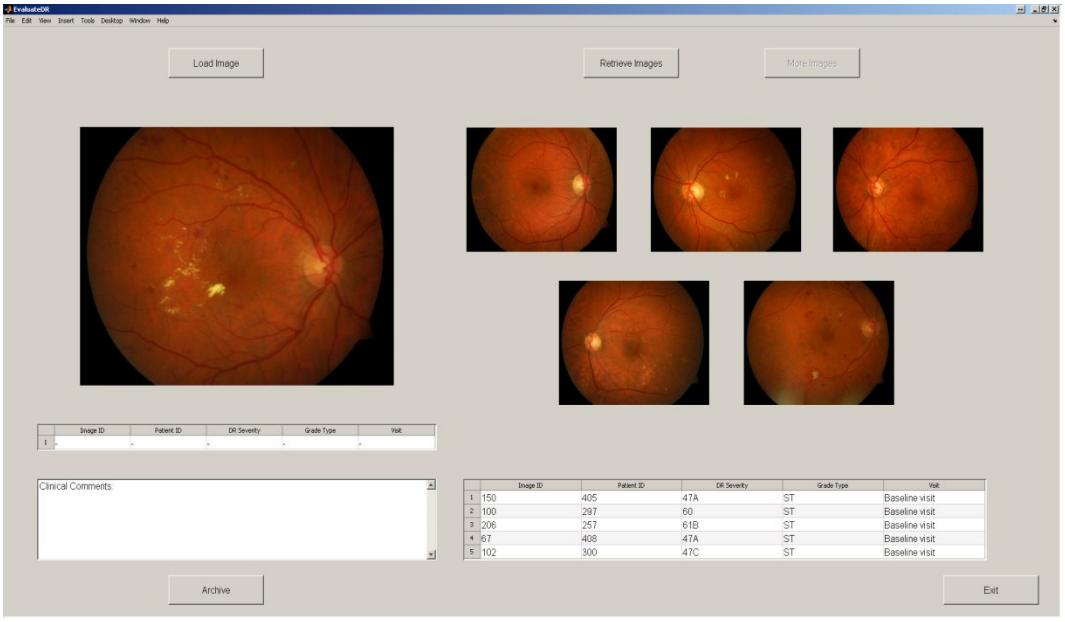



Fig. 33. Design of a Simple DR Retrieval System

# 5. CONCLUSION

This thesis proposes modified color correlogram features which are spectrally-tuned towards DR images. Based on these features and multiple instance learning approach, a classification framework is designed which is able to outperform prior arts and commonly used image classification approaches. This shows that MIL approach is suitable for this problem. This DR classification system needs to be validated on a bigger dataset and then it could be very useful for reducing the load on the screening stage.

It is also necessary to explore and make use of the vast knowledge in archived DR images. Therefore, this thesis proposes a novel DR image retrieval approach based on color correlogram and steerable Gaussian filter features. To exploit the special structure of DR images, a new KNN approach called *rank-citation KNN is* proposed. The proposed retrieval framework achieves high accuracy while maintaining consistency in the retrieval results.

Finally, it is required to integrate all the above algorithms into a user-friendly system so that it will useful for ophthalmologists. Therefore, a primary design of a DR retrieval system is proposed. It takes a query image as an input and gives 5 visually similar and more importantly clinically relevant images as output along with the details of respective patients.

## REFERENCES


[1] M. Li, "Texture moment content-based image retrieval," in *IEEE international conference on multimedia and expo (ICME)*, 2007.

[2] Centers for Disease Control and Prevention, "Prevalence and incidence of diabetes mellitus in United States, 1980-1987.," vol. 39, pp. 809-812, 1990.

[3] A. Amos, D. McCarty and P. Zimmet, "The rising global burden of diabetes and its complications: estimates and projections to the year 2010," Diabetic medicine, vol. 14, no. S5, pp. S7-S85, 1997.

[4] X. Zhang, J. Saaddine, C. Chou, M. Cotch, Y. Cheng, L. Geiss, E. Gregg, A. Albright, B. Klein and R. Klein, "Prevalence of Diabetic Retinopathy in the United States, 2005-2008," JAMA: The Journal of the American Medical Association, vol. 304, no. 6, pp. 649-656, 2010.

[5] J. Kempen, B. O'Colmain, M. Leske, S. Haffner, R. Klein, S. Moss, H. Taylor, R. Hamman and others, "The prevalence of diabetic retinopathy among adults in the United States," Archives of Ophthalmology, vol. 122, no. 4, p. 552, 2004.

[6] "Early Treatment and Detection of Diabetic Retinopathy Study Group. Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs -An Extension of the Modified Airlie House Classification," EDTRS Report Nr. 10. Ophthalmology, vol. 98, pp. 786-806, 1991.

[7] D. C. a. C. T. R. G. a. others, "Effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: Diabetes Control and Complications Trial," J Pediatr, vol. 125, no. 2, pp. 177-188, 1994.

[8] S. Philip, A. Fleming, K. Goatman, S. Fonseca, P. Mcnamee, G. Scotland, G. Prescott, P. Sharp and J. Olson, "The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme," British Journal of Ophthalmology, vol. 91, no. 11, pp. 1512-1517, 207.

[9] G. Scotland, P. McNamee, S. Philip, A. Fleming, K. Goatman, G. Prescott, S. Fonseca, P. Sharp and J. Olson, "Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland," British Journal of Ophthalmology, vol. 91, no. 11,

pp. 1518-1523, 2007.

[10] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Transactions on medical imaging, vol. 8, no. 3, pp. 263-289, 1989.

[11] S. Lee, Y. Wang and E. Lee, "Computer algorithm for automated detection and quantification of microaneurysms and hemorrhages (HMAs) in color retinal images," in SPIE, 1999.

[12] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," IEEE Transactions on Biomedical Engineering, vol. 51, no. 2, pp. 246-254, 2004.

[13] C. Sinthanayothin, J. Boyce, H. Cook and T. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," British Journal of Ophthalmology, vol. 83, no. 8, pp. 902-910, 1999.

[14] H. Kalviainen, P. Hirvonen, L. Xu and E. Oja, "Probabilistic and non-probabilistic Hough transforms: overview and comparisons," Image and vision computing, vol. 13, no. 4, pp. 239-252, 1995.

[15] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," Medical Imaging, IEEE Transactions on, vol. 22, no. 8, pp. 951-958, 2003.

[16] M. Foracchia, E. Grisan and A. Ruggeri, "Detection of optic disc in retinal images by means of a geometrical model of vessel structure," IEEE Transactions on Medical Imaging, vol. 23, no. 10, pp. 1189-1195, 2004.

[17] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, E. Fletcher and L. Kennedy, "Lowell, J. and Hunter, A. and Steel, D. and Basu, A. and Ryder, R. and Fletcher, E. and Kennedy, L.," IEEE Transactions on Medical Imaging, vol. 23, no. 2, pp. 256-264, 2004.

[18] B. Lay, C. Baudoin and J. Klein, "Automatic detection of microaneurysms in retinopathy fluoro-angiogram," in Society of Photo-Optical Instrumentation Engineers (SPIE), 1983.

[19] T. Spencer, R. Phillips, P. Sharp and J. Forrester, "Automated detection and

quantification of microaneurysms in fluorescein angiograms," Graefe's archive for clinical and experimental ophthalmology, vol. 230, no. 1, pp. 36-41, 1992.

[20] T. Spencer, J. Olson, K. McHardy, P. Sharp and J. Forrester, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," Computers and biomedical research, vol. 29, no. 4, pp. 284-302, 1996.

[21] M. Cree, J. Olson, K. McHardy, P. Sharp and J. Forrester, "A fully automated comparative microaneurysm digital detection system," Eye, vol. 11, no. 5, pp. 622-628, 1997.

[22] B. Ege, O. Hejlesen, O. Larsen, K. Moller, B. Jennings, D. Kerr and D. Cavan, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," Computer Methods and Programs in Biomedicine, vol. 62, no. 3, pp. 165-175, 2000.

[23] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," Diabetic Medicine, vol. 19, no. 2, pp. 105-112, 2002.

[24] D. Usher, M. Dumskyj, M. Himaga, T. Williamson, S. Nussey and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening," Diabetic Medicine, vol. 21, no. 1, pp. 84-90, 2004.

[25] G. Gardner, D. Keating, T. Williamson and A. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool," British Journal of Ophthalmology, vol. 80, no. 11, pp. 940-944, 1996.

[26] A. Osareh, M. Mirmehdi, B. Thomas and R. Markham, "Automated identification of diabetic retinal exudates in digital colour images," British Journal of Ophthalmology, vol. 87, no. 10, pp. 1220-1223, 2003.

[27] N. Patton, T. Aslam, T. MacGillivray, I. Deary, B. Dhillon, R. Eikelboom, K. Yogesan and I. Constable, "Retinal image analysis: concepts, applications and potential," Progress in retinal and eye research, vol. 25, no. 1, pp. 99-127, 2006.

[28] W. Cai, D. Feng and R. Fulton, "Content-based retrieval of dynamic PET functional images.," IEEE transactions on information technology in

biomedicine., vol. 4, pp. 152-158, 2000.

[29] W. Chu, I. Leong and R. Taira, "A semantic modeling approach for image retrieval by content.," The VLDB journal-The international journal on very large databases., vol. 3, pp. 445-477, 1994.

[30] P. Kelly, T. Cannon and D. Hush, "Query by image example: the comparison algorithm for navigating image databases (CANDID) approach.," in Proceedings of the SPIE, 1995.

[31] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel and Z. Protopapas, "Fast and effective retrieval of medical tumor shapes.," IEEE transactions on knowledge and data engineering, vol. 10, pp. 889-904, 1998.

[32] Nah, Yunmook and P. C.-Y. Sheu, "Image content modeling for neuroscience databases.," in Proceedings of the 14th international conference on software engineering and knowledge engineering., Ischia, Italy, 2002.

[33] C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. Aisen and L. Broderick, "ASSERT: a physcian-in-the-loop content-based retrieval system for HRCT image databases.," Computer vision and image understanding., vol. 75, pp. 111-132, 1999.

[34] M. Goldbaum, N. Katz, S. Chaudhuri and M. Nelson, "Image understanding for automated retinal diagnosis," in Proceedings of the annual symposium on computer application in medical care., 1989.

[35] A. Hoover, V. Kouznetsova and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp. 203-210, 2000.

[36] A. Gupta , S. Moezzi, A. Taylor, S. Chatterjee, R. Jain, L. Goldbaum and S. Burgess, "Content-based retrieval of opthalmological images.," in International conference on image processing (ICIP), 1996.

[37] E. Chaum, T. Karnowski, V. Govindasamy, M. Abdelrahman and K. Tobin, "Automated diagnosis of retinopathy by content-based image retrieval," Retinal, vol. 28, no. 10, p. 1463, 2008.

[38] T. K.W., M. Abramoff, E. Chaum, L. Giancardo, V. Govindasamy, T. Karnowski, M. Tennant and S. Swainson, "Using a patient image archive to diagnose retinopathy," in 30th Annual International IEEE EMBS

Conference, Vancouver, British Columbia, Canada, 2008.

[39] M. Swain and D. Ballard, "Indexing via color histograms," in Third International Conference on Computer Vision, 1990.

[40] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," in Third IEEE workshop on Applications of Computer Vision, 1996.

[41] J. Han and K. Ma, "Fuzzy color histogram and its use in color image retrieval," IEEE Transactions on Image Processing, vol. 11, no. 8, pp. 944-952, 2002.

[42] J. Huang, S. Kumar, M. Mitra, W. Zhu and R. Zabih, "Image indexing using color correlograms," in IEEE computer society conference on computer vision and pattern recognition (CVPR), 1997.

[43] B. Ege, O. Hejlesen, O. Larsen, K. Moller, B. Jennings, D. Kerr and D. Cavan, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," Computer Methods and Programs in Biomedicine, vol. 62, no. 3, pp. 165-175, 2000.

[44] M. Swain and D. Ballard, "Indexing via color histograms," in International Conference on Computer Vision, 1990.

[45] O. Chapelle, P. Haffner and V. Vapnik, "Support vector machines for histogram-based image classification," IEEE Transactions on Neural Networks, vol. 10, no. 5, pp. 1055-1064, 1999.

[46] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[47] D. Lowe, "Distinctive image features from scale-invarient keypoints.," International journal of computer vision, vol. 60, pp. 91-110, 2004.

[48] A. Bovik, M. Clark and W. Geisler, "Multichannel texture analysis using localized spatial filters," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 55-73, 1990.

[49] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," IEEE transactions on pattern analysis and machine intelligence

(PAMI-special issue on digital libraries), vol. 18, pp. 837-842, August 1996.

[50] Q. Chen, X. Tai, Y. Dong, S. Pan, X. Wang and C. Yin, "Medical image retrieval based on sementic of neighborhood color moment histogram," in The 2nd international conference on bioinformatics and biomedical engineering (ICBBE), 2008.

[51] T. Dietterich, R. Lathrop and T. Lozano-Perez, "Solving the multiple instance problem with axis-parallel rectangles," Artificial Intelligence, vol. 89, no. 1-2, pp. 31-71, 1997.

[52] O. Maron and T. Lozano-Perez, "A framework for multiple-instance learning," in Advances in neural information processing systems, 1998.

[53] Q. Zhang and S. Goldman, "EM-DD: An improved multiple-instance learning technique," in Advances in neural information processing systems, 2001.

[54] J. Wang and J.-D. Zucker, "Solving the multiple-instance problem: A lazy learning approach," in 17th International conference of Machine Learning, 2000.

[55] Q. Zhang, S. Goldman, W. Yu and J. Fritts, "Content-based image retrieval using multiple-instance learning," in International conference on machine learning, 2002.

[56] S. Andrews, I. Tsochantaridis and T. Hofmann, "Support vector machines for multiple-instance learning," in Advances in neural information processing systems, 2002.

[57] O. Maron and A. Ratan, "Multiple-instance learning for natural scene classification," in International Conference on Machine Learning, 1998.

[58] C. Yang and T. Lozano-Perez, "Image database retrieval with multiple-instance learning techniques," in International conference on Data Engineering, 2000.

[59] E. Garfield and R. Merton, Citation indexing: Its theory and application in science, technology, and humanities, vol. 8, Wiley New York, 1979.

[60] T. Kauppi, V. Kalesnykiene, J. Kamarainen, L. Lensu, I. Sorri, H. Uusitaloh, H. Kalviainen and J. Pietila, "Diaretdb0: Evalutaion database and

methodology for diabetic retinopathy algorithms," 2006.

[61] T. Kauppi, V. Kalesnykiene, J. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kalviainen and J. Pietila, "Diaretdb1: Diabetic retinopathy database and evaluation protocol," Procdeeings of medical image understanding and analysis (MIUA), pp. 61-65, 2007.

[62] B. McCormick and M. Goldbaum, "STARE=Structered analysis of the retina: Image processing of TV fundus image.," 1975.

[63] Chang, Chih-Chung and C.-J. Lin, "{LIBSVM}: A library for support vector machines," ACM transactions on intelligent systems and technology, vol. 2, no. 3, pp. 27:1 - 27:27, 2011.

[64] R. Venkatesan, P. S. Chandakkar, B. Li and H. K. Li, "Classification of Diabetic Retinopathy Images Using Multi-Class Multiple-Instance Learning Based on Color Correlogram Features," in 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), San Diego, 2012.

[65] W. Freeman and E. Adelson, "The design and use of steerable filters," IEEE Transactions on pattern analysis and machine intelligence, vol. 13, pp. 891-906, 1991.

[66] P. Locher and C. Nodine, "Symmetry catches the eye," Eye movements: From physiology to cognition, pp. 353-361, 1987.

[67] L. Kaufman and W. Richards, "Spontaneous fixation tendencies for visual forms," Attention, Perception, & Psychophysics, vol. 5, no. 2, pp. 85-88, 1969.

[68] W. Richards and L. Kaufman, "“Center-of-gravity” Tendencies for fixations and flow patterns," Attention, Perception, & Psychophysics, vol. 5, no. 2, pp. 81-84, 1969.

[69] D. Reisfeld, H. Wolfson and Y. Yeshurun, "Context Free Attentional Operators: The Generalized Symmetry Transform," International Journal of Computer Vision, vol. 14, no. 2, pp. 119-130, 1995.

[70] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 959-973, 2003.

[71] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 959-973, 2003.

[72] B. Sigurbjornsson and R. Van Zwol, "Flickr tag recommendation based on collective knowledge," in Proceedings of the 17th international conference on world wide web, 2008.

[73] Z. Wang and B. Li, "Learning to Recommend Tags for On-line Photos," Social Computing and Behavioral Modeling, pp. 1-9, 2009.

[74] A. Kollias and M. Ulbig, "Diabetic retinopathy: early diagnosis and effective treatment," Deutsches Arzteblatt International, vol. 107, no. 5, p. 75, 2010.