Sādhanā Vol. 40, Part 6, September 2015, pp. 1715–1736. © Indian Academy of Sciences

Automatic segmentation of blood vessels from retinal fundus images through image processing and data mining techniques

R GEETHARAMANI and LAKSHMI BALASUBRAMANIAN*

Department of Information Science and Technology, Anna University, Chennai 600025, India
e-mail: rgeetha@yahoo.com; lakshmi@auist.net

MS received 16 June 2014; revised 24 March 2015; accepted 24 April 2015

Abstract. Machine Learning techniques have been useful in almost every field of concern. Data Mining, a branch of Machine Learning is one of the most extensively used techniques. The ever-increasing demands in the field of medicine are being addressed by computational approaches in which Big Data analysis, image processing and data mining are on top priority. These techniques have been exploited in the domain of ophthalmology for better retinal fundus image analysis. Blood vessels, one of the most significant retinal anatomical structures are analysed for diagnosis of many diseases like retinopathy, occlusion and many other vision threatening diseases. Vessel segmentation can also be a pre-processing step for segmentation of other retinal structures like optic disc, fovea, microneurysms, etc. In this paper, blood vessel segmentation is attempted through image processing and data mining techniques. The retinal blood vessels were segmented through color space conversion and color channel extraction, image pre-processing, Gabor filtering, image post-processing, feature construction through application of principal component analysis, k-means clustering and first level classification using Naïve–Bayes classification algorithm and second level classification using C4.5 enhanced with bagging techniques. Association of every pixel against the feature vector necessitates Big Data analysis. The proposed methodology was evaluated on a publicly available database, STARE. The results reported 95.05% accuracy on entire dataset; however the accuracy was 95.20% on normal images and 94.89% on pathological images. A comparison of these results with the existing methodologies is also reported. This methodology can help ophthalmologists in better and faster analysis and hence early treatment to the patients.

Keywords. Blood vessel segmentation; Gabor filtering; classification; clustering; principle component analysis; STARE database.

*For correspondence

1. Introduction

The concept of Machine Learning has proved its usefulness in almost every field of concern. Mining the data to identify useful patterns from it, has profound effect in many domains such as Medicine and Biology, Social Networking, Transaction analysis, Software defect analysis and many others. In this paper, the power of data mining in analysing medical images is reported (Geetha Ramani *et al* 2012a). Retinal fundus images (Abràmoff *et al* 2010) are the major source for ophthalmologists in segmenting the anatomical structures of the retina viz. blood vessels, optic disc, macula and fovea (Niail *et al* 2006) to identify the eye pathologies related to retina. Characteristics of the blood vessel network play a very important role in identification of many diseases like Proliferative Diabetic Retinopathy, Hypertensive Retinopathy, Retinal Artery Occlusion, Retinal Vein Occlusion, etc. Segmentation and elimination of blood vessel tree from the fundus images also serve as a pre-processing step for localization of other structures such as optic disc, fovea, microneurysms, lesions, etc. The orientation and position of blood vessels also provide detection of left or right eye. Segmentation of retinal blood vessel network thus has many utilities and advantages. Manual tracing of blood vessel network is a cumbersome, time-consuming and error-prone task demanding expertise in this field of study. Hence automatic segmentation of the vessel network can be very useful to the ophthalmologists and thereby to the society.

In this paper, segmentation of blood vessels from the retinal fundus images is attempted through image processing and data mining techniques. Retinal image data, which is given as input for data mining process is considered as Big Data since every pixel forms a tuple. Blood vessel network is segmented through color space conversion and channel extraction, image pre-processing, Gabor filtering, application of Principle Component Analysis, Clustering, First-level-classification and Second-level-classification techniques. To the best of our knowledge, this is the first attempt to segment the blood vessels through this approach.

The rest of the paper is organized as follows. Section 2 presents the previous works in retinal blood vessel segmentation. Section 3 describes the proposed methodology to automatically segment the blood vessels. Section 4 highlights the experimental results. Section 5 concludes the paper and provides insight on future enhancements.

2. Previous work

Computational approaches are sought for the purpose of automatic retinal blood vessel segmentation. A retinal fundus image and its vessel segmented image are shown in figure 1 and figure 2 respectively.

A few attempts have been made in the past to automatically segment the retinal blood vessels. The works in the literature mainly adopt techniques of supervised and unsupervised learning, matched filtering, morphological operations, model based approaches and multi-scale operators. These works are briefly presented here.

Among the supervised and unsupervised learning techniques, artificial neural networks which determined the probability of the input pixel as either belonging to vessel or non-vessel through the adjusted weights obtained from training the network with sample instances (Akila & Kuga 1982). In 1999, application of Principal Component Analysis followed by Neural Networks achieved a specificity and sensitivity of 91% and 83.3%, respectively on 112 retinal fundus images (Sinthanauothin *et al* 1999). Adopting K- Nearest Neighbor algorithm on feature vector constituting the green channel image and the responses of the Gaussian matched filter with varying parameters, followed by thresholding yielded a binary vessel segmented image yielding

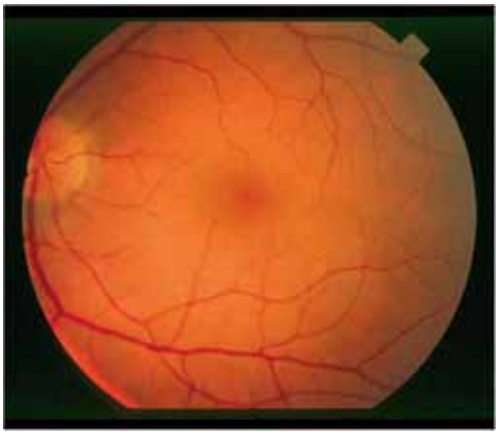

Figure 1. Retinal fundus image.

an average accuracy of 94.16% on DRIVE database images (Niemeijer *et al* 2004). Bayesian classifiers with class-conditional probability density function as combination Gaussian functions on feature vector comprising of pixel intensities and Gabor wavelet co-efficient obtained from multiple scales yielded an accuracy of 94.66% and 94.80% on DRIVE and STARE database images, respectively (Soares *et al* 2006). In 2010, a methodology was proposed that uses adaptive local thresholding to extract large vessels and Support Vector Machine on 12-dimensional feature vector from the residual image excluding the large vessels, to classify the thin vessel pixels (Xu & Luo 2010). This is followed by application of a tracking methodology to the thin vessel pixels thus forming the blood vessel network with an average accuracy of 93.20% on DRIVE database. In 2011, radial projection was used to extract the thin vessels and support vector machine learning on wavelet co-efficient on different scales for extracting major vessels (You *et al* 2011) achieving an accuracy, sensitivity and specificity of 94.34%, 74.10% and 97.51% on DRIVE images and 94.97%, 72.60% and 87.56% on STARE database mages, respectively.

Amongst the filtering techniques adopted for vessel segmentation, a few studies are briefed here. In 1989, a two-dimensional linear kernel with Gaussian profile which is rotated every 12 degrees, to extract the blood vessels in different orientation (Chauduri *et al* 1989) yielding an accuracy of 87.73% on DRIVE images was proposed. In 1998, local and region based properties were used in combination with threshold probing on matched filtered response image (Hoover *et al* 2000) yielding an average accuracy of 92.67% on STARE images. A hybrid method of matched filtering and ant colony system were proposed where the pre-processed image is given as input to both the matched filtering and ant colony in parallel and the results are combined to yield the segmented blood vessels (Cinsdikici & Aydin 2009). The hybrid technique achieved an accuracy of 92.93% on DRIVE images. In 2010, a methodology to reduce false positives (non-vessels as vessels) was proposed (Zhang *et al* 2010). The technique was extended to the matched filter approach by including its first order derivative. The vessel pixels will yield a high response to the matched filter but almost zero response to the first order derivate while the false step edges would yield high response to both the filters. This proposal achieved sensitivity, specificity and accuracy of 71.20%, 97.24% and 93.82% on DRIVE images and 71.77%, 97.53% and 94.84% on STARE images. In 2013, another attempt on ant colony system to identify vessel and non-vessel pixels was proposed (Asad *et al* 2013) with the system consisting of four phases namely, pre-processing, feature selection, ant colony-based segmentation and post-processing. The system achieved an accuracy of 91.39% on STARE images.

The concept of morphological operations is also extensively used for the purpose of vessel segmentation. A few works in this regard is briefly highlighted here.

In 2006, a methodology was adopted to extract the centerlines, enhancement of the blood vessels and segmentation of the vessels through Difference of Offset Gaussian filters, top hat operator and iterative region growing (Mendonca & Campilho 2006). This procedure yielded an average accuracy of 94.63% and 94.40% on STARE and DRIVE images, respectively. In 2008, an approach to segment blood vessels through enhancement by top hat operator and segmentation using fuzzy c-means clustering was proposed (Yang *et al* 2008). In 2011, employment of Fast Discrete Curvelet Transform, multi-structure mathematical morphology and adaptive connected component analysis to segment the blood vessels (Min & Mahloojifar 2011) achieved an accuracy of 94.58% on DRIVE database images. In 2012, a hybrid approach combining the centerline extraction and morphological bit plane slicing (Fraz *et al* 2012) yielded a specificity, sensitivity and accuracy of 71.52%, 97.69% and 94.30% on DRIVE database images and 73.11%, 96.80% and 94.42% on STARE database images, respectively. In 2012, yet another attempt was made to detect the blood vessels through Isotropic Undecimated Wavelet Transform, thresholding, morphological operations and spline fitting (Bankheard *et al* 2012) yielded an accuracy of 93.71% on DRIVE images.

Model based approaches were also investigated for the purpose of retinal blood vessel segmentation. These works are concisely highlighted here. In 2004, a model based method for blood vessel detection based on a Laplace and thresholding segmentation, followed by classification Vermeer *et al* (2004) reported an average accuracy of 92.87% on STARE database images. In 2008, Laplacian operator is employed to extract vessel-like object in which the false detections were pruned using the centerlines detected through the normalized gradient vector field (Lam & Hong 2008). The method yielded an accuracy of 94.74% on pathological images of STARE database.

Retinal blood vessel segmentation was also attempted through multi-scale operators. A few works in this field is presented here. Segmentation of retinal blood vessels based on multi-scale feature extraction is proposed in which feature information and spatial information are used for region growing (Martinez-Perez *et al* 2007). The approach yielded sensitivity, specificity and accuracy of 72.46%, 96.55% and 93.44% on DRIVE images and 75.06%, 95.69% and 94.10% on STARE images, respectively. In 2010, a multi-scale line-tracking procedure was proposed which

Figure 2. Segmented blood vessels of the retinal fundus image as in figure 1.

initiated from a small group of pixels obtained from brightness selection rule and terminated when a cross-sectional profile condition becomes invalid (Vlachos & Dermatas 2010). Median filtering and morphological reconstruction when applied to the initial vessel network, obtained through quantization of the multi-scale image map, yielded an accuracy of 92.90% on DRIVE images. In 2014, an approach was proposed to detect vessels through application of perspective transforms, K-means clustering and multi-scale line operator (Saffarzadeh *et al* 2014) achieving 94.83% on STARE database and 93.87% on DRIVE database.

Thus, the vast research that is being carried out in this field validates the significance of the current work. In this paper, image pre-processing, Gabor Filtering, Principal Component Analysis, Clustering and Ensemble of classification techniques have been adopted to automatically segment the blood vessels from the retinal fundus images which is presented in the next section.

3. Proposed methodology

The proposed methodology comprises of retinal fundus image data collection, image processing and data mining techniques. Image processing includes color space conversion and color channel extraction, image pre-processing, Gabor filtering and image post-processing, Data Mining techniques incorporated feature construction through principal component analysis, clustering, first level classification using Naïve–Bayes classification procedure, second level classification using C4.5 enhanced with Bagging technique and performance evaluation. The proposed framework is depicted in figure 3. The modules in the proposed framework are described below.

3.1 *Retinal fundus image data collection*

In this paper, segmentation of blood vessels from retinal fundus images is presented. STARE, a publicly available database is used for evaluation of the proposed approach. The STARE database (Goldbaum 1975) consists of 20 images, out of which ten are healthy images and the other ten are affected by various retinal diseases. The images were captured using TopCon TRV-50 fundus camera at 35 degrees field of view. The image resolution was 605×700 pixels with 8 bits per color channel. Thus, totally the image contains **423500 pixels**. Ground truth for blood vessel segmentation is available from two human observers. In the literature, the ground truth obtained from the first observer is considered for performance evaluation and hence we also have considered it as the ground truth. A sample image from the STARE database is shown in figure 1. The image processing techniques that contribute to the proposed framework is described in the subsequent sections.

3.2 *Color space conversion and colour channel extraction*

The STARE images are RGB images. RGB color model is not perceptually uniform and Euclidean Distances in 3D RGB space do not correspond to color differences as perceived by humans. Hence perceptually uniform color spaces namely L*a*b* color space and Gaussian color space along with the RGB color model have been adopted to extract the Gabor features. These color spaces are also very efficient in rotation invariant color texture analysis. The RGB color space consists of red, green and blue components. The L*a*b* color space (Brainard 1989) consists of L (lightness), *a* and *b* (color opponent) components. The RGB is converted to L*a*b* color space in accordance with the algorithm presented in figure 4.

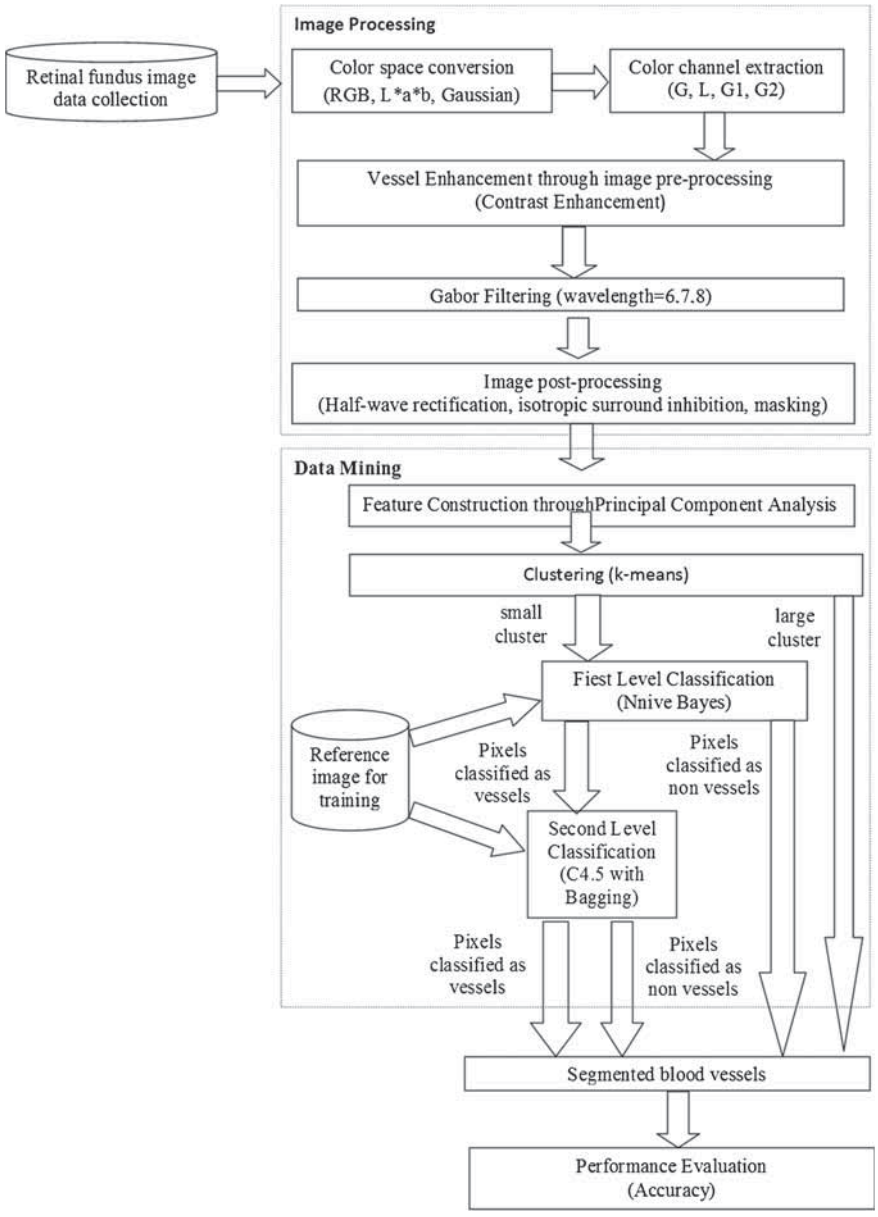

Figure 3. Proposed framework for retinal blood vessel segmentation.

Further, the RGB color space is converted to Gaussian color model. Gaussian model consists of É, $É_\lambda$ and $É_{\lambda\lambda}$ components and can be extracted from RGB color space using the following equation (Geusebroek *et al* 2001).

$$\begin{bmatrix} É \\ É_\lambda \\ É_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Step 1: Normalize the value of RGB between the range 0 and 1.

$$R = \frac{R}{255} \quad G = \frac{G}{255} \quad B = \frac{B}{255}$$

Step 2: Adjust the values of R, G and B as follows:

$$R = \begin{cases} \dfrac{R}{12.92}, & R \leq 0.04045 \\ \left(\dfrac{R+\alpha}{1+\alpha}\right)^{2.4}, & R > 0.04045 \end{cases}$$

$$G = \begin{cases} \dfrac{G}{12.92}, & G \leq 0.04045 \\ \left(\dfrac{G+\alpha}{1+\alpha}\right)^{2.4}, & G > 0.04045 \end{cases}$$

$$B = \begin{cases} \dfrac{B}{12.92}, & B \leq 0.04045 \\ \left(\dfrac{B+\alpha}{1+\alpha}\right)^{2.4}, & B > 0.04045 \end{cases}$$

Step 3: The adjusted RGB is converted to XYZ space with D50 illuminant according to the following transformation.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.6326696 & 0.2045558 & 0.1269946 \\ 0.2284569 & 0.7373523 & 0.0341908 \\ 0.0000000 & 0.0095142 & 0.8156958 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Step 4: Update X, Y and Z values as follows:

$$X = \frac{X}{X_n} \qquad Y = \frac{Y}{Y_n} \qquad Z = \frac{Z}{Z_n}$$

Where $[X_n \quad Y_n \quad Z_n] = [0.9642\ 1.0000\ 0.8251]$ **are the** tristimulus value for the reference white with d50 illuminant

Step 5: Adjust the values of X, Y and Z

$$X = \begin{cases} X^{1/3}, & X > 0.008856 \\ (7.767 * X) + \dfrac{16}{116}, & X \leq 0.008856 \end{cases}$$

$$Y = \begin{cases} Y^{1/3}, & Y > 0.008856 \\ (7.767 * Y) + \dfrac{16}{116}, & Y \leq 0.008856 \end{cases}$$

$$Z = \begin{cases} Z^{1/3}, & Z > 0.008856 \\ (7.767 * Z) + \dfrac{16}{116}, & Z \leq 0.008856 \end{cases}$$

Step 6: XTZ space is converted to the L*a*b* apace according to the following equations:

$$L^* = (116 * Y) - 16$$

$$a^* = 500 * (X - Y)$$

$$b^* = 200 * (Y - Z)$$

Figure 4. Conversion of RGB color space to L*a*b* color space.

The channels are then extracted from these color spaces. The channels that reveal high contrast are considered for further processing. The Green channel from the RGB color space, the Lightness channel from the RGB color space and É and $É_\lambda$ channels (denoted as G1 and G2, respectively in this paper) from the Gaussian color space are extracted and considered for further investigation, the process of which is discussed in the following section.

3.3 *Image pre-processing*

The quality of the extracted channel images are improved through contrast enhancement. The Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer *et al* 1987) is applied to enhance the blood vessels. It operates on small data regions rather than the whole image. Contrast

Step 1: The initial parameter values are set as follows

$$num\_of\_tiles = [M\ N] = [8\ 8]$$

$$num\_bins = 256$$

$$clip\_limit = 0.01$$

$$Distribution: Uniform$$

$$Alpha = 0.4$$

Step 2: Split the image into tiles. Pad the image if needed. (If rows and columns are not a multiple of 8 in this case)

Step 3: Compute the actual clip limit as follows

$$num\_of\_pixels\_in\_tile = M * N\ (64\ in\ this\ case)$$

$$\min\_clip\_linit = \frac{num\_of\_pixels\_in\_tile}{num\_bins}$$

$$actual\_clip\_limit = \min\_clip\_limit + round(clip\_limit * (num\_of\_pixel\_in\_tile - \min\_clip\_limit))$$

Step 4: Each tile is processed as follows to produce the corresponding grey level mapping:

Step 4a: Extract a tile and construct its histogram

Step 4b: Clip the histogram using the computed $actual\_clip\_limit$.

Step 4c: A mapping transformation function for the tile is formulated as follows:

$$sum\_of\_hist = cummulativesum(image\_histogram)$$

$$Range = \max(tile) - \min(tile)$$

$$scale = \frac{Range}{num\_of\_pixels\_in\_tile}$$

$$Mapping = \min(\min(tile) + sum\_of\_hist * scale, \max(tile))$$

Step 5: The final contrast enhanced image is formed through interpolation of grey level mappings as follow.

Step 5a: For each pixel find the four closest neighboring tiles that surround that pixel.

Step 5b: With the pixel intensity as an index, find its mapping at the four neighboring tiles.

Step 5c: Apply bilinear interpolation among these values to obtain the mapping of the currently considered pixel.

Step 5d: This pixel value forms the pixel intensity of the output contrast enhanced image.

Figure 5. Enhancement of extracted channel images through CLAHE.

of each small region is enhanced followed by employment of various interpolations for combining the neighboring small regions. The contrast of the extracted channel images are improved through CLAHE whose operation is given in figure 5.

After enhancing the images, Gabor filtering is applied to these four images G_{clahe}, L_{clahe}, $G1_{clahe}$ and $G2_{clahe}$, whose procedure is detailed in the next section.

Step 1: Initialize the filter parameters

$wavelength = 6, 7, 8; orientation = 0\ degree;$
$phase\ oofset = [-pi\ pi]; aspectratio = 0.5; bandwidth = 1;$
$NumberOfOrientation = 24;$

Step 2: Compute $sigma$ as follows:

$$slratio = \frac{1}{\pi}\sqrt{\frac{\ln(2)}{2}} * \frac{2^{bandwidth}+1}{2^{bandwidth}-1}$$

$$sigma = slratio * wavelength$$

Step 3: Calculate the size of 2D Gabor filter kernel as below.

$$n = 2.5 * \frac{sigma}{aspectratio}$$

Truncate the value to the lowest whole number
Size of the kernel matrix $=2n+1$

Step 4: Create a mesh grid $[x\ y] = meshgrid(-n:n)$.

Step 5: Change the direction of y: $y = -y$

Step 6: 2D Gabor filter kernel is constructed as follows.

$$f = \frac{2\pi}{wavelength}\quad b = \frac{1}{2*sigma^2}\quad a = \frac{b}{\pi}$$

$$xp = x\cos\theta + y\sin\theta \quad yp = -x\sin\theta + y\cos\theta \quad cosfunc = \cos\left((f * xp) - \pi\right)$$

$$result = a * e^{\left(-b*\left(xp^2+(aspectratio^2*yp^2)\right)\right)} * cosfunc$$

Step 7: The positive and negative values are normalized to ensure that the integral of Gabor kernel is 0.

$pos = sum(postive\ values\ in\ kernel)$
$neg = abs(sum(negative\ vaues\ in\ kernel))$
Divide every positive and negative value in the kernel with pos and neg respectively.

Step 8: Convolve the image with the filter kernel

Step 9: Repeat Step 2 to Step 8 by incrementing the orientation by 15 degrees till 180 degrees is reached

Step 10: The peak responses from every iteration is accumulated and the image is constructed.

Figure 6. Procedure for Gabor filtering.

3.4 *Gabor filtering*

The contrast enhanced images G_{clahe}, L_{clahe}, $G1_{clahe}$ and $G2_{clahe}$ are considered for further investigation. The blood vessels take the shape of Gaussian approximation. Hence Gaussian based filters can help in enhancing the blood vessels of the retinal image. 2-D Gabor filters (Fogel & Sagi 1989), which are sinusoidally modulated Gaussian functions, have been used to enhance the blood vessels. These filters have optimal localization in both frequency and space domains. The filter parameters greatly affect the performance of the Gabor Filters. Gabor filters work on the retinal images as stated in figure 6.

Each time, the procedure is applied, it results in three images (Gabor filtered image at wavelength $= 6, 7, 8$). Since the procedure is applied on four channel images, this step yields 12 Gabor images. The twelve Gabor images along with the four enhanced channel images are considered for further examination.

3.5 *Image post-processing*

The Gabor output yields strong response to the high frequency components of the image. In the retinal fundus images, along with the vessels, there also exists some variation in background, or the outer rim of the field of view, which are captured as high frequency components. Hence a few post-processing techniques namely; half wave rectification, surround inhibition and masking

Step 1: Find the maximum value of the input image matrix I

$$maxIntensity = \max(I)$$

Step 2: Calculate the percentage Value

$$perVal = maxIntensity * \frac{hwpercent}{100}$$

Step 3: Update the values in Matrix I.

$$I = \begin{cases} 0 & I < perVal \\ I & I \geq perVal \end{cases}$$

Figure 7. Halfwave rectification.

Step 1: Initiaize the parameters

$sigma$: standard deviation of gaussian factor (Figure 6)

$alpha$: suppresion of inhibition $= 1$

$k1$: factor for negative gaussian $= 1$

$k2$: factor for positive gaussian $= 4$

$inhibterm = -inf$

Step 2: Calculate inhibterm as follows:

$$\boldsymbol{inhibterm = max(abs(I, inhibterm)))}$$

Step 3: Construct inhibkernel

Step 3a: $diff\_of\_gauss = DoG(sigma, k1, k2)$

Step 3b: Negative values are made to zero

Step 3c: norm_L1=sum(diff_of_guass)

Step 3d: if $norm\_L1 \neq 1$, then $inhibkernel = diff\_of\_guass/norm\_L1$ else $inhibkernel = 0$

Step 4: Perform inhibition as follows:

Step 4a: $t = convolution(ihibterm, inhibkernel)$

Step 4b: $b = I - (alpha * t)$

Step 4c: $ResultantImage = b$ (with negative values set to 0)

Figure 8. Procedure for inhibition.

techniques were employed on the twelve Gabor responses to eliminate the false vessels as far as possible.

Firstly, Halfwave rectification is operated on the image based on a percentage value of the maximum intensity of the image. This process removes all the Gabor responses which are lesser than the specific percentage (10 in this case). Hence minor variations in texture of the images, which are mistaken for vessels, are eliminated. Setting too high a value for threshold will eliminate the true vessels (mostly the thinner vessels) whereas setting too low a value will lead to false spurs. Hence an optimal value is needed. The parameter (threshold $= 10\%$) was chosen based on experimental analysis. The procedure for Halfwave Rectification is shown in figure 7.

After the process of halfwave rectification, there exist still many false edges as their Gabor responses lie close to that of the true vessels. Hence surround inhibition (Grigorescu *et al* 2004) is applied to suppress the edges which act as noise, while leaving relatively unaffected the boundaries of vessels. Isotropic surround suppression is rotation invariant and does not get influenced by the orientation of the surrounding edges. The inhibition procedure is explained in figure 8.

Table 1. The feature vector obtained through image processing techniques.

Feature vector
The contrast enhanced Green channel pixel intensities
Post-processed Gabor response of green channel at wavelength 6
Post-processed Gabor response of green channel at wavelength 7
Post-processed Gabor response of green channel at wavelength 8
The contrast enhanced Lightness channel pixel intensities
Post-processed Gabor response of lightness channel at wavelength 6
Post-processed Gabor response of lightness channel at wavelength 7
Post-processed Gabor response of lightness channel at wavelength 8
The contrast enhanced G1 channel of Gaussian space
Post-processed Gabor response of G1 channel at wavelength 6
Post-processed Gabor response of G1 channel at wavelength 7
Post-processed Gabor response of G1 channel at wavelength 8
The contrast enhanced G2 channel of Gaussian space
Post-processed Gabor response of G2 channel at wavelength 6
Post-processed Gabor response of G2 channel at wavelength 7
Post-processed Gabor response of G2 channel at wavelength 8

Step 1: Calculate mean of every attribute.
Step 2: Subract attribute's mean from each of the data in the attribute
Step 3: Calculate the covariance matrix (16*16 matrix)
Step 4: Calculate Eigen vectors and eigen values for the covariance matrix.
Step 5: Choose necessary components
Step 6: Transpose the eigen vector and multiply it on the left of the transposed original data set,

Figure 9. Procedure for application of principal component analysis.

Finally, masking is performed to the resultant images after application of surround inhibition to remove the disturbances caused due to Gabor filtering outside the field of view of the fundus image. The pixels in the region outside the field of view are always considered as background pixels. Gabor filtering could have resulted in strong responses in these regions, since variations in the background and the edge of the field of view might be considered as high frequency components. Eliminating these responses would help to increase the accuracy of the proposed system. During masking, the processed image obtained from the previous step is superimposed with the mask of the original image so that the noise outside the field of view is suppressed.

The image data obtained from the post-processing step is given as input to the data mining procedures explained in the subsequent sections.

3.6 *Feature construction through principal component analysis*

The four contrast enhanced images and the twelve post-processed images are further analysed through data mining techniques to segment the blood vessels. The feature vector includes the sixteen images which are listed in table 1.

Each pixel contributes 16 features to the feature vector. In total there are 423500 pixels. Hence the data mining techniques are applied on a data with **423500 tuples and 16 attributes.** The ground truth forms the class field.

Input:
D: Dataset containing N instances
k: Number of clusters =2
m: number of attributes

Method:
Step 1: Randomly choose *(k=2)* instances from D as the initial cluster centers
Step 2: Assign each instance to the cluster to which it is most similar as follows:
Step 2a:For i=1 to N do
Step 2a1: $clusterlabel(i) = \min(\text{dist}(x_i, c_j))$

$where\ j = 1, 2\ (number\ of\ clusters)$

$x_i\ is\ the\ i^{th}\ instance\ from\ the\ dataset$

$dist(x_i, c_j) = euclidean\ distance\ between\ the\ two\ instances\ calculated\ as$

$$\sqrt{(x_{i_1} - c_{j_1})^2 + \cdots + (x_{i_m} - c_{j_m})^2}$$

Step 2a2: Update the cluster means
For all j=1,2
$average_j = Mean(x_i\ whose\ clusterlabel(i) = j)$
Step 3: For all j=1,2
$prevAverage_j = average_j$
Step 4: Iterate steps 2 and 3 until there is no change between $prevAverage_j$ and $average_j$.

Figure 10. K-Means clustering algorithm.

Principal Component Analysis (PCA) (Jolliffe 1986) is applied to the feature vector to arrive at a new set of features. It is a statistical method that employs an orthogonal transformation to convert data with possibly correlated attributes into a set of values of linearly uncorrelated attributes called principal components. The number of principal components can be less than or equal to the number of attributes (in this case, number of components is equal to the number of attributes = 16). Principal Component analysis is applied to the feature vector according to the procedure shown in figure 9.

Subsequently, clustering and classification techniques are applied in succession on the new feature set and data obtained through application of PCA. The process of clustering is described in the next section.

3.7 *Clustering*

Clustering, an unsupervised learning algorithm groups a set of data values such that instances in the same group have more similar properties than the instances between the groups. With segmentation of vessels from the retinal image under consideration, the clustering can be viewed as a technique for grouping the pixels into two groups namely, vessel cluster and non-vessel cluster.

K-Means clustering algorithm (Lloyd 1982) with McQueen's method of average computation is found to outperform the other clustering algorithms in view of distinguishing vessels and non-vessels. K-Means clustering algorithm divides the instances into two groups with each group's centre being represented by the mean value of the attributes of the instances. Since only two groups (either a vessel or non-vessel) are possible, the number of cluster is assigned as two. The procedure of K-Means clustering is presented in figure 10.

The clustering algorithm results in two clusters with unequal size. Since the non-vessels occupy major part (around 90%) of the image and vessels occupy only less part (around 10%), it is acceptable to consider the cluster with greater number of instances as non-vessel cluster and the cluster with less number of instances as vessel cluster. The pixels in the non-vessel cluster are concluded as non-vessel pixels. The methodology further proceeds with the vessel cluster. Classification techniques are applied to the vessel cluster to further identify the non-vessel pixels that may be present in the vessel cluster.

3.8 *First level classification*

Supervised classification techniques predict the class of a new data instance based on the training set of data which comprises of data instances whose class label is known. Supervised classification technique (Geetha Ramani & Shomonna Gracia Jacob 2013b) thus demands training data to form the decision rules. One of the images from the dataset is used for training the classifier, while the pixel data from the vessel cluster is given for classification. PCA is applied to the original feature vector of the training image and the vessel cluster pixels before the process of classification. Naïve–Bayes Classification algorithm (John *et al* 1995) produced the best results in this regard to classify vessel and non-vessel pixels. It is a probability based algorithm applying Bayes' theorem with the assumption of strong independence between the attributes. The Naïve–Bayes Classification Algorithm is shown in figure 11.

The Naïve–Bayes classification gives a prediction of vessels and non-vessels. The pixels predicted as non-vessels are concluded as non-vessels. The pixels that are predicted as vessels are further examined for better results.

3.9 *Second level classification*

The pixels predicted as vessels during the first level classification are taken for a second level classification. The same image, which was given as training image to the first level classification is again given for training to the second level classification. PCA is applied to the training image data and the vessel pixel data prior to classification process. C4.5 algorithm (Steven 1994) was employed for this second level of classification. C4.5 constructs decision trees through top-down approach from a set of training data using the concept of information entropy. The performance of C4.5 algorithm was enhanced using Bagging Technique (Breiman 1996) that aids in increasing the stability and accuracy of the classification algorithms. Procedure for building C4.5 decision trees enhanced with bagging technique is presented in figure 12. When an instance is given for classification, each of the decision trees built from the bagging technique tries to classify the instance and the final classification is the class that gets the majority vote.

Step 1: Calculate the prior probability of the classes as follows:

$$p_k = \frac{number\ of\ instances\ with\ k\ as\ class}{total\ number\ of\ instances}$$

Step 2: Calculate the conditional average and standard deviation for every class for all the attributes

Step 3: Classification function for each class is as follows:

$$d(x) = \ln(P_x) + \sum_j \left(-\frac{1}{2 * {\sigma_{kj}}^2} {x_j}^2 + \frac{\mu_{kj}}{{\sigma_{kj}}^2} x_j - \left(\frac{{\mu_{kj}}^2}{2 * {\sigma_{kj}}^2} + \ln(\sigma_{KJ})\right)\right)$$

where k denotes the class and j denotes the attribute

Step 4: A new instance is classified as that class whose function is maximized by its feature vector (x_1,x_2).

Figure 11. Naïve–Bayes classification procedure.

Input:
D: Dataset containing N instances along with their class label
A: set of attributes
numModels: Number of models In the ensemble (10 in this case)

Method: C4.5withBagging(D,A,numModels)
Step 1: For I = 1 to numModels do
Step 2: Create Bootstrap samples D_i by sampling $D$with replacement.
Step 3: Call C4.5(D_i, A)

Method: C4.5(D,A)
Step 1: Create a node N.
Step 2: If all instances in D belong to the same class C, then
Return N as the leaf node labeled with class C.
Step 3: If A is empty, then
Return N as a leaf node labeled with the majority class in D.
Step 4: For all attributes a in A, compute gain ratio as follows:

$$GainRatio(a) = \frac{Gain\ (a)}{SplitInfo\ (a)}$$

Where $splitinfo_a(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} * log_2(\frac{|D_j|}{|D|})$

Step 5:Assign a_{best} = attribute with maximum gain ratio
Step 6: Label node N with a_{best} and let it test the splitting criterion.
Step 7: For each outcome j of the splitting criterion,
D_j = data instances in D satisfying outcome j.
Step 8: If D_j $is\ empty$, then
Attach a leaf labeled with the najority class in D to node N.
Step 9: Else, attach the node returned by recursively calling C4.5(D,A).
Step 10: return N.

Figure 12. Procedure for growing decision trees using C4.5 with bagging.

The prediction from the second level classification is considered as the final prediction. The results from clustering, first level classification and second level classification are combined to produce the segmented image.

3.10 *Performance evaluation*

The performance of the blood vessel segmentation techniques can be evaluated using accuracy, specificity, sensitivity (Geetha Ramani *et al* 2012b) and area under ROC curve (Geetha Ramani *et al* 2013a). In this paper, accuracy metric is used for the purpose of evaluation. In the view of blood vessel segmentation, True positive (TP) signifies vessels correctly predicted as vessels; True negatives (TN) signifies non-vessels correctly classified as non-vessels; False Positive (FP) denotes non-vessels misclassified as vessels and False Negative (FN) denotes vessels wrongly predicted as non-vessels. The performance metric is calculated as given in the following equation.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$

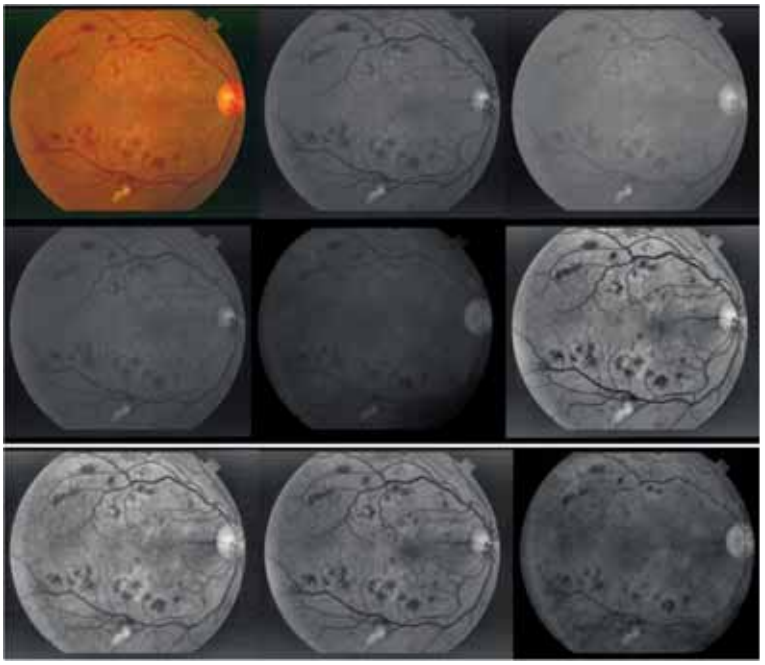

Figure 13. RGB image, its green channel, Lightness channel, G1 channel, G2 channel images and their contrast enhanced versions (in order from left to right).

The experimental results are reported with average accuracy. Results for the proposed framework are presented in the next section. After careful analysis and experimentation, the set of parameters which yielded the highest accuracy is considered. A detailed explanation of the various trials and experiments is mentioned in the subsequent section.

4. Experimental results

STARE database (Geetha Ramani & Lakshmi Balsubrmanian 2013c) is used for the evaluation of the proposed framework. The image processing techniques were implemented using Matlab r2008a and data mining techniques were implemented through Tanagra, an open source data mining tool. Experimental analysis is presented below.

4.1 *Experimental analysis on image processing techniques*

The RGB images were used for experimentation. Image processing techniques were applied to enhance the blood vessels on the retinal images so that it would be easier for segmentation. The channel images which showed better contrast were selected for processing. Hence the Green, Lightness, G1 and G2 channels were selected and processed. Contrast Limited Adaptive Histogram Equalization (CLAHE) was employed to these images so that the contrast was still increased. Unlike other contrast enhancement techniques, CLAHE limits the contrast and thus avoids over emphasizing the noise present in the image. Sample images of RGB, its Green channel, Lightness channel, G1 and G2 channels and its CLAHE enhanced images are shown in figure 13.

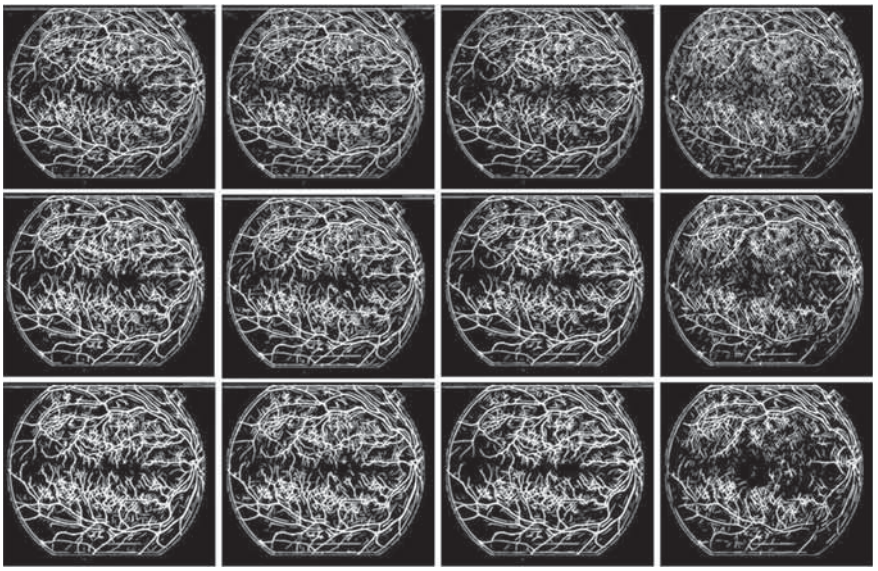

Figure 14. Post-processed images of Gabor reponses of green, lightness, G1 and G2 channels at wavelength 6, (first row), 7(second row) and 8 (third row).

Subsequent to the contrast enhancement on the images, Gabor filtering was applied to the images. Since blood vessels are piece-wise linear, they can be enhanced using a Gaussian based filter. Gabor responses were collected for every 15 degree change in orientation. The 15 degrees change is justified due to the responsiveness of the filter for 7.5 degrees in either direction. The peak responses during each orientation were included in the output image. The noise in the output image from the Gabor responses was eliminated through image post-processing namely, Half wave rectification, Isotropic surround inhibition and masking. The resultant post-processed images were given as input to the Data Mining processes. Sample images of the post-processed Gabor reponses are depicted in figure 14.

4.2 *Experimental analysis on data mining techniques*

The post-processed images are given as input to the data mining processes. Different combinations of techniques were attempted.

Initially, Gabor images alone were considered and given as input to the processes. Due to the noise in the Gabor responses even after the post-processing, the segmentation accuracy was not appreciable. Then the contrast enhanced images were also given in addition to the Gabor responses. The intensities were given with and without the application of PCA. Accuracy yielded through PCA features was higher than that of the original features. Then, data mining methodologies were employed with significant principal components and all the components. Again, the results with all the components achieved better. Hence all the sixteen components were included for the mining process.

The segmentation of retinal blood vessels was then attempted through classification and clustering techniques separately. The classification was attempted with the first image as training and the second image as test. Classification algorithms such as C4.5, Random Tree (Gall & Jean- Francois Le 2005) and Naïve–Bayes Classification were attempted. All the three achieved

promising training accuracies above 99%. But their performances in predicting the vessels and non-vessels of all the instances from a test image were around 85 to 90% with performance of C4.5 ranking the highest followed by Naïve–Bayes and Random Tree classification.

Clustering techniques were then attempted individually to a single image. Grouping of vessel and non-vessel pixels were tried using K-Means clustering and EM clustering. When the clusters were compared with the ground truth, it was noticed that K-Means clustering grouped the pixels as vessels and non-vessels better than the EM clustering technique. Among the variations of K-Means clustering algorithms, Forgy and McQueen approach of average computation was analysed on the data. Forgy approach computes the mean of the cluster after assigning all the instances to a cluster while the McQueen approach computes the average after assigning every instance in the data to a cluster. It was observed that the McQueen approach yielded better result than the Forgy approach.

When analysing on the number of clusters, attempt was made with varying number of clusters so that they can be further combined and/or processed to arrive at better segmentation results. But the pattern of clusters differed for each image and hence a generalized methodology could not be arrived at. Hence the number of clusters was fixed to two. Uniformly in all images, thus yielded two unequal clusters with one having more than 75% of the instances and the other having less than 25% of the instances. Therefore, the cluster with greater number of instances was assumed as non-vessel cluster and the cluster with less number of instances were assumed as vessel-cluster. Best results were achieved when the algorithm was run for 5 iterations and 20 trials. The trial which yielded the maximum R-square was considered as the final grouping by the clustering algorithm. However, the accuracy of only about 80% was reached. Hence, combination of clustering and classification techniques was attempted for segmentation.

The methodology involving clustering followed by classification was tried. The first image was the training image and the other images were given for class prediction. Analysing the number of vessels which were wrongly grouped in the non-vessel cluster, it was only around 5400 (1.5%) pixels on an average. Hence it was decided not to give the non-vessel cluster for classification and thus the instances in the non-vessel cluster were concluded as the non-vessel pixels.

When the vessel cluster was given as input to the classification procedure, the performance of the classification algorithms proved to increase the accuracy. This is attributed to the fact that vessels occupy only 10% of the image while the vessel cluster consists of pixels which occupy around 25% of the image. Hence false detections (non-vessel grouped as vessel) in the vessel cluster out-numbered the false detection (vessels classified as non-vessels and non-vessels classified as vessels) produced from the classification algorithms. The increase in accuracy is given in table 1. During this stage, Naïve–Bayes Classification yielded better results than the other two procedures with less computation time. This technique of clustering followed by classification on vessel cluster reached an accuracy of about 93%. To improve the accuracy still further, a second level of classification was adopted. It is to be noted that, during every step PCA was applied to the data.

The first level classification yielded increased number of false positives. A more number of non-vessels still were labelled as vessels. Hence an attempt to further classify the vessel labelled pixels was made. Moreover, the pixel data of this problem is biased towards non-vessels as its cardinality is extremely high when compared to that of the vessels. Hence, during every step (clustering and first level classification) the non-vessels identified are eliminated to arrive at an unbiased data with almost equal number of vessels and non vessels. In most of the cases, the vessel class of the first level classification contained almost equal number of vessels and

Table 2. Performance of the proposed methodology on STARE database.

Image number	Clustering accuracy	First level classification accuracy	Overall accuracy
01		Training image	
02	80.03%	94.55%	96.00%
03	75.27%	90.23%	94.67%
04	80.70%	94.67%	94.57%
05	76.97%	90.75%	93.77%
06	75.64%	88.05%	92.83%
07	78.73%	91.17%	95.40%
08	76.43%	95.43%	96.63%
09	80.15%	94.22%	95.94%
10	77.67%	93.88%	94.77%
11	79.67%	92.74%	94.96%
12	81.40%	93.55%	96.06%
13	79.67%	92.74%	94.96%
14	79.35%	95.14%	95.14%
15	77.96%	93.96%	93.93%
16	77.92%	92.75%	92.71%
17	79.32%	93.38%	95.16%
18	80.75%	94.29%	96.25%
19	80.93%	96.53%	96.68%
20	80.83%	94.80%	94.50%
Average		95.05%	

non-vessels in it. An attempt to further classify this unbiased data was made to improve the classification accuracy.

Hence, the instances with vessel predictions from the Naïve–Bayes classification procedure were given to a second level classification. It was observed that C4.5 yielded good results during this step. Ensemble techniques were attempted to attain still better accuracies.

When attempting with boosting and bagging techniques, bagging technique which builds 10 models of C4.5 yielded the best results which are presented in table 2.

The results of blood vessel segmentation are reported in table 2. Table 2 gives (i) the accuracy contribution from clustering, which is calculated as the ratio of number of correctly assigned non-vessel pixels to the total number of image pixels. (ii) Accuracy improvement through first level classification, which is calculated as ratio of sum of correctly predicted non-vessels from clustering and first level classification to the total number of image pixels. (iii) Overall accuracy which is calculated as ratio of correctly predicted non-vessels from clustering, first and second level classification and the correctly predicted vessels from the second level classification to the total number of image pixels.

From table 2, it can be seen that the accuracy obtained through two-level classification outperforms the single level classification for retinal blood vessel segmentation in retinal fundus images of STARE database. However, during second level of classification, it should be noted that the correctly labelled vessel from the first-level classification can now be labelled as non-vessels. The chance of this possibility increases when the number of classification level increases. Hence, as a trade-off between accuracy and sensitivity, the analysis was stopped with two levels of classification.

Table 3. Comparison of perfoamance with the existing methodologies.

Methodology	Accuracy %
Proposed methodology	95.05
Zhang *et al* (2010)	94.84
Saffarzadeh *et al* (2014)	94.83
Soares *et al* (2006)	94.80
Lam & Hong (2008)	94.74
Mendonca & Campilho (2006)	94.63
Fraz *et al* (2012)	94.42
Martinez-Perez *et al* (2007)	94.10
Vermeer *et al* (2004)	92.87
Hoover *et al* (2000)	92.67
Asad *et al* (2013)	91.39
You *et al* (2011)	84.97

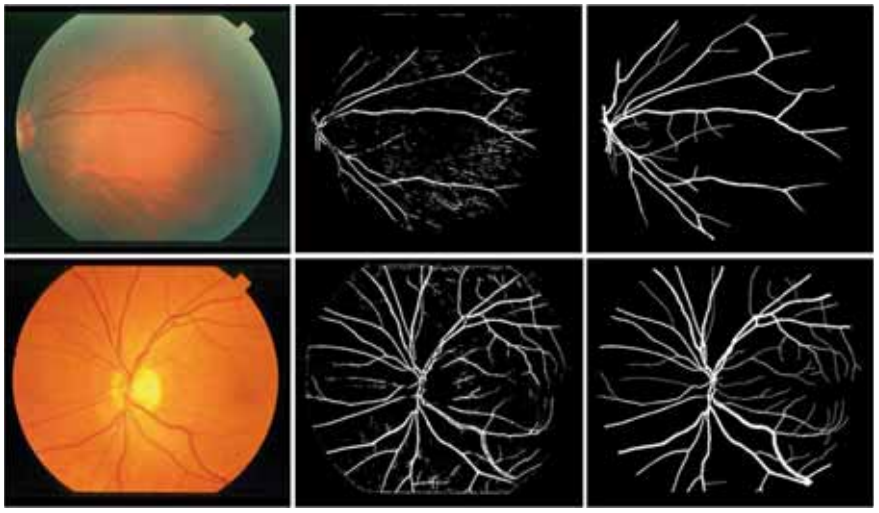

Figure 15. Sample images of diseased retina, segmented output and ground truth (first row) normal retina, segmented result and the ground truth (second row).

On further investigation, it was noticed that the average segmentation accuracy of healthy images was higher when compared to that of pathological images. The accuracy in detecting the vessels and non-vessels in healthy images was 95.20% and the average accuracy of vessel segmentation in pathological images was 94.89%. Though the training image itself was a diseased image, it should be noted that, each image was affected by a different pathology and hence the decrease in accuracy. The results were promising when compared to the existing blood vessel segmentation algorithms. Comparison of the results with the existing algorithms is reported in table 3.

On further analysing the outcome of the segmented images from the data mining process, it was found that many vessels remained unconnected to the vessel network. Hence a few simple connecting operations like bridging and cleaning were performed yielding a sensitivity of

68.59% and specificity of 97.17%. The sensitivity performance was 71.34% on healthy images and 66.11% for the pathological images. A maximum sensitivity of 81.46% was achieved. The higher specificity value indicates that the false vessels are greatly reduced in the proposed method. Moreover, the accuracy levels depicting the prediction of vessels and non-vessels are higher when compared to the other existing techniques.

Sample of the RGB images, segmented result of the proposed methodology and the ground truth is illustrated in figure 15.

The proposed methodology is thus efficient in segmenting blood vessels from the retinal fundus images. It can be of great use to the ophthalmologists in analysing the patient's retina. The methodology can perform still more efficiently when adopting parallel and distributed computing techniques.

5. Conclusion

The power of computational techniques viz. image processing and data mining are exploited for the purpose of retinal image analysis. The blood vasculature of the retinal fundus image is analysed to diagnose many sight threatening diseases. It is also useful in localization and segmentation of other retinal structures. In this paper, segmentation of retinal blood vessel is attempted through color space conversion and color channel extraction, Contrast Enhancement, Gabor Filtering, Image post-processing, Principal Component Analysis, Clustering, first and second level of classification through Naïve–Bayes and C4,5 with bagging, respectively. Mapping of every pixel to a feature vector makes it a Big Data. Results reported outperform the existing retinal segmentation algorithms. An average accuracy of 95.05% on the entire dataset, 95.20% on normal images alone and 94.89% on pathological images is reported. The insight for future directions includes studying the impact of the training image and refining the resultant segmented image obtained from data mining process to yield better performance measures. The proposed framework can assist the eye doctors in better retinal analysis, early diagnosis and hence better treatment to patients.

References

Abràmoff M D, Garvin M K and Milan Sonka 2010 Retinal imaging and image analysis. *IEEE T. Med. Imaging* 1(3): 169–208

Akila K and Kuga H 1982 A computer method for understanding ocular fundus images. *Pattern Recognit.* 15: 431–443

Asad A H *et al* 2013 An improved ant colony system for retinal blood vessel segmentation, *Proceedings of the 2013 Federal Conference on Computer Science and Information Systems*, 199–205

Bankheard P, Scholfield C N, McGeown J G and Curtis T M 2012 Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLoS One* 7(3): e32435

Brainard D H 1989 Callibration of computer controlled color monitor. *Color Res. Appl.* 14(1): 23–34

Breiman Leo 1996 Bagging predictors. *Mach. Learn.* 24(2): 123–140

Cinsdikici M G and Aydin D 2009 Detection of blood vessels in ophthalmoscope images using MF/ant (matched Filter/ant colony) algorithm. *Comput. Methods Programs Biomed.* 96: 85–95

Chauduri S, Chatterjee S, Katz N, Nelson M and Goldbaum M 1989 Detection of Blood Vessels im retinal images using two-dimensional matched filters. *IEEE Trans. Med. Imaging* 8: 263–269

Fogel I and Sagi D 1989 Gabor filters as texture discriminator. *Biol. Cybern.* 61(2): 103–113

Fraz N M, Barman S A, Remagnino P, Hoppe A, Basit A, Uyyanonvara B, Rudhicka A R and Owen C G 2012 An approach to localisze the retinal blood vessels using bit planes and centreline detection. *Comput. Methods Programs Biomed.* 108(2): 600–616

Gall and Jean- Francois Le 2005 Random trees and applications. *Probability Surveys* 2: 245–311

Geetha Ramani R, Lakshmi Balasubramanian and Shomona Gracia Jabob 2012a Automatic Prediction of Diabetic Retinopathy and Glaucoma through Image processing and Data Mining Techniques. *Proc. of Int. Conf. on Machine Vision and Image Processing,* 163–167

Geetha Ramani R, Lakshmi Balasubramanian and Shomona Gracia Jacob 2012b Data mining method of evaluating classifier prediction accuracy in retinal data, *Proc. of IEEE Int. Conf. on Computational Intelligence and Computing Research*, 426–429

Geetha Ramani R, Lakshmi Balasubramanian and Shomona Gracia Jacob 2013a ROC Analysis of classifiers in automatic detection of diabetic retinopathy using shape features of fundus images, *Proc. Int. Conf. Advances in Computing, Communications and Informatics*, 66–72

Geetha Ramani R and Shomonna Gracia Jacob 2013b Prediction of P53 mutants (multiple sites) transcriptional activity based on structural (2D&3D) properties. *PloS one* 8(2): e55401

Geetha Ramani R and Lakshmi Balsubrmanian 2013c Multi-Class Classification for Prediction of Retinal Diseases (Retinopathy and Occlusion) from Fundus Images. *Proceedings of ICKM'* 13: 122–134

Geusebroek J M, Van den Boomgaard R, Smeulders A W M and Geerts H 2001 Color Invariance. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(2): 1338–1350

Goldbaum M 1975 *Structured Analysis of the Retina. Available at* http://www.parl.clemson.edu/~ahoover/stare/index.html

Grigorescu C, Petkov N and Westenberg M A 2004 Contour and boundary detection improved by surround suppression of texture edges. *Image Vision Comput.* 22(8): 609–622

Hoover A D, Kouznetsovz V and Goldbaum M 2000 Locating Blood Vessels in retinal images by piecewise threshold probing of a matched flter respinse*, IEEE Trans. Med. Imaging* 19: 203–210

John, George H, Langley and Pat 1995 Estimating Continuous Distributions in Bayesian Classifiers, *in the Proc.of the 11th conf.on University in Artificial Intelligence* 338–345

Jolliffe I T 1986 Principal Component Analysis, *Springer-Verlag*, 487. ISBN 978-0-387-95442-4

Lloyd S P 1982 Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* 28: 128–137

Lam B S Y and Hong Yin 2008 A Novel Vessel Segmentation Algorithm for Pathological Retinal Images Based on the Divergence of Vector Fields. *IEEE Trans. Med. Imaging* 27(2): 227–246

Martinez-Perez M E, Hughes A D, Thom S A, Bharath A A and Parker K H 2007 Segmentation of blood vessels from red-free and fluoroscein retinal images. *Med. Image Anal.* 11: 47–61

Mendonca A M and Campilho A 2006 Segmentaion of retinal blood vessels by cobining the detection of centerlines and morphological reconstruction. *IEEE Trans. Med. Imaging* 25: 1200–1213

Min M S and Mahloojifar A 2011 Retinal Image Analysis using curvelet transform and multistructure elements morphology by reconstruction. *IEEE Trans. BioMed. Eng.* 58: 1183–1192

Niail Patton, Aslam T M, MacGillivray T, Deary I J, Dhillon B, Eikelbhoom R H, Yogesan K and Constable I J 2006 Retinal image analysis: Concepts, applications and potential. *Progr. Retinal Eye Res.* 25: 99–127

Niemeijer M, Staal J J, Van Ginneken B, Loog M and Abramoff M 2004 Comparative study on retinal vessel segmentaion methods on a new publicly available database. *SPIE* 648–656

Pizer S M, Amburn E P, Austin J D, Cromartia R, Gesselowitz A, Greer T, Romeny B T H and Zimmerman J B 1987 Adaptive histogram equalization and its variations. *Comput. Vision, Graphics, Image Process.* 39: 355–368

Saffarzadeh V M, Osareh A and Shadgar B 2014 Vessel Segmentation in Retinal Images Using Multi Scale Line Operator and K-Means Clustering. *J. Med. Signals Sens.* 4(2): 122–129

Sinthanauothin C, Boyce J F, Cook H L and Williamsom T H 1999 Automated localisation of the optic-disc, fovea and retinal blood vessels from digital colour fundus images. *Br. J. Opthalmol.* 83: 902–910

Soares J V B, Leandro Cesar R M, Jelinek H F and Cree M J 2006 Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* 25: 1214–1222

Steven L Salzberg 1994 C4.5: Programs for Machine Learning by J. Ross Quinlan, Morgan Kaufmann Publishers, Inc. 1993. *Mach. Learn.* 16(3): 235–240

Vermeer K A, Vos F M, Lemij H G and Vossepoel A M 2004 A model based method for retinal blood vessel detection. *Comput. Biol. Med.* 34: 209–219

Vlachos M and Dermatas E 2010 Multi-Scal Retinal vessel Segmentation using line tracking. *Comput. Med. Imaging Graphics* 34: 213–227

Xu L and Luo S 2010 A novel method for blood vessel detection from retinal images. *BioMed. Eng. Online* 9: 14

Yang Y, Huang S and Rao N 2008 An automatic hybrod method for retinal blood vessel extraction. *Int. J. Appl. Mathemat. Comput. Sci.* 18: 399–407

You X, Peng Q, Yuan Y, Cheung Y and Lei J 2011 Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognit.* 44: 2314–2324

Zhang B, Zhang L, Zhang L and Karray F 2010 Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Comput. Biol. Med.* 40: 438–445