CHAPTER 6

RESULTS AND DISCUSSION

346 million people worldwide have diabetes. More than 80% of diabetes deaths occur in low and middle income countries (WHO, 2011). Diabetic retinopathy(DR) is a major cause of blindness today. Blindness from DR is responsible for about 20 percent of new cases of blindness between the age group of 45 and 74. Laser photocoagulation can slow down the progression to blindness, if DR is detected in its early stages. However this is not an easy task because DR patients do not perceive symptoms until visual loss develops and this happens in the later stage of the disease, when treatment is less effective. In order to ensure that diabetic patients receive treatment on time, yearly fundus eye examination is advised by physicians. However growing incidents of diabetes increase the number of patients and as a consequence the number of images that need to be reviewed by experts. In addition, the high cost of examinations and the lack of specialists prevent many patients from receiving effective treatment. Due to these reasons, an expert system for the automatic detection of such anomalies has inspired much research in this direction.

Segmentation of major anatomical features such as optic disc, macula and retinal vasculature is of paramount importance in developing automated diagnosis expert system for DR. This feature segmentation is a key step in almost all algorithms used to identify fundus anomalies automatically. Furthermore automatic segmentation of these landmark features is important for the diagnosis of other ophthalmic pathologies also. Unfortunately most of

the algorithms used today for the detection of fundus features are computationally intensive, interdependent, and are less accurate, particularly in the presence of pathologies in the human retina.

6.1 ALGORITHM EVALUATION IN RETINAL DATABASES

The design, implementation and testing of algorithms which are capable of analyzing the retinal morphology or to diagnose pathologies require annotated data. In the retinal imaging field, there are few publicly available annotated datasets of retinal images which have different goals, characteristics, and level of completeness. However, the images by themselves are not enough to make a dataset particularly useful for algorithm development or method evaluation. The key aspect is the ground truth (GT) data which provides the gold standard against which the algorithms can be trained and evaluated. Whenever a single common dataset (with a common GT) is employed by different research groups, the advantages and drawbacks of the proposed methods can be easily measured and compared as demonstrated by Niemeijer et al. (2010). Luca Giancardo (2011) shows a comparison of the retina datasets most commonly used based on their GT and number of images and is shown in Table 7.1. The GT of the STARE database is available upon request. The GT of the ARIA database is not available for all the images. The number of experts varies between images. If there are multiple experts for a dataset, the GT for an image comes from a single expert hence the number of experts is still one for an image. Depending on the type of algorithm tested, different evaluation strategies need to be adopted.

A diagnostic system that outputs a binary healthy/unhealthy decision about an image can be evaluated against a GT (or gold standard) with sensitivity and specificity. The sensitivity is a ratio measured as the number of images classified as positive that effectively were positives (true positives (Tp)), divided by the “real” positives (i.e. false negatives (Fn) plus

Tp). The specificity is a ratio measured as the number of images classified as negative that effectively were negative (true negative (Tn)), divided by the "real" negatives (i.e. false positives (Fp) plus Tn). Both measures range between 0 and 1. These types of clear cut decisions are also defined as categorical.

Table 6.1 Utility of fundus image databases

DATA BASE	No. of Images.	Vessels segmentation GT	Optic nerve GT	Microaneurisms GT	Hard exudates GT	Other lesions GT	Diabetic Retinopathy GT	Macular Edema GT	No of Experts.
MESSIDOR (Messidor, 2010)	1200				√		√	√	1
DRIVE (Niemeijer et al. 2004)	40	√							1
ROC (Niemeijer et al. 2010)	100			√					4
DIARETDB1(Kauppi et al. 2007)	89			√	√	√			4
ARIA (ARIA 2006)	212	√	√						
HEI-MED (Giancardo et al. 2011)	169		√	√	√	√		√	1
STARE (Hoover et al. 2003)	81	√							

Another criteria to evaluate the categorical concordance between an algorithm and a gold standard is the Kappa value (k), which is widely used in the literature (Landis and Koch 1977). The equation for k is obtained as:

$$\Pr(a) = \frac{Tp + Tn}{No.of\ images} \tag{6.1}$$

$$\Pr(e) = \frac{Tp + Fn}{No.of\ images} X \frac{Tp + Fp}{No.of\ images} + \frac{Fp + Fn}{No.of\ images} X \frac{Fn + Tn}{No.of\ images} \quad (6.2)$$

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (6.3)$$

Some of the drawbacks in k is identified by Gwet (2002) by showing its sensitivity to trait prevalence in the subject population and proposes AC1 as an alternative:

$$P_{+} = \frac{(Tp \times 2 + Fn + Fp)}{2 \times No.of\ Images} \quad (6.4)$$

$$\Pr(eAc1) = 2 \times P + \times (1 - P +) \quad (6.5)$$

$$Ac1 = \frac{\Pr(a) - Pr(eAc1)}{1 - Pr(eAc1)} \quad (6.6)$$

Whenever the system tested still takes a binary decision but outputs a probability value rather than a categorical one, it can be better evaluated through a receiver operating characteristic (ROC) curve (Fawcett 2004), a graphical curve which represents the sensitivity (on the vertical axis) and the complement to 1 of the specificity on the horizontal axis. Since standard receiver operating characteristics (ROC) analysis is a frequently used tool to assess algorithm performance and it is well in accordance with the diagnostic procedures in medical practice, it was adopted in the framework to provide a means for statistical performance evaluation. Note that the ROC curve is not a single point (i.e. a single sensitivity/specificity pair), because the probability is thresholded at various points. A very compact representation of the ROC curve analysis is given by the area under the ROC curve (AUC), which is a

single number between 0 and 1, where 1 denotes perfect performances and 0.5 a completely aleatory set of decisions. The AUC is also defined as "the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one" (Fawcett 2006).

In some occasions, the use of the Tn is either not possible or not suggested. This is the case for the evaluation of lesion segmentation algorithms where the negative samples (that can be represented by pixels or other image areas) represent the vast majority of the image area. Hence, it is either not possible or misleading to calculate the specificity. A suited substitute is the positive predictive value (PPV), which is computed as Tp/(Tp+Fp). The free-response receiver operating characteristic (FROC) analysis (Metz 1986) is the equivalent to the ROC curve with the PPV instead of 1-specificity. Some authors (such as Niemeijer et al. 2010), plot directly the number of Fp instead of the PPV. In this work, both approaches are used.

Algorithm performance can also be measured with accuracy, a metric commonly employed in the field of science, engineering, industry and statistics especially for measurement systems. It evaluates the degree of closeness of measurements of a quantity to its actual (true) value.

6.2 THE OPTIC DISC DETECTION

Observing the state of the optic disc is not only useful in diagnosis of diabetic retinopathy, but also important for diseases, such as glaucoma, diabetic optic neuropathy and other optic nerve related pathologies. Since optic disc extraction is usually part of a more comprehensive diagnosis system, it is important that the method is efficient and accurately estimates the location of the optic disc. Based on the fact that optic cup part of the disc being the brightest part in the image, detection of optic disc boundary becomes important for the diagnosis of glaucoma. Changes in the optic disc

to cup ratio can be used as a measure of glaucoma. Since the OD may be easily confounded with large exudative lesions by image analysis techniques, its detection is also important to exclude it from the set of possible lesions. Moreover, OD detection is fundamental for establishing a frame of reference within the retinal image, and is thus important for any image analysis application. The detection of OD position is also a prerequisite for the computation of some important diagnostic indexes. Difficulty in finding the optic disc boundary is due to its highly variable appearance in retinal images. Optic disc boundary drawn manually by an expert is normally used as ground truth (gold standard) in this type of evaluation.

The performance of the OD-detection algorithm is evaluated based on the determined OD-location with regard to the manually segmented optic disc. To investigate inter observer differences in optic disc segmentation; two observers have segmented the optic disc in most cases. An optic disc is segmented by selecting three points on the border of the optic disc. Then a circle is drawn through these points. The methodology adopted by Frank ter Haar (2005) is followed in this evaluation.

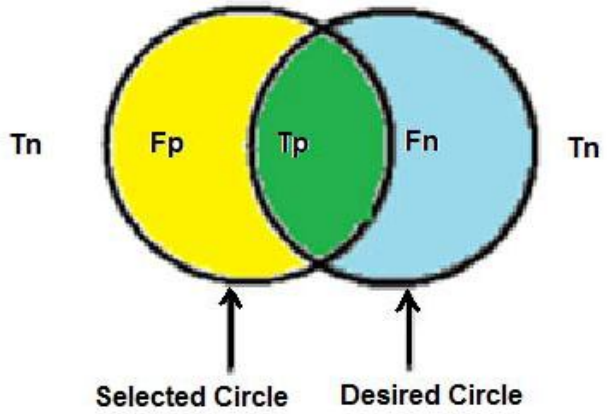

Figure 6.1 Performance measure

- T_P: The area contained in the selected circle as well as the desired circle.
- F_P: The area contained only in the selected circle.
- T_n: The area not contained in the selected circle or the desired circle.
- F_n: The area contained only in the desired circle.

The optic disc radii are collected from the segmentation results. To be able to draw conclusions from these radii, the average, the minimum, the maximum, the variance and the standard deviation are calculated. The segmented optic discs consist of a center point (odx, ody) and a radius. For each image, the center point location and the radius of the segmented optic disc will differ in case of different observers. The distances between the center point locations of the first and second observer are determined, and also the overlap between the segmented circles. The distance measure and the overlap measure used are defined in equations 6.7 and 6.8.

$$Distance = \sqrt{(Od1x - Od2x)^2 + (Od1y - Od2y)^2} \qquad (6.7)$$

$$Overlap = \frac{Tp}{(Tp + Fp + Fn)} \qquad (6.8)$$

The overlap between two circles is defined as the true positive (T_P) area divided by summed area of the true positives, the false positives (F_P) and the false negatives (F_n). That is, the overlap is defined as the shared area divided by the sum of shared and non-shared areas of the two circles.

The average optic disc dislocation , overlap and other statistics are noted. The maximal distance between two observed OD-centers does not exceed the minimal optic disc radius for all images, which means that the

center of the segmented optic disc of the first observer is always included in the optic disc segmentation of the second observer and vice versa. This means that both observers have a 100% successful localization of the optic disc if the other's segmentation is used as the golden standard. This is also confirmed by the overlap values for the dataset used.

For the automatic boundary provided by the algorithm, the centre is determined. In order to determine the performance of the algorithm, outcomes are classified as being a success or a failure. The classification performed is as follows:

- An outcome is classified as successful if the resulting OD-center lies within the manually segmented optic disc.

- An outcome is classified as failure if the resulting OD-center lies outside the manually segmented optic disc or if the outcome is inconclusive.

Whether the outcome is classified as successful or failure largely depends on which observer's segmentation is selected as the gold standard. Therefore, each outcome is classified according both observers' segmentations for each dataset individually.

As this is a broad classification, most automatic optic disc localization algorithms perform well in this measure. The only problem with some algorithms as it is not conclusive. On a sample set of randomly chosen 20 images from the DRIVE database and 120 images from MESSIDOR database, the result is as shown in Table 6.2.

Table 6.2 Performance of the algorithm in randomly selected images

Database	Pathology	No. of Samples taken	No. of Successful Segmentation	Percentage	Total percentage of Segmentation
MESSIDOR	No	87	87	100%	98.33%
MESSIDOR	Yes	33	31	93.9%	
DRIVE	No	16	15	93.7%	90%
DRIVE	Yes	4	3	75%	

It has been observed that for the MESSIDOR database images with no pathology, the success rate is 100%. The failure in the two images is due to the presence of severe large bright lesions such as exudates and cotton wool spots. While in the case of the DRIVE dataset, it is assessed that, in images with no pathology, the cause of failure is due to uneven illumination. This suggests that our illumination correction methodology has to be refined further. Also in the case of images with pathology, the cause of failure can be attributed to pathological lesions and uneven illumination. However it is worth to be noted that there were no false alarms. In all cases of failure (4 numbers in two datasets), the algorithm became inconclusive.

In another experiment with 120 images from MESSIDOR image database and 7 images from DRIVE database with various visual qualities and severe pathologies, the algorithm accurately identifies 109 optic discs out of the 120 images in MESSIDOR database (including training and testing sets) meaning an average accuracy of 90.8% and 5 optic discs out of the 7 images in DRIVE database meaning an accuracy of 71.4%. A summary of the test is depicted in Table 6.3.

Table 6.3 Performance of the algorithm in pathological images

Dataset	No. of images samples	No. of images with pathology	Success percentage	False alarm Percentage
MESSIDOR	120	120	90.8%	3.3%
DRIVE	7	7	71.4%	0%

The false alarm rate is found to be less than 5% even in the presence of a number of micro-aneurysms, hard exudates and intra-retinal hemorrhages, which is considered as one of the greatest advantage in any algorithms. Also, even though the success rate is less in DRIVE database, there were no false alarms. Several images with symptoms of non proliferative diabetic retinopathy and proliferative diabetic retinopathy from MESSIDOR dataset were also tested using the algorithm.

It is rather difficult to compare the performance of the algorithms for optic disc detection due to the non uniformity of datasets used and the availability and non uniformity of ground truth images. Also the computational speed largely depends on the computing platform used. An assessment of the timing profile of the algorithm using 'profiler' and 'timeit' is given in Appendix. Hence whenever computational speed is mentioned, the computing platform is also specified throughout the thesis. However an attempt was made to compare the performance of this algorithm on 81 images in a publicly available STARE database. The performance of the algorithm proposed by Hoover et al. (2003) for optic nerve localization is obtained from their website and compared with the proposed method . The optic disc or optic nerve head is the location where ganglion cell axons exit the eye to form the optic nerve. The optic disc represents the beginning of the optic nerve (second cranial nerve) and is the point where the axons of retinal

ganglion cells come together. The optic disc is also the entry point for the major blood vessels that supply the retina. The success in optic nerve localization is considered to be more or less equivalent to the localization of optic disc. The performance indices of the benchmarking algorithm used for comparison is reported in the website http://www.parl.clemson.edu/~ahoover/stare/nerve/index.html. The performance comparison is as shown in Table 6.4.

Table 6.4 A comparison of OD detection algorithms

Method	No. of images samples	No. of inconclusive cases	No. of failures	Success Percentage
Hoover et al. (2003)	81	2	7	91.1%
Proposed method	81	3	5	96.1%

As reported by the authors through their webpage, the algorithm by Hoover et al. (2003) fails in images im003, im007, im013, im020, im026, im041 and im139 while the inconclusive images were im019 and im027. The proposed method shows a success in im007 and im139, where the benchmarking algorithm fails. But the inconclusive cases has been increased to 3 (im019, im027 and im042) against a lower value of 2 (im019 and im027) for the benchmarking algorithm. The performance is graphically depicted in Figure 6.2 in terms of the reported accuracy.

In order to demonstrate the weakness of the algorithm, an example of a bad marking is given.

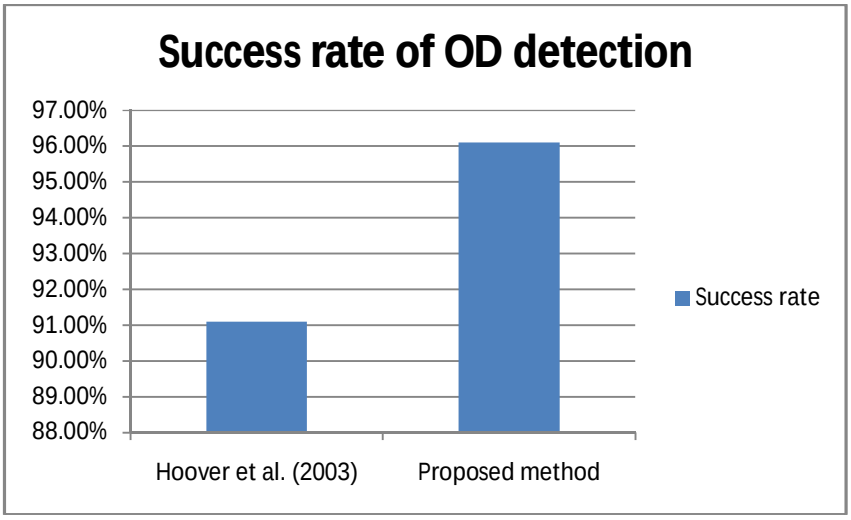

Figure 6.2 A Comparison of Accuracy of OD Detection

Figure 6.3 shows a bad marking of optic disc by the algorithm on an image with severe pathologies from MESSIDOR dataset. Figure 6.3 (a) shows the original image, Figure 6.3 (b) shows the extracted pixels corresponding to OD and Figure 6.3(c) shows the incorrect marking of OD by the algorithm.

Figure 6.3 (a) Original Image

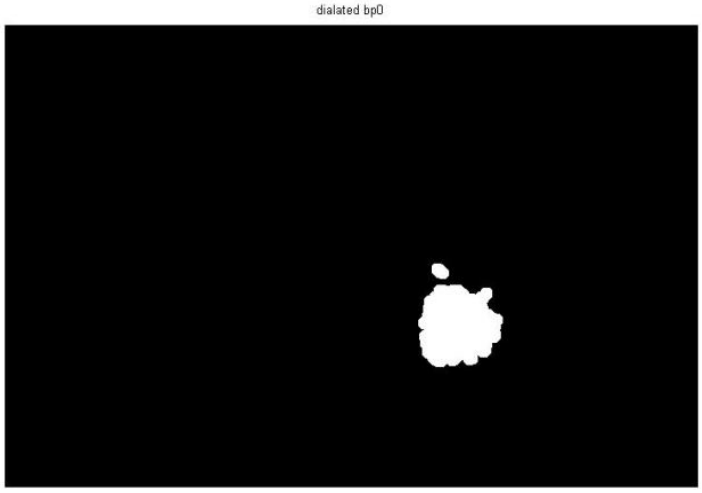

Figure 6.3 (b) Pixels corresponding to OD

Figure 6.3 (c) Incorrectly marked OD

While, Figure 6.4 depicts a correctly marked optic disc along with extracted image pixels corresponding to the optic disc. In this Figure 6.4 (a) shows the original image, Figure 6.4 (b) shows the extracted pixels corresponding to OD and Figure 6.4(c) shows the marked optic disc on the image.

Figure 6.4 (a) Original Image

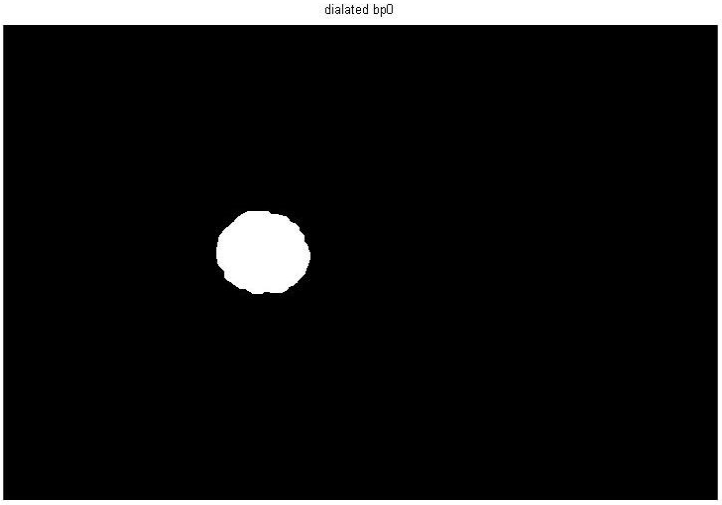

Figure 6.4 (b) Extracted pixels for OD

Figure 6.4 (c) Correctly marked OD

In the case of incorrect marking in the above image also, it can be classified as a success based on the criteria accepted such that the center of the marking lies inside the true OD area. The major problem here is another small marking near to the true OD. However, note that the methodology could avoid other lesions present in the image.

We have tested the algorithm on images with serious anomalies like *macular edema* . In all cases from MESSIDOR database, it is found to be useful for segmenting the optic disc with a good speed of only 11 seconds on a Core i3 machine with 4 GB of RAM.

The proposed method does not require any prior knowledge of retinal blood vessels and hence it is simple to implement and there is no error accumulation. The algorithm is found to be superior to the existing ones in terms of computational load and accuracy. It has high value in clinical practice for pre processing and automatic screening of early diabetic retinopathy. This algorithm demonstrates its strong ability to differentiate the

true optic disc from other bright lesions. The method works pretty well even when the input image is a low-contrast one. The experimental results demonstrate that the proposed algorithm is fast and robust especially with the presence of light lesions and pathologies.

6.3 SEGMENTATION OF MACULA AND FOVEA

The *macula lutea* or macula is a yellow spot near the center of the retina, with a diameter of about 1.5 mm and is often defined as having two or more layers of ganglion cells. Macula is a highly sensitive region of the retina responsible for the detailed central vision. The fovea is the macula center, which is the retina central zone. Unlike the peripheral retina, it has no blood vessels; instead, it has a very high concentration of cones, allowing for the appreciation of colour. It is the darkest part in most fundus images; in some images it is not obvious to human eyes due to bright lighting or being covered by lesions. The diagnosis of age related macular degeneration is typically undertaken through the inspection of macula. Also macula detection is inevitable for the detection of abnormalities such as hemorrhages. Macular oedema is a special case of diabetic retinopathy caused by leakage of blood vessels in the macular region.

The segmentation of macula and fovea are relatively less studied. The fovea appears as a large dark disc and is centered at the image approximately 2.5 times the optic disc diameter from the optic disc. The location of macula and fovea can be used to judge the severity of an abnormality detected in the retina. Abnormal lesions which are closer to the fovea (where our central vision resides) are more dangerous in terms of potential vision loss.

In our experiments to evaluate the algorithm, the following three parameters have been defined and a macula marked by an expert ophthalmologist is used as GT.

$$Sensitivity\ (Sn) = True\ positive\ rate(Tpr) = \frac{Tp}{(Tp+Fn)} \qquad (6.9)$$

$$Specificity\ (Sp) = True\ negative\ rate(Tnr) = \frac{Tn}{(Tn+Fp)} \qquad (6.10)$$

The Sn and Sp are obtained as follows: Both measures are evaluated using the four metric values; true positive (Tp), false positive (Fp), false negative (Fn), true negative (Tn). The sensitivity and the specificity are computed from Equation 6.9 and 6.10 respectively.

Table 6.5 shows the performance of the algorithm in MESSIDOR dataset taking an arbitrarily chosen 45 images with and without pathologies.

Table 6.5 Performance of the algorithm in MESSIDOR database

Macula Detection in MESSIDOR dataset	**Performance of the Algorithm (No. of cases)**
True detection (TD)	42
False alarm (FA)	1
Missed detection (MD)	2
Total no. of cases	45

The detection scheme provided an accuracy of 93.3%, while the false alarm rate is only 2%. The percentage of missed detection is as low as 4.4%.

In another experiment using the images from the DRIVE database with 22 image samples, the following values as demonstrated in Table 6.6 have been obtained. In the sample chosen, 15 images were normal ones and 7 were images with pathologies. For this dataset, the sensitivity and specificity are calculated using equation 6.9 and 6.10. The sensitivity values are found to be better than most existing methods but there exists a lower specificity value which is nowhere near to the state of the art methodologies in use today. This can happen due to the peculiarity of the GT and the number of samples chosen for the experiment.

Table 6.6 Performance of the algorithm in DRIVE database

Database	Pathologies	Sensitivity	No of samples
DRIVE	No	0.916	15
DRIVE	Yes	0.75	7

It is hard to compare the algorithm with the existing ones as the database and gold standard used may be different in different evaluations. Moreover, in most cases, the gold standard or ground truth images are obtained locally by an expert for evaluation which are not available publicly. So ,due to the non availability of a unified gold standard, a comparison of performance of different algorithms for macula detection often becomes difficult. Also, the computational speed mostly depends on the computing platform used and the size of the image. Hence, whenever computational speed is stated, the computing platform is also mentioned throughout the thesis. However due to less mathematical complexity and with the result we have obtained, it is evident that the proposed algorithm outperforms most existing ones. In a computing platform with Core i3 processor and 4 GB

memory, it took an average of 11 seconds to segment the macula for an image from MESSIDOR database.

Finally an ROC analysis was conducted using the proposed method for a small set of samples of size 20 images taken from MESSIDOR database with and without visible pathologies. The sensitivity and the specificity are computed from Equation 6.9 and 6.10 respectively. The obtained ROC curve is demonstrated in Figure 6.5.

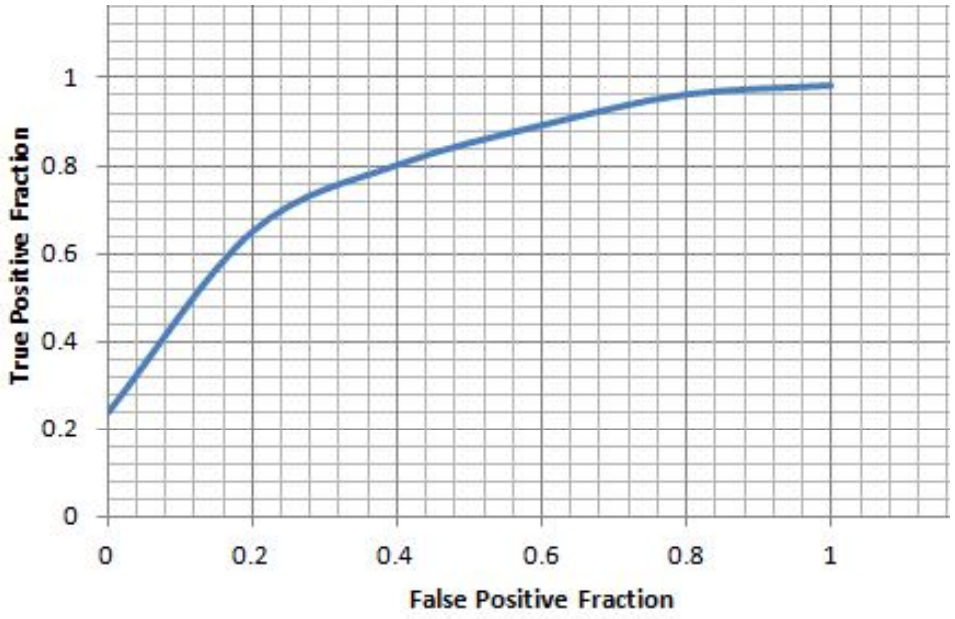

Figure 6.5 ROC curve for MESSIDOR database

In order to show the weakness of the method, a bad marking of the macula by the algorithm in an image from the MESSIDOR database is as shown in Figure 6.6 along with the extracted macula pixels and original image. Figure 6.6 (a) shows the original image while Figure 6.6 (b) and Figure 6.6 (c) shows the extracted pixels corresponding to the macula and the marking corresponding to macula on the original image.

Figure 6.6 (a) Original image

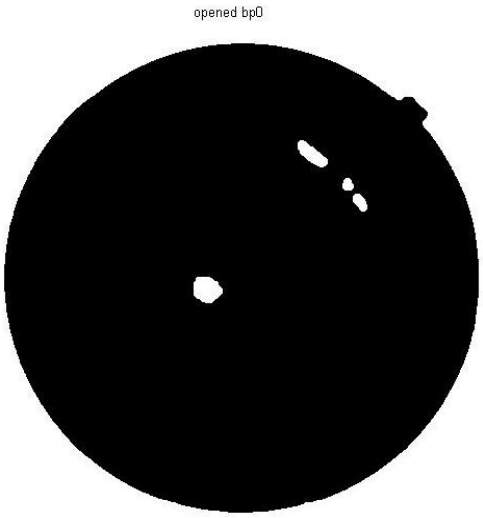

Figure 6.6 (b) Extracted macula pixels

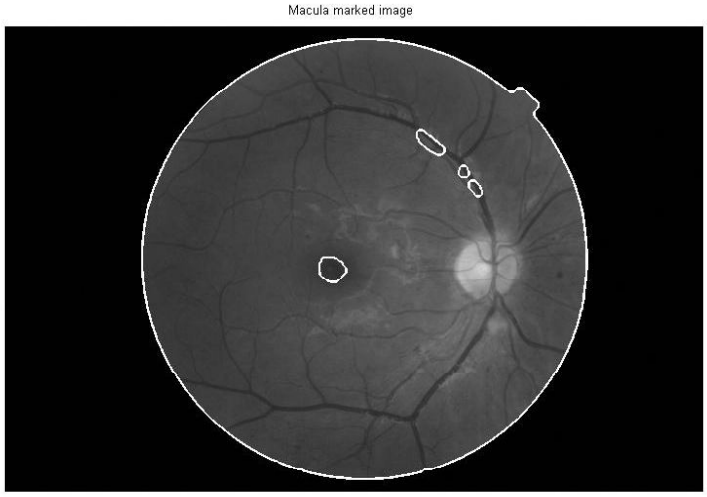

Figure 6.6 (c) Wrongly marked macula

It is to be noted that although there is some wrong marking by the automated method, the correct marking of the macula has not been missed in the outcome. This cannot be considered as a missed detection but some sort of ambiguity has been created by the algorithm.

Figure 6.7 (a) Original Image

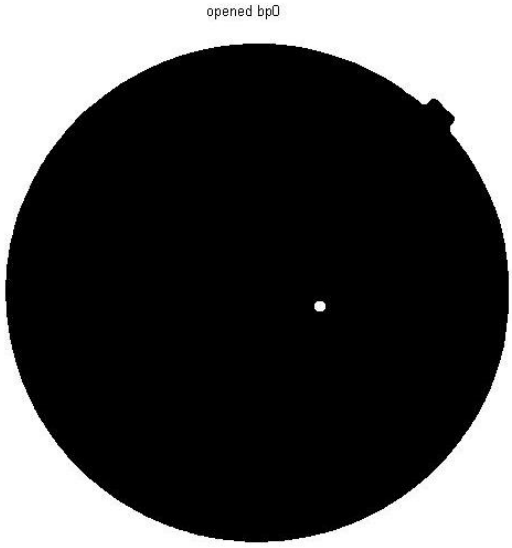

Figure 6.7 (b) Extracted Macula Pixels

Figure 6.7 (c) Correctly marked macula

Figure 6.7 shows a correct marking of the macula by the algorithm. Figure 6.7 (a) shows the original image while Figure 6.7 (b) shows the

extracted pixels and Figure 6.7 (c) shows the macula marked by the methodology.

The algorithm has been implemented by using *Matlab* version 7.9 (Release 2009 b) . Moreover, this method does not just detect the macula but can extract the shape and size of it also. It could differentiate dark lesions and hemorrhages present in fundus images of the human retina with that of macula. The results are promising even when it is applied to localize macula on images with varying lighting or exposure levels and with varying pathologies like macular edema. The main attraction of the proposed method is its simplicity, accuracy and saving in computational time. Moreover this algorithm does not require a prior knowledge of other retinal features for the detection of macula. An analysis of the timing profile of the algorithm using 'profiler' and 'timeit' is given in Appendix. The experimental results demonstrate that the proposed algorithm is fast and robust.

6.4 EXTRACTION OF RETINAL VASCULATURE

Blood vessel segmentation is a key step in almost all algorithms used to identify fundus features automatically. Furthermore blood vessel detection is important for automatic diagnosis of other ophthalmic pathologies also. Many automatic techniques for vessel identifcation techniques have been proposed in the literature, with various degrees of complexity and accuracy. The motivation that leads to the development of a new method can be summarized in the following points

- computational speed
- robustness

- Flexibility to accommodate a wide range of contrast in retinal images.

Three performance measures *viz* maximum average accuracy, area under the receiver operator characteristics (ROC) and the type of algorithm, whether it is an unsupervised algorithm or a supervised one are taken to evaluate the performance of the proposed method and compared it with known best algorithms. The performance indices of the algorithms used in this comparison are obtained from their corresponding published literature. These performance indices are then compared with the result we have achieved in similar environment using the proposed method of vasculature detection and is shown in Table 6.7. It is evident from the table given that the proposed method eclipses even the best known algorithms in terms of maximum average accuracy and area under ROC curve. Moreover the algorithm proposed is an unsupervised one.

Table 6.7 A comparison of vasculature detection algorithms

Method	Maximum Average Accuracy	Area under ROC	Comments
Human observer	0.9473	--	--
Staal et al.	0.9442	0.952	Supervised
Neimeijer et al.	0.9416	0.9294	Supervised
Kande et al.	0.9437	0.9515	Unsupervised
Zana et al.	0.9377	0.8984	Unsupervised
Jiang et al.	0.9212	0.9114	Unsupervised
Martnez-Prez et al.	0.9181	--	Unsupervised
Chaudhuri et al.	0.8773	0.7878	Unsupervised
Onkaew et al.	0.9388	0.8557	Unsupervised
Proposed Method	**0.9392**	**0.86805**	**Unsupervised**

A comparison of the maximum average accuracy of different methods is as shown in Figure 6.8.

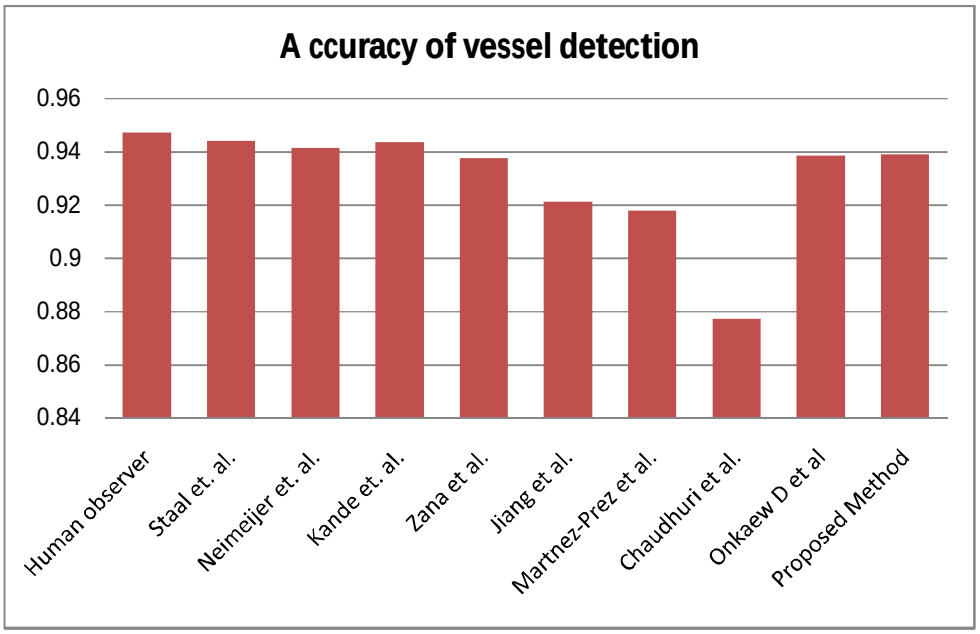

Figure 6.8 A Comparison of Accuracy of Vessel Detection

The first performance measure is receiver operator characteristics (*ROC*). An *ROC* space is defined by false positive rate (*Fpr*) and true positive rate (*Tpr*) as x and y axes respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). Since Tpr is equivalent with *sensitivity* (*Sn*) and *Fpr* is equal to (*1 - specificity*), the *ROC* curve is sometimes called the *sensitivity vs (1 - specificity)* plot. The *Sn* and *Sp* are obtained as follows: Both measures are evaluated using the four metric values– true positive (*Tp*), sum of pixel marked as vessel in both result and ground truth image; false positive (*Fp*), sum of pixel marked as a vessel in result image but not in ground truth image; false negative (*Fn*), sum of pixel marked as a background in result image but not in ground truth image; true negative (*Tn*), sum of pixel marked as a background in both result and ground

truth image. The sensitivity and the specificity are computed from Equation 6.9 and 6.10 respectively.

The algorithm has been implemented by using *Matlab* version 7.9 (Release 2009 b) and is found to be reasonably fast and accurate than the existing computationally intensive methods. The results are promising even when it is applied to segment the vasculature on images with varying lighting or exposure levels and with varying pathologies like macular edema.

The best possible prediction method would yield a point in the upper left corner or coordinate (0,1) of the ROC space, representing 100% *sensitivity* (no false negatives) and 100% *specificity* (no false positives). The (0,1) point is also called a perfect classification. The second is the area under ROC. The larger the area under the ROC curve, the greater the discriminating ability of the segmentation method. The third measure is maximum average accuracy (Maa). The accuracy of an image is calculated by taking the sum for the Tn and Tp divided by sum of the total number of nonvessel pixels (n) and total number of vessels (p) as illustrated in Equation 6.11. In our experiments, we used the manual segmentation by 1st observer of DRIVE database as a gold standard or GT for calculating all these three measures ROC, *area under* ROC, and Maa. Only pixels inside the field of view (FOV) are taken into account.

$$Maximum\ averageaccuracy = \frac{(Tn + Tp)}{(p + n)} \qquad (6.11)$$

The ROC curve for DRIVE database when the proposed method is employed is as shown in Figure 6.9. The area under the ROC curve is considered to be one of the key performance indicator for an algorithm. Note that the area under the ROC curve for the proposed method is one of the

maximum available today in the context of blood vessel extraction methodologies from fundus images of the human retina.

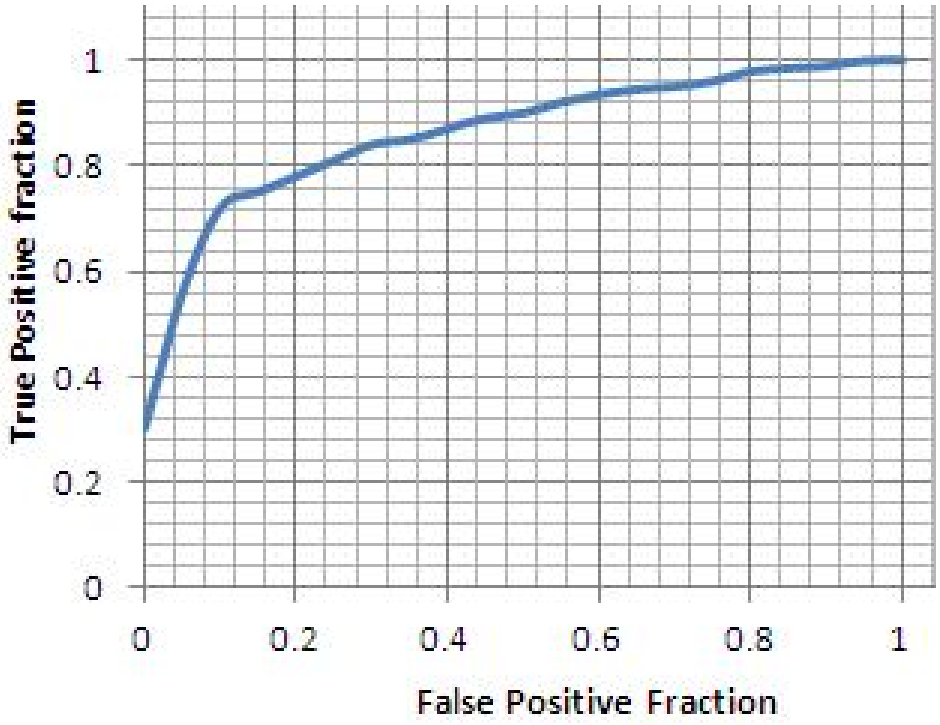

Figure 6.9 ROC curve for DRIVE database

The method used by Niemeijer et al. for comparison of vasculature extraction algorithms relies on determination of the Kappa(k) value as described by Landis and Koch (1977). Although this is not a robust measure of algorithm performance, for the sake of completeness it is given in Table 6.7. The limitations of this method have been pointed out in literature, mainly by Gwet (2002) by showing its sensitivity to trait prevalence in the subject population and he proposes AC1 as an alternative to this approach. However no attempt has been made to evaluate the algorithm using AC 1 in this work.

Table 6.8 Comparison based on Kappa value

Method	Kappa value
2nd Human observer	0.7589
Niemeijer et al.	0.7145
Zana et al.	0.6971
Jiang et al.	0.6399
Martinez-Perez et al.	0.6389
Chaudhury et al.	0.3357
Proposed method	**0.7216**

The data in Table 6.8 shows Kappa values obtained by numerous researchers in this field is drawn from the comparative study done by Niemeijer et al. along with the Kappa value calculated from the proposed method. As can be seen here, the Kappa values are far from ideal even in the case of the second human observer. However according to the criteria developed by Landis and Koch (1977), it does however indicate that this series of algorithms produce agreements with the human observer that are classified as "substantial" according to the strength of agreement standard developed by them. Most of the results were in the range 0.6 to 0.8, which is the range categorized to be "substantial".

Figure 6.10 shows a blood vessel segmentation performed by the algorithm on an image in MESSIDOR database along with the original image. Figure 6.10 (a) shows the original RGB image and Figure 6.10 (b) shows the extracted vasculature.

Figure 6.10 (a) Original image

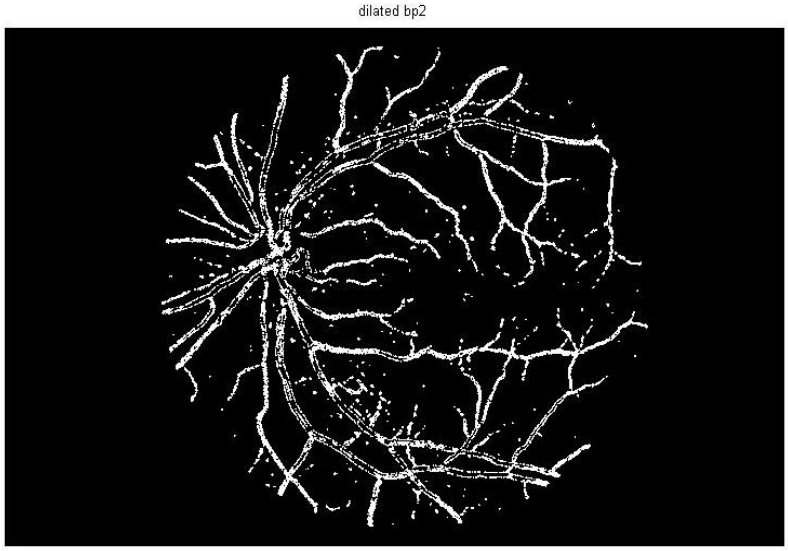

Figure 6.10 (b) Vasculature extracted by the algorithm

Figure 6.11 shows the result of segmentation using the DRIVE dataset. In the test dataset of DRIVE database, image no 16 is chosen for demonstrating the result.

Figure 6.11 (a) Original image

Figure 6.11 (b) Vasculature extracted by the algorithm

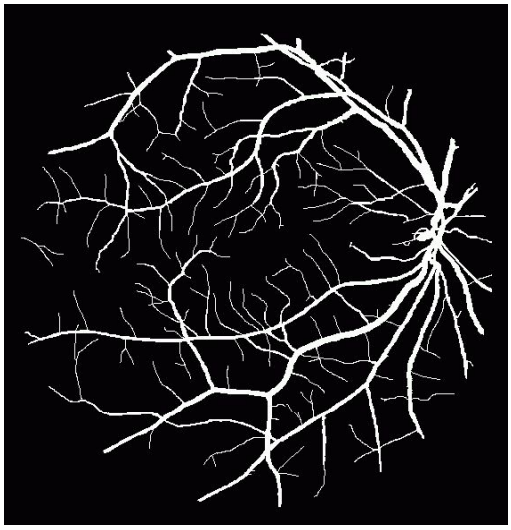

Figure 6.11(c) Gold standard

Figure 6.11 (a) shows the original image, Figure 6.11 (b) shows the result of automatic segmentation by the algorithm and Figure 6.11 (c) shows the manual segmentation by second observer (Gold standard or GT).

The proposed algorithm is simple and more efficient for automation. There is no mathematical complexity as in other methods and hence there is a significant improvement in computational time also. Moreover, this method does not just detect the vasculature but can extract the pixels corresponding to vascular tree also. The algorithm executed on a Core i3 system with 4 GB memory took only 14 seconds for the segmentation of an image from DRIVE database. An attempt has been made to assess the time complexity of the algorithm using 'profiler' and 'timeit'. The computational time largely depends on the image size and the computing platform. The timing profile analysis using a typical image can be seen in the Appendix. The algorithm has been evaluated on a subset of the MESSIDOR and DRIVE image databases with various visual qualities. It is found to be superior to the existing ones in terms of computational speed and accuracy. Several images with both eyes and with or without pathologies were also tested using

the algorithm. The false alarm rate is found to be very less even in low contrast images with multiple defects. Our algorithm has a very promising average accuracy as depicted in Table 6.7. The main attraction of the proposed method is its simplicity, accuracy and saving in computational load. Moreover this algorithm does not require a prior knowledge of other retinal features for the detection of vasculature .The method works pretty well even when the input image is a low contrast one. The experimental results demonstrate that the proposed algorithm is fast and robust.