Chapter 5

Heuristic Based Clustering For Macula Segmentation and Fovea Localization

Dr. Malpe Kalpana Devidas[1], Mr. Shewale Shubham Shivaji[2]

[1]Assistant Professor, [2]M.Tech Student
Department of Computer Science and Engineering, Gurunanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.

shubh.shewale18@gmail.com

ABSTRACT

Macula is a significant structural pattern responsible for high resolution vision. Existing approaches towards extraction of macula and fovea mostly involve segmentation of optic disc and identifying macula with optic disc as reference. This is a computationally complex and error prone task. This work attempts to utilize data mining techniques for this purpose. Since macula lacks well defined border definition and no expert opinion is provided in this regard, supervised classification techniques cannot be adopted. Unsupervised clustering technique is sought for this purpose. This chapter introduces an unsupervised clustering algorithm for segmentation of macula. The algorithm targets at incorporating a heuristic based on measures indicating the statistical distribution of data for selecting the initial cluster centers that play a significant role in performance of the clustering algorithms. Initially, the clustering algorithm is explained. Then, the significance of extraction of macula region is quoted which is followed by the proposed methodology to serve the purpose. Then, the results are presented with respect to the literature. The following section deals with the heuristic based clustering algorithm.

Keywords: - *Macula, Segmentation, Data Mining Techniques, Clustering, Unsupervised Clustering, fovea, heuristic based Clustering.*

INTRODUCTION

5.1 HEURISTIC BASED CLUSTERING ALGORITHM

This section discusses the proposed clustering algorithm. On analysing clustering algorithms, it is observed that the initial cluster centre selection influences the performance of clustering process greatly (Hartigan 1975). Commonly and extensively used heuristics for this purpose include random selection of all the cluster centres or random selection of one centre followed by choosing the other centres through a heuristic or choosing every centre through heuristics. To begin with, datasets with less number of instances and that belong to two class category were chosen from public repositories for investigation. Then, all combinations of cluster centres were worked out to analyse the performance of the partitioning. It was interesting to note that a very few combination of initial cluster centres yielded very high accuracy in partitioning. It was higher than that obtained from the previous commonly adopted heuristics. Hence, attempts are made to find the cluster centre combination through some formulated heuristic. At first, trials are made to find one of the two cluster centres. After obtaining one cluster centre, the logic for identifying the other cluster center is devised. On analysing, the metrics that defined the statistical distribution viz., minimum, median,

mean, maximum and skew appear appealing for the purpose of partitioning. The following paragraphs explain the proposed heuristic.

The proposed clustering algorithm utilises the metrics that represent the statistical distribution of data for initial selection of cluster centres. It is well suited for the cases where there exists two groups and the data attributes are continuous. Hence, the algorithm can be applied to image segmentation where the anatomical structure is one group and the background forms the other group. Additionally, the features representing each pixel of the image belong to the category of continuous attributes. The proposed algorithm attempts to choose two cluster centres and the outcome results in two groups such that the intra cluster similarity is less and the inter cluster distance is high. The heuristic first chooses the first centre followed by the selection of the second centre. The steps followed in the algorithm to choose the first cluster centre is shown in Figure 5.1.
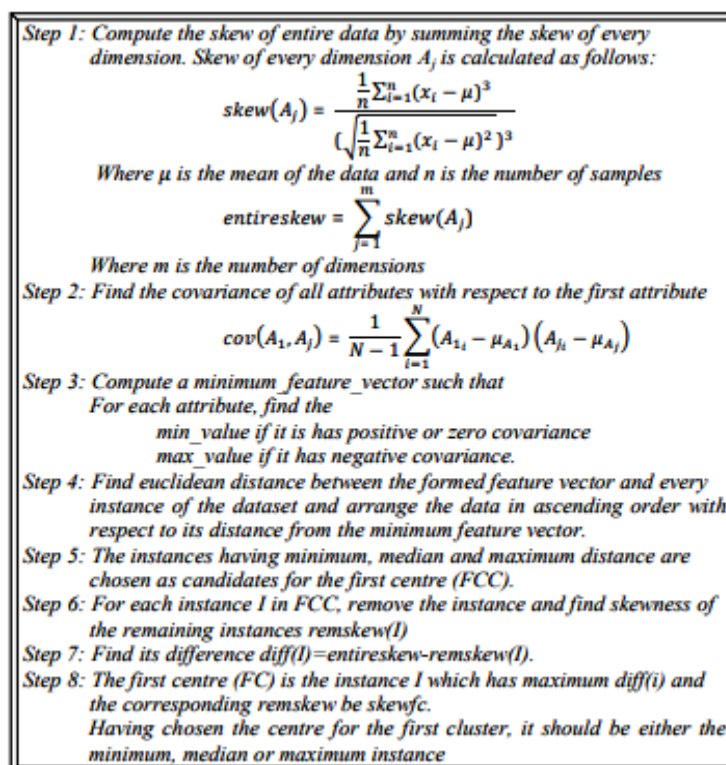

Step 1: Compute the skew of entire data by summing the skew of every dimension. Skew of every dimension A_j is calculated as follows:

$$skew(A_j) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^3}{(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2})^3}$$

Where μ is the mean of the data and n is the number of samples

$$entireskew = \sum_{j=1}^{m} skew(A_j)$$

Where m is the number of dimensions

Step 2: Find the covariance of all attributes with respect to the first attribute

$$cov(A_1, A_j) = \frac{1}{N-1}\sum_{i=1}^{N}(A_{1i} - \mu_{A_1})(A_{ji} - \mu_{A_j})$$

Step 3: Compute a minimum_feature_vector such that
For each attribute, find the
    min_value if it is has positive or zero covariance
    max_value if it has negative covariance.

Step 4: Find euclidean distance between the formed feature vector and every instance of the dataset and arrange the data in ascending order with respect to its distance from the minimum feature vector.

Step 5: The instances having minimum, median and maximum distance are chosen as candidates for the first centre (FCC).

Step 6: For each instance I in FCC, remove the instance and find skewness of the remaining instances remskew(I)

Step 7: Find its difference diff(I)=entireskew-remskew(I).

Step 8: The first centre (FC) is the instance I which has maximum diff(i) and the corresponding remskew be skewfc.
Having chosen the centre for the first cluster, it should be either the minimum, median or maximum instance

---

Figure 5.1 Procedure for selection of first cluster centre

The proposed algorithm initially computes the skew of the entire data. Then, it calculates the covariance of all attributes of the data with the first attribute. Then, a minimum feature vector is formulated based on the covariance of each attribute with the first attribute. It is formed such that if the attribute has a positive covariance, then the minimum value of the attribute is taken; else the maximum value of the attribute is chosen. Then, the distance of every data instance is computed with respect to the minimum feature vector formulated. The data instances are then sorted in ascending order based on the distance obtained from the previous calculation. The instances that are at minimum, median and maximum distance from the minimum feature vector are shortlisted as candidates for first cluster centre and investigated further. Again, skew of the data is calculated after removing the minimum , median and maximum instances with replacement .The difference between the skew of the data after removing the instances and skew of the entire data is computed. The instances, whose removal led to the minimum difference, is chosen as the first cluster centre (FC). Having chosen the first cluster centre, it should either be the minimum, median or maximum instances. Now, the other centre is chosen based on the first centre as given in Figure 5.2.
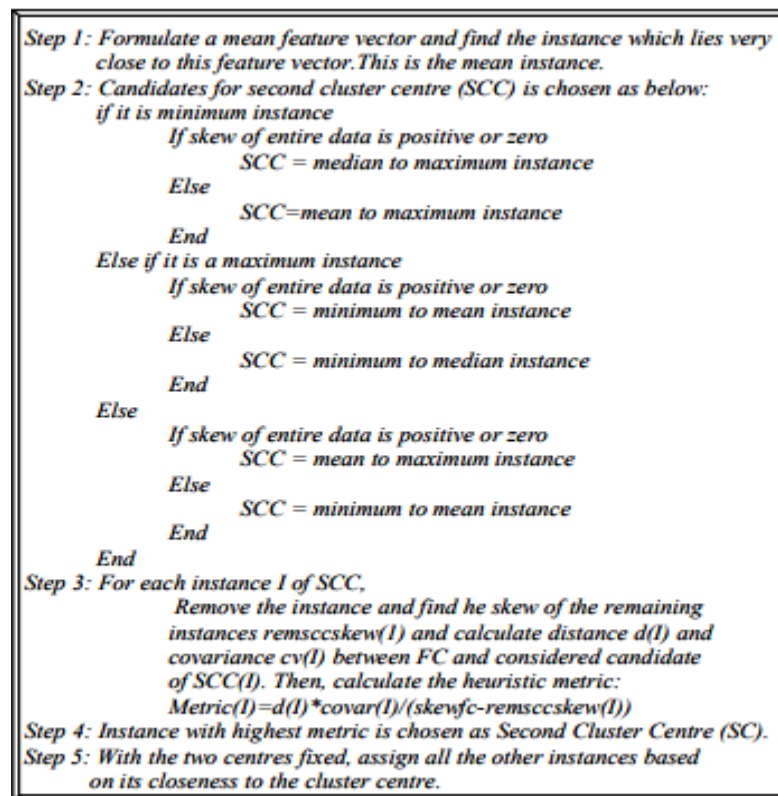
Chapter 5

Figure 5.2 Procedure for selection of second cluster centre

The second cluster centre is chosen based on the following notion. Having chosen the FC, it should be either the minimum, maximum or median instance. Then, a mean instance is formulated for the subsequent investigations. The mean instance will lie between the median and maximum instance, if the data is positively skewed. It will be located between the minimum and median instance if the data is negatively skewed. So, if FC is a minimum or maximum instance, the other centre that produces best grouping is expected to be in another subset, where the subsets are separated by median or by mean instance. The superset between the two subsets is always chosen as the candidates for second cluster centre (SCC).

Based on this concept, the candidates for the second cluster centre are chosen as follows: If FC is a minimum instance and entire skew is positive, the second cluster centre (SC) that yields best partitioning is expected to lie between the median instance and the maximum instance (superset between subsets formed from (i) median instance and maximum instance and (ii) mean instance and maximum instance). Otherwise, if FC is minimum and skew is negative, then SC is expected to lie between the mean and the maximum instance. Similarly, if FC is a maximum instance, then the candidates for SC will be located between minimum and mean instance if the data is positively skewed while it is expected to lie between minimum and median instance if the data is negatively skewed. If FC is a median instance in a positively skewed data, then the candidates for SC are expected to lie between mean and maximum instance. If FC is a median instance in the negatively skewed data, then, the instances from minimum to mean are included in SCC.

The proposed clustering is very useful in grouping continuous data as skew is meaningful for continuous data only. This algorithm can therefore be utilised for the task of image segmentation, where a particular region have to be segmented and the remaining areas are considered as the background. This algorithm is adopted to segment the macula of the retinal fundus images. The proposed methodology incorporating the proposed clustering algorithm for segmentation of macula is presented in the following section.

5.2   MACULA SEGMENTATION AND FOVEA LOCALISATION

Chapter 5

This section discusses the proposed methodology to segment macula, a structural pattern. To begin with, the importance of extracting macula is highlighted. Followed by this, the proposed methodology is detailed. The subsequent sub-section deals with the significance of macula segmentation.

5.2.1   Significance of macula segmentation

The macula is a round area in the central region of the retina, which measures about 3 to 4 mm in diameter (Helga Kolb 2011). It provides high resolution vision and is responsible for central vision. There is a small depression in the centre of the macula measuring around 1 mm in diameter and appears as a round dark area called the fovea (Patton et al 2006). The macula exhibits non-specific structure and varies greatly across individuals due to variations in the levels of pigmentation associated with factors such as ethnicity, age, diet and disease conditions. Anatomically, the fovea centre is located at 2.5 optic disc diameter (DD) from the optic disc (OD) centre. The radius of the macula region is approximately equal to 1 DD (Schwiegerling 2004). This region is devoid of vessels. A sample fundus image showing an annotated macula is illustrated in Figure 5.3.
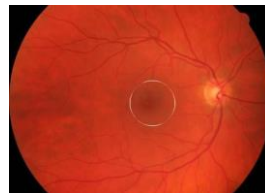

Figure 5.3: Sample fundus image showing annotated macula

5.2.2   Proposed methodology for macula segmentation and fovea localisation

The proposed framework segments the macula from the retinal fundus image through three phase's viz. image pre-processing, data mining and image post-processing phases. The proposed framework is portrayed in Figure 5.4
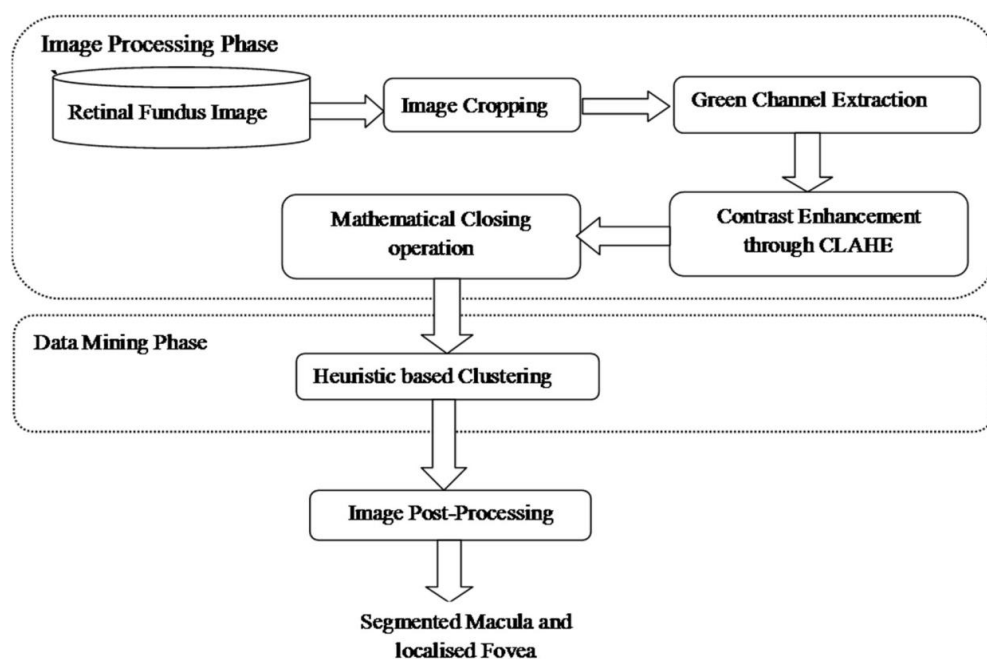

Figure 5.4   Proposed framework for macula segmentation

Chapter 5

The proposed methodology begins with the image pre-processing phase, during which the image is cropped to the desired extent and the green channel image is extracted and contrast enhanced through CLAHE. The intensities of this image are given as input to the proposed clustering algorithm which results in a binary map with many candidate components for macula. Then, image post-processing is performed to eliminate the unwanted components and choose the macula component. The centre of the macula component is the fovea.

To start with, the proposed framework employs image pre- processing techniques. Initially, the image is cropped to delineate the field of view. Then, the upper quarter and lower quarter of the image is eliminated in the view that the macula is mostly found only in the central region of the fundus image, along the line which separates the inferior and superior part of the fundus. The elimination of lower and upper quarters reduces the computational complexity in the forthcoming steps. After cropping the image, the green channel image of the RGB color model is extracted. As this channel exhibits the contrast very well, macula is clearly visible. Then, the green channel image is contrast enhanced through CLAHE procedure (Piezer et al 1987) explained in the previous chapter, to exhibit the macula more prominently.

Sample illustration of the outcomes of the various processes adopted during image pre-processing phase is depicted in Figure 5.5.

(a)   (b)

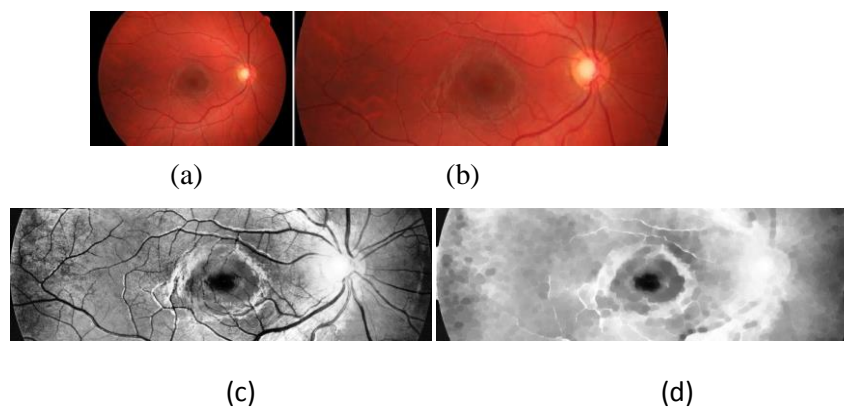(c)   (d)

Figure 5.5 Sample images of outcomes of image pre-processing phase in macula segmentation (a) RGB, (b) cropped RGB, (c) contrast enhanced Green channel, (d) contrast enhanced Green channel after performing closing operation

The intensities of the processed Green channel image are given as input to the data mining phase. The data mining phase involves clustering of the input data into two groups namely background and candidate components for macula. The proposed heuristic based clustering algorithm is adopted for this purpose. The outcome of clustering algorithm is a binary map with a few components. This binary image is then post-processed to find the macula.

The post-processing phase initially eliminates the components whose eccentricity is higher than 0.95. Then, the extreme components, at places out of the field of view, which could not be eliminated during cropping, also appear dark and are exposed as candidates for macula. These regions are eliminated.

Out of the remaining components, each component is superimposed on the green channel image and the minimum intensity of each part is found. The component that corresponds to the minimum of the minimum intensity is chosen as the macula. The holes, if present in the identified component are filled and the component is concluded as the macula. The outcome of the data mining phase, image post-processing phase, and the RGB image showing the detected macula boundary, detected fovea centre and the annotated fovea centre are provided in Figure 5.6.

(a)   (b)

Chapter 5
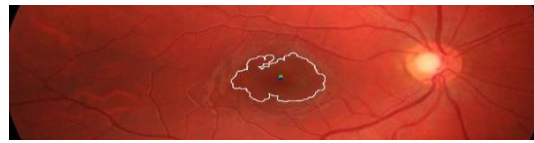
(c)

Figure 5.6 Output of data mining and image post-processing phase in macula segmentation and fovea localization

(a) Data mining phase (b) Image post-processing phase (c) RGB image showing detected macula boundary, detected (+ in blue color) and marked (* in green color) fovea centre

The performance of the proposed methodology is evaluated on various benchmark datasets. The performance on segmentation of macula is exhibited in the following section.

5.3   Performance of proposed methodology in macula segmentation and fovea centre identification

The proposed methodology is evaluated using HRF (Budai et al 2011), DRIVE (Niemeijer et al 2004; Staal et al 2004), DIARETDB0 (Kauppi T et al), DIARETDB1 (Kauppi et al 2007), HEI-MED (Giancardo et al 2012), STARE (Hoover et al 2000) and MESSIDOR (Decenciere et al 2014) datasets. With regard to the MESSIDOR dataset, annotations of fovea centres created and published by the University of Huelva for 1136 images are used for performance evaluation. In contrast, since there are no publicly available annotations offered for the HRF, DRIVE, DIARETDB0, DIARETDB1, HEI- MED, STARE and remaining 64 images of MESSIDOR dataset, fovea centres were marked by experts to make performance evaluation feasible in these datasets.

This sub-section presents the performance of the proposed methodology in identifying the macula and hence the fovea centre. The proposed methodology is also implemented using K-Means clustering algorithm (MacQueen 1967; Lloyd 1982; Forgy 1965; Hartigan 1975) and the performance of K-Means and proposed clustering algorithm is exhibited revealing the better performance of the proposed clustering in this regard. Table 5.1 presents the performance of the proposed methodology in terms of number of images with correctly identified fovea location with respect to the criterion specified in the previous sub-section.

Table 5.1 Performance of the proposed methodology in macula and fovea detection

Dataset	Clustering	Excellent	Good	Fair	Poor	Excellent-Fair
HRF	K-Means	9	19	15	2	43
	Proposed	22	19	4	0	45
DRIVE	K-Means	24	7	1	3	32
	Proposed	23	12	0	0	35
DIARETDB0	K-Means	65	39	18	8	122
	Proposed	72	38	16	4	126
DIARETDB1	K-Means	21	26	31	11	78
	Proposed	46	33	8	2	87
HEI-MED	K-Means	43	55	62	9	160
	Proposed	47	55	65	2	167
STARE	K-Means	6	3	5	6	14
	Proposed	12	2	4	2	18
MESSIDOR	K-Means	996	91	75	38	1162
	Proposed	1020	109	63	8	1192

Chapter 5

On examination of the results of the proposed methodology in segmenting macula, the results justify the better performance of the proposed clustering algorithm in comparison with the K-Means clustering procedure. On computation of accuracy, it is found that the fovea centre is correctly located (satisfying the IR criterion) in 100% of HRF, 100% of DRIVE, 96.92% of DIARETDB0, 97.75% of DIARETDB1, 98.82% of HEI-MED, 90% of STARE and 99.33% of MESSIDOR images. Overall, in 98.93% of images, the fovea centre is identified within the one time OD radius from the real fovea centre. Figure 7.7 graphically represents the improved performance of the proposed clustering with regard to K-Means clustering.
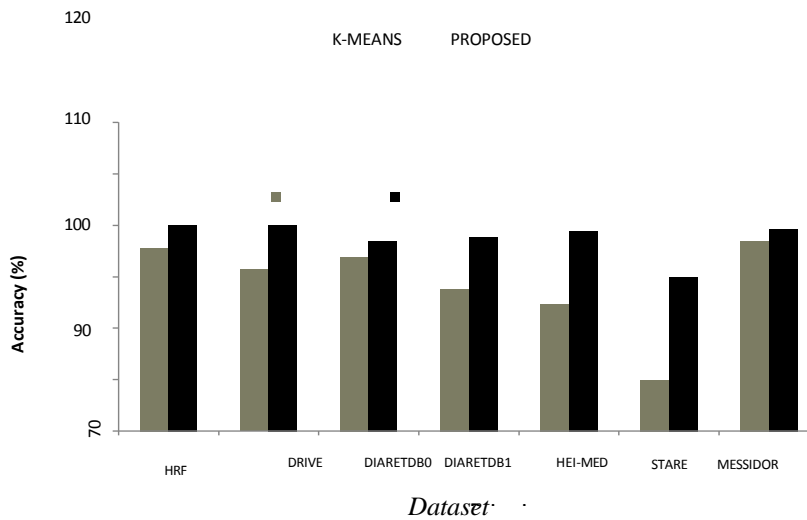

Figure 5.7 Performance comparison of proposed clustering algorithm with K-Means algorithm for macula segmentation and fovea detection

On further investigation of the results owing to the quality and the health status of the images, the following results are reported. With regard to the quality categories viz. Good, Average and Poor, the performance of the proposed methodology is exhibited in terms of satisfying Excellent (<0.25R), Good (>0.25 and <0.5R), Fair (>0.5R and <1R) and Poor (>1R) criteria. The cumulative result projecting correct fovea locations in the span Excellent to fair with respect to various quality categories is also presented in Table 5.2. The results are projected for the entire collection of 1688 images from all the datasets.

Table - Performance of the proposed methodology in macula and fovea detection owing to quality categories.

Quality Category	Excellent	Good	Fair	Poor	Excellent-Fair
Good	1007/1293	186/1293	96/1293	4/1293	1289/1293
Average	218/356	74/356	56/356	8/356	348/356
Poor	17/39	8/39	8/39	6/39	33/39

The results show the number of images in which the fovea centre is correctly located to the number of images in a particular quality category. In terms of accuracy, the results of correct prediction of the fovea centre are reported as follows: 77.88% of Good quality images satisfies the 0.25R Criterion, remaining 14.39% of Good quality images satisfies the 0.5R Criterion and left out 7.43% satisfies the 1R criterion while in 0.31% of good quality images, the detected fovea do not lie within 1R distance from the real fovea centre. This shows that in 99.69% of good quality images, the detected fovea centre lies within the IR distance of the real fovea centre. Similarly, in average quality images, 61.24%, 20.79% and 15.73% of images satisfies the 'Excellent', 'Good' and 'Fair' criteria respectively while

in 2.25% of images, the correct fovea location is not identified. Thus, in the case of average quality images, 97.75% of images have their detected fovea centres at less than the value of OD radius from the fovea centres marked by the experts. On examination of the results with respect to poor quality images, it is found that 43.59%, 20.52% and 20.52% of images satisfies 'Excellent', 'Good' and 'Fair' criteria while in 15.39% of images, the fovea centre is at a distance greater than the OD radius from the real fovea centre. This shows that in 84.62% of poor quality images, the fovea centre lies within the distance of OD radius from the annotated fovea centre.

Subsequently, analysis is performed with respect to health status of the images. Table 5.3 presents the related results.

Table- Performance of the proposed method in macula and fovea detection owing to health of the images

Health Category	Excellent	Good	Fair	Poor	Excellent-Fair
Healthy	551/727	103/727	67/727	6/727	721/727
Diseased	691/961	165/961	93/961	12/961	949/961

The results recorded in Table 7.3 presents the number of images in which the fovea centre is correctly located to the number of images in healthy and diseased images respectively. In terms of accuracy, the results of correct prediction of the fovea centre is reported as follows: 75.79% of healthy images satisfies the 'Excellent' criterion, remaining 14.17% of healthy images satisfies the 'Good' criterion and left out 9.22% satisfies the 1R criterion while in 0.82% of healthy images, the detected fovea do not lie within 1R distance from the real fovea centre. This shows that in 99.18% of healthy images, the detected fovea centre lies within the 1R distance of the real fovea centre. Similarly, in diseased images, 71.90%, 17.17% and 9.68% of images satisfies the 'Excellent', 'Good' and 'Fair' criteria respectively while in 1.25% of images, the correct fovea location cannot be found. Thus, in the case of diseased images, 98.75% of images have their detected fovea centres less than the OD radius from the fovea centres marked by the experts.

The results justify the utilisation of the proposed methodology in automated retinal image system. The following sub-section presents a comparison of the performance of the proposed methodology with regard to the earlier works in macula segmentation and fovea localisation.