3
The Fundus Image Quality Assessment (FIQuA) Data-set

As discussed in Chapter 2, required number of categories of quality and few subjective inputs are the limitations of existing fundus IQA works. Most of the fundus IQA works has not included the appropriate subjective opinion of medical doctors about the quality of images. Although, in a few of the works [Wang *et al.*, 2016; Fu *et al.*, 2019a], the authors have included the subjective evaluation of the fundus images. However, it lacks in terms of participation of an appropriate number of ophthalmologists and local quality evaluation. The EYE-Q data-set [Fu *et al.*, 2019a] is prepared using the opinions taken from only *two* medical experts, which is too low to claim that it matches the generalized perception of ophthalmologists about a fundus image. It is important to mention here that there are no guidelines available for the subjective quality evaluation of medical images. However, for natural images, such procedures are known by the International Telecommunication Union (ITU) in [Union, 2015]. According to the guidelines, at least 15 subjects (persons) should participate in an SQA process of images/videos. Considering these guidelines for fundus images, a data-set prepared with the inputs taken only from *two* medical experts is a significant limiting factor.

Furthermore, these data-sets contain perceptions only for the overall image quality. However, before reaching any conclusion, an ophthalmologist first inspects all the required information that must be present locally in different spatial locations in an image. As, our objective is to build a fundus IQA algorithm that can efficiently mimic the perception of ophthalmologists. Therefore, to properly understand the medical doctor's visual system, it is necessary to carefully observe and analyze the subjective evaluation of local quality parameters such as the structural parameters present in a fundus image. A new fundus image quality assessment (FIQuA) data-set is prepared in view of the limitations as mentioned above and challenges. In this Chapter, a detailed description of the FIQuA data-set is provided.

The remainder of the chapter is organized as follows. Section 3.1 includes the description and peculiarities of the FIQuA data-set. The analysis of the subjective quality assessment is given in Section 3.2 and Section 3.3 summarizes the chapter.

Table 3.1 : Classification of the six quality parameters for the subjective quality assessment

S.No.	Structural Parameters	Generic Parameters
1	Visibility of Optic Disc (F1)	Color (F4)
2	Visibility of Macula (F2)	Contrast (F5)
3	Visibility of Blood Vessel (F3)	Blur (F6)

Figure 3.1 : Samples of the fundus images from FIQuA dataset.

Table 3.2 : Sample of the subjective scores and corresponding quality class graded by the ophthalmologists for the respective images shown in Fig. 3.1

Image	F1	F2	F3	F4	F5	F6	Class
Image 1	10	9	10	9	9	10	Good
Image 2	8	6	7	5	6	6	Fair
Image 3	1	0	1	1	1	1	Poor

3.1 DESCRIPTION AND PECULIARITIES OF THE PROPOSED FIQUA DATA-SET

The fundus images used for the subjective evaluation are taken from a publicly available large data-set provided by EyePACS [Kaggle, 2015] for the DR detection challenge. The ophthalmologists were asked to grade all the pictures into one of the following three categories: Good, Fair, and Poor. The definitions for the overall quality classes are given below:

- **Good:** *The quality of the given fundus image satisfies all the necessary expectations based on quality parameters, and the image is deemed reliable for the diagnosis.*

- **Fair:** *The quality of the given fundus images does not satisfy all the necessary expectations, but at the same time the image may support a diagnosis in some contexts.*

- **Poor:** *The quality of the given fundus images is not at all satisfying the necessary expectations and surely not reliable for the diagnosis.*

As mentioned earlier in Chapter 2, the quality indicators for fundus images are: focus and clarity, field definition, visibility of the macula, optical disc and blood vessels. These include both structural (local) and generic (global) fundus image properties. Also, a careful study of the previous work leads us to conclude that the quality parameters that are used by the researchers can be divide into two categories: (i) Structural, and (ii) Generic. In this work, we have identified a total of six quality parameters from these two categories. Table 3.1 provides the classification of the quality parameters. Further, F1-F6, as mentioned in Table 3.1, will be used to represent the respective quality parameter. The ophthalmologists were asked to provide ratings for these six different quality parameters of fundus images on the scale of 0 to 10, where a higher number indicates better quality. A total of fifteen ophthalmologists have participated in the subjective quality assessment (SQA) process. The ophthalmologists are from prestigious medical institutes in India with more than 5 years

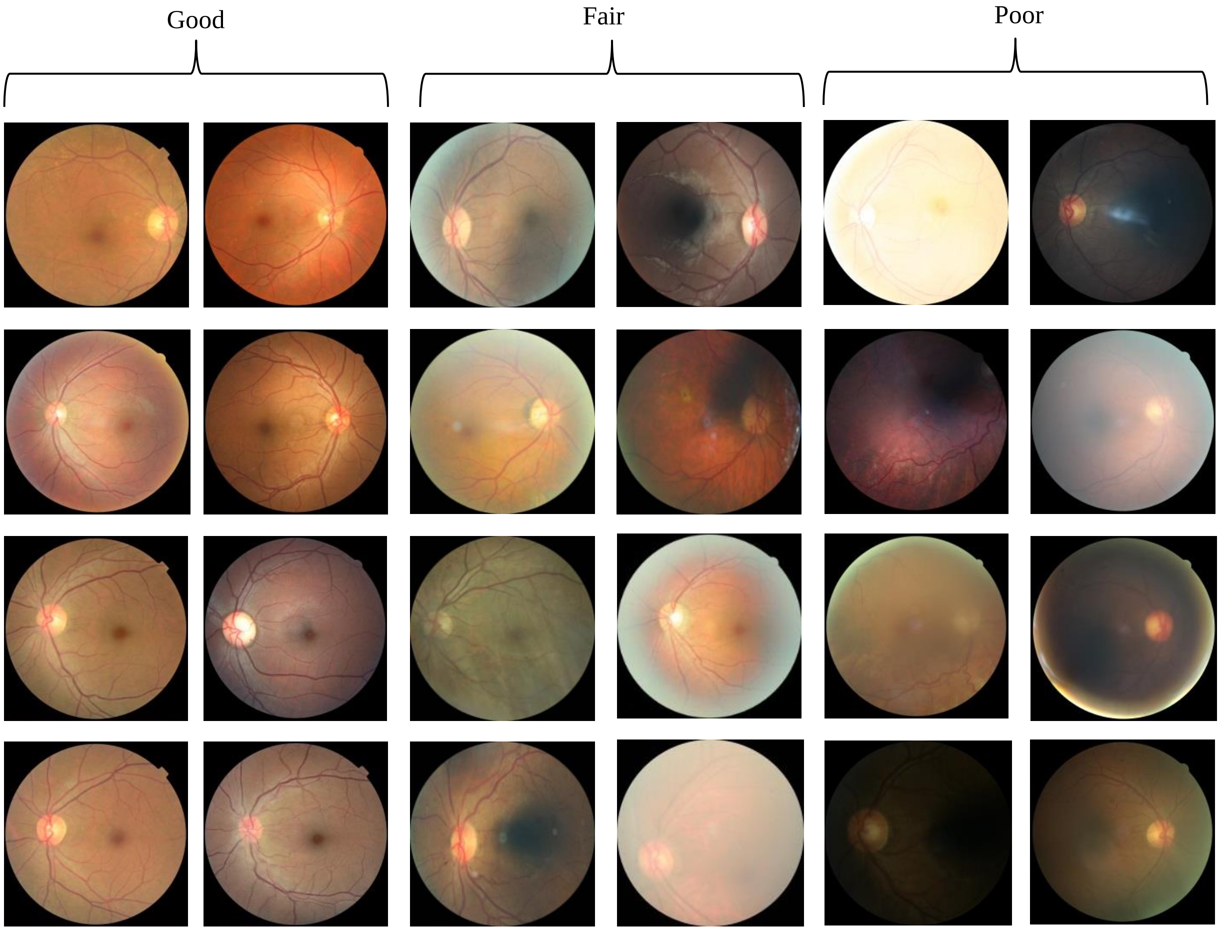

Figure 3.2 : Samples of the fundus images taken from each category of FIQuA dataset.

of experience. The two participating hospitals are All India Institute of Medical Sciences (AIIMS), Jodhpur and Mathura Das Mathur (MDM) hospital Jodhpur, India. Here, AIIMS is an institute of national importance. The number of experts was selected according to the recommendations of the International Telecommunication Union (ITU) for the subjective evaluation of images given in ITU-R Rec. BT.500 [Union, 2012]. We sought the services of medical doctors across the spectrum of expertise and experience to provide inputs on the quality of the images. A maximum of 40 images has been used for subjective quality evaluations at a time to get the required data from the ophthalmologists. For illustration purposes, through Fig. 3.1 and Table 3.2 the output of the SQA process is presented. The Fig. 3.1 contains samples of the fundus images from the FIQuA dataset and Table 3.2 the respective subjective ratings. In addition to Fig. 3.1, a few more samples of the fundus images from each category of quality are also shown in Fig. 3.2.

3.2 ANALYSIS OF SUBJECTIVE QUALITY ASSESSMENT

The details of the subjective study are reported below, together with an analysis of the results. The study aims at:

- validating the collected data inputs.

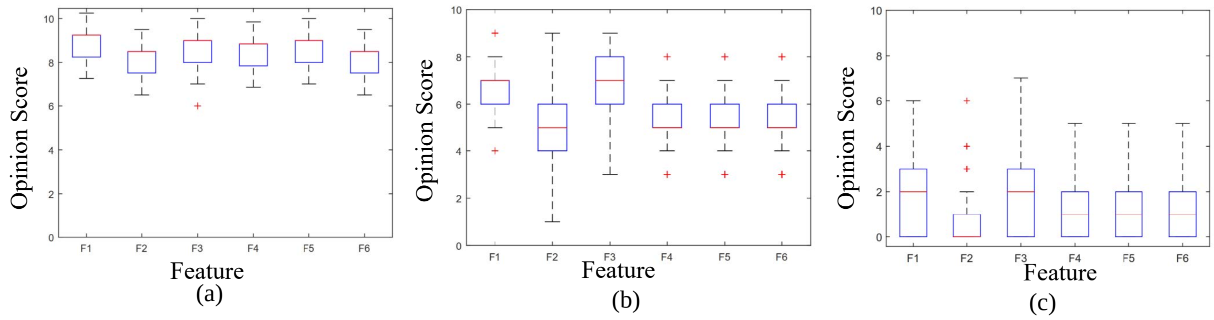

Figure 3.3 : Graph showing the range of Opinion Score values for all the six features for the three classes of quality: (a) Good, (b) Fair, and (c) Poor. Here (+) indicates the outlier values.

- providing a better understanding of ophthalmologist's visual perception, by analyzing the relationship between the physical changes in quality parameters and the corresponding changes in visual perception.

Due to human errors and variability across subjects, dissimilarities across the opinion scores still exist. The outliers were detected and removed using the Median Absolute Deviation, given in the equation below:

$$MADN = c \times median(|X_i - median(X)|) \tag{3.1}$$

where c=1.483 and X_i is the score provided by the medical expert i, with $i = 1, 2, ..., N$ where N is the number of medical experts (i.e., the median is calculated over the opinion scores of the different subjects on the considered image/feature). After outlier removal, the final subjective score value for a particular feature is derived by averaging the remaining values. The MAD method considers an element as an outlier if it is more than three times the MAD from the median value. The MAD method is preferred over the mean plus-minus three standard deviation method because it does not pre-assume the distribution of the data and is efficient for a small sample size [Leys *et al.*, 2013]. The ground truth for the overall image quality class was selected by choosing the median value from the inputs provided by all the medical doctors. Fig. 3.3 illustrates the range of subjective values obtained for the features for each class. We can observe that the majority of the subjective values for each feature are in the range of $10 \geq S_V > 7$, $7 \geq S_V \geq 5$, and $5 > S_V \geq 1$ for the good, fair, and poor classes, respectively. Here, S_V represents subjective scores. The values of F1-F6 have been used to train various classifiers to classify the fundus image into the Good, Fair, and Poor category. The data was split into an 80-20% ratio for training and testing, i.e., 1200 for training and 300 for testing. The results in Table 3.3 show that the feature set made using the subjective score values gives high classification accuracy. It is important to mention here that all the cases of wrong classification occurred between the Good & Fair and Fair & Poor classes. The confusion matrices for all the four classification algorithms mentioned in Table 3.3 is shown in Fig. 3.4. It validates that there is no single sample with wrong classification between Good and Poor classes. The obtained results are important because the objective was to reduce the number of the wrong classification between Good and Poor classes and simultaneously determine the need for enhancement in the fundus images. Furthermore, the contribution of each quality parameter towards the perceptual quality is investigated. The coefficient values derived for each quality parameter using the classification algorithms have been applied for the analysis mentioned above. In Table 3.3 it can be observed that one of the highest classification accuracy is achieved by the SVM algorithm. The coefficient values obtained using the SVM method are shown in Table 3.4. The obtained coefficient

Table 3.3 : Comparison table of accuracy (in %) of various classifiers for individual classes and overall. SVM: Support Vector Machine (Polynomial Kernel); NB: Naive Bayesian; RF: Random Forest; SF: SoftMax.

S.no.	Classifier	Good	Fair	Poor	Overall Accuracy
1	SVM	100	99.1	98.8	99.33
2	NB	99.0	98.3	98.8	98.7
3	RF	100	99.1	98.8	99.33
4	SF	100	98.2	100	99.33

Table 3.4 : Coefficient values obtained for F1-F6 from SVM (Polynomial Kernel) classification method.

F1	F2	F3	F4	F5	F6
1.463	2.836	1.7532	2.563	2.463	2.281

values indicate that "Visibility of Macula (F2) and Color (F4)" are the two parameters that mostly affect the perceptual quality of the fundus images. Also, the least importance is given by the ophthalmologists to "Visibility of Optical Disc (F3) and Visibility of Blood Vessels (F1)".

Predicted Class

Actual Class

Class	Good	Fair	Poor
Good	98	0	0
Fair	1	115	0
Poor	0	1	85

(a) Support Vector Machine

Class	Good	Fair	Poor
Good	97	1	0
Fair	2	114	0
Poor	0	1	85

(b) Naïve Bayesian

Class	Good	Fair	Poor
Good	98	0	0
Fair	1	115	0
Poor	0	1	85

(c) Random Forest

Class	Good	Fair	Poor
Good	98	0	0
Fair	2	114	0
Poor	0	0	86

(d) Soft Max

Figure 3.4 : Confusion Matrices for the each of the four classification results shown in Table 3.3.

3.3 SUMMARY

Ophthalmologists assess the quality of fundus images based on two quality parameters: Structural and Generic. This work aims at assessing the quality of fundus images on similar grounds. In this chapter, a new fundus image quality assessment data-set (FIQuA) of 1500 images is prepared and the peculiarities and specifications are presented.

- For each image, a total of *seven* subjective inputs taken from the *fifteen* ophthalmologists. Out of the seven inputs, the first six are subjective scores in the range of [0,10] for following six quality parameters: (1) Visibility of Optic Disc (F1), (2) Visibility of Macula (F2), (3) Visibility of Blood Vessel

(F3), (4) Color (F4), (5) Contrast (F5), (6) Blur (F6). The last input is the overall class of fundus image quality: good, fair, and poor.

- The outliers in the subjective scores are identified and removed using the median absolute deviation method.
- It has been observed that most of the subjective opinion scores for each feature lie in the range of $10 \geq S_V > 7$, $7 \geq S_V \geq 5$, and $5 > S_V \geq 1$ for the Good, Fair, and Poor classes, respectively.
- The significance of the subjective scores is analyzed using values of F1-F6 to train various classifiers to classify the fundus image into the above-mentioned categories of quality.
- For experiment purposes, a total of four classifiers (Support vector machines, Naive Bayesian, Random forest, Softmax) are trained using values of F1-F6. The obtained results achieved more than 98% accuracies. It proves that the feature set prepared using the subjective score is highly effective in identifying the quality of a fundus image correctly.
- Another important finding is that there is no wrong classification occurred between the Good & Poor classes. The obtained results are significant because our objective is to reduce the wrong classification between Good and Poor classes and simultaneously determine the need for enhancement in the fundus images.
- Finally, to closely understand the ophthalmologists' visual perception, the coefficient values obtained from the SVM classifier are analyzed. It has been observed that medical doctors have given preference to "Visibility of Macula (F2) and Color (F4)" parameters while judging the fundus image quality. Also, the least importance is given to the "Visibility of Optical Disc (F3) and Visibility of Blood Vessels (F1)" quality parameters.

An ideal fundus IQA method is expected to produce quality predictions closely similar to the ophthalmologists visual perception. Keeping this objective in mind, in the next chapter, we propose a new Multivariate regression based neural network model for fundus image quality assessment.

...