# NO-REFERENCE QUALITY ASSESSMENT FOR RETINAL FUNDUS IMAGES

*Chander Dev, Sharang MS, Soumya Jana*

Indian Institute of Technology Hyderabad, India

## ABSTRACT

Fundus imaging is the most commonly used modality to collect information about the human eye background. The primary structures that can be visualized on Fundus images are the central and peripheral retina, optic disc and macula, thus making them very important from a diagnostic perspective. The problem of automatically evaluating the quality of such an image in terms of its medical utility is therefore an important one, and has the potential to speed up the imaging and diagnostic processes.
In this project we have developed an algorithm that can asses these images in real-time for their quality and classify them as gradable or ungradable. Our approach is to train a classifier based on a comprehensive set of features that are representative of the following quality indicators: colour, focus, contrast, symmetry and illumination.

***Index Terms—*** Fundus Image, Quality Estimation, No-reference, generic features

## 1. INTRODUCTION

The problem of Image Quality Assessment(IQA) has garnered a lot of attention in the recent years due to the proliferation of image data generated. Most of the focus is on the problem of Natural Image IQA, which forms the bulk of the images generated these days. The problem of Medical Image IQA is comparatively less studied, but nonetheless very significant as it has the potential to significantly improve the quality and speed of medical diagnosis.

The problem in medical IQA is to estimate the level of distortion present in an image which affects its diagnostic utility. Broadly, Image Quality Assessment problems are of three types: Full Reference (FR) problems where the pristine image is available during assessment time, Reduced Reference (RR) problems where some features of the pristine image are available during assessment and No Reference(NR) problems where there is no information about the pristine image available during assessment time. The latter case is the most interesting, and is what we have in the Fundus image IQA discussed here. There are three typical approaches to NR-IQA problems :
1) Image Statistics Based Approaches: Here, the statistical model generating pristine images is estimated, and the quality of an image is determined by how much its parameters differ from the parameters of the estimated model. This is a well studied problem for natural images, but difficult in practice for medical images due to the scarcity of available data to model the generative distribution.
2) Distortion Specific Approaches : These assume specifics forms for the distortion, such as blur, aberration etc. and attempt to quantify them, obtaining a quality metric.
3) Data Driven Discriminative Approaches: These involve training machine learning models such as DNNs to predict the quality of images.

In case of Fundus image quality estimation, depending on whether an analysis of the anatomical structure in the Fundus image is primarily targeted or not, Fundus image quality evaluation methods can be divided into two categories. The first category is based on generic image statistical indicators. The other approach is to analyze the anatomical structure of the Fundus image to evaluate quality. Both the categories have their respective strengths and applicability. Generic statistical techniques are widely applicable, and they work well even if the image quality is poor. In contrast, because the Fundus image structure analysis is complicated and error-prone, anatomical structures analysis methods may fail when there are unusual changes in Fundus image or the image is of very poor quality.

## 2. BACKGROUND

A lot of significant work has been done in this field before. As per our knowledge, [1] is the first paper on Fundus image quality estimation in which the notion of Image Statistics is explored. Pixel histograms for high quality and low quality images are computed for the training data. Given a previously unseen image, the similarity between its pixel histogram and those of the two classes is used to classify it. This idea is still the foundation for many of current Fundus image quality methods.
In some works like [2], the focus is on the perceptual quality of the images rather than their clinical utility. However, it is not necessary that a perceptual bad quality image is not-gradable in clinical sense.
In [3] vessel segmentation is performed, and the quality of the extracted vessels is used as a proxy for the quality of the Fun-

dus Image.
In this project, our approach is along the lines of [4] and [5]. They extract a wide variety of features that contribute to both the perceptual quality and diagnostic utility of the Fundus image, such as pixel statistics, texture based features, wavelet features and features that capture image symmetricity. [4] uses gray scale images and extracts 113 generic features after which a classifier is trained and [5] extracts more or less the same features in all three color channels.

## 3. PROBLEM STATEMENT

Given a colored Retinal Fundus image, our objective is to classify it with respect to its diagnostic utility. The point of critical importance is that diagnostic utility, not perceptual quality is desired. Thus, the classifier must be able to distinguish between image distortions which affect quality, and pathological features which are vital in terms of its diagnostic utility.

(a) Gradable (b) Non-gradale

**Fig. 1**: Good and bad quality Fundus images

## 4. MATERIAL AND METHODOLOGY

In this section, we describe, how the features determining the quality of Fundus image are obtained and what dataset have been used for training purpose.

### 4.1. Material

The model needs to be trained with quality labels provided by medical experts. The volume of the dataset should be as large as possible. A proprietary dataset, obtained from the LV Prasad Eye Institute was used for training our model.

Due to the limited number of patients and the bias issues arising due to the fact that the patients and the imaging techniques used in one institute are very similar, it was necessary to augment the dataset with other images. Hence, a public retinal image dataset called the DRIMB dataset[6] with 125 good and 69 bad images was also used.

### 4.2. Methodology

We extract generic features andg these features are used to train classifiers. Flowchart in Fig 2 gives a brief overview of the whole process.

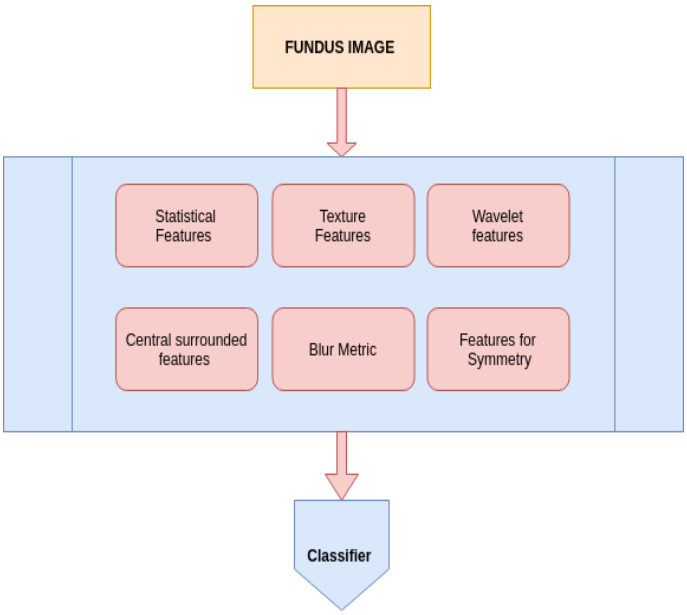


**Fig. 2**: Flowchart of proposed method

It is observed that the pixel information of the poor and good quality images have specific behaviors. Overexposure would produce a distinct bright area in the image while underexposure would cause the appearance of dark areas. Both kinds of degradation lead to a change in the pixel distributions.
Defocusing in image will cause the loss of image details, and hence the high-frequency components will be lost. Such anomalies can be captured by wavelets features.
Non-uniform illumination and chromatic aberrations will change the contrast of the image. Such changes can be captured using image texture features.
The following are are the features that are extracted:

#### *4.2.1. Statistical characteristic of the pixel value*

In order to characterize the pixel distribution of the images, we extracted the first four moments of the distribution as features.

$$\overline{x}(mean) = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma(Deviation) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})} \tag{2}$$

$$s(Skew) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^3}{\sigma^3} \tag{3}$$

$$k(Kurtosis) = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^4}{\sigma^4} \tag{4}$$

Where N is the total number of pixels. Along with these, 3/4 quantile and median are also measured. To quantify the degree of "randomness" in the image, the Image Entropy is calculated:

$$E = -\sum_{i=1}^{255} p_i * \log_e(p_i) \tag{5}$$

in which the $p_i$ is the probability of the pixel intensity taking the value i.

#### *4.2.2. Haralick features for Texture*

We extracted 24 features that contain the information about the image texture based on the co-occurrence matrix, that contains the probability of coexistence of any two pixels values in the specific mode. The features obtained are: contrast, correlation, Energy, ASM, Dissimilarity and Homogeneity in four different modes, which are horizontal, vertical and two diagonal directions.

$$Contrast = \sum_{i,j} |i-j|^2 p_{i,j} \tag{6}$$

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-_j)p_{i,j}}{\sigma_i \sigma_j} \tag{7}$$

$$Energy = \sum_{i,j} p_{i,j}^2 \tag{8}$$

$$Homogeneity = \sum_{i,j} \frac{p_{i,j}}{1+|i-j|} \tag{9}$$

$$AngularSecondMoment = \sum_{i,j} p_{i,j}^2 \tag{10}$$

$$Dissimilarityâ = \sum_{i,j} p_{i,j}|i-j| \tag{11}$$

#### *4.2.3. Central Statistical characteristics and center- surrounded difference features.*

The retinal Fovea is the most significant regions of the Fundus image for clinical diagnosis. Roughly saying it is positioned at the center of the image. So we extract all statistical features for the middle of the image. Also, we take the difference between the effective mean intensity of the vertically-opposite patches. These features will give us information about the symmetricity of the image. It is noticeable that in case of uneven illumination the quality of the image is reduced and the symmetric features will capture that.

#### *4.2.4. Wavelet-based features*

If the image is blurred or overexposed, there will be a loss of fine details will be reflected in the loss of high-frequency components of the image. Thus, a frequency domain analysis of the image will be useful in detecting such distortions. Wavelet analysis is used for this purpose.
The importance of wavelet analysis stems from its adoption of the concept of multiresolution. Different parts of the data are viewed through different size windows of resolution. High frequency parts of the data use a small window to obtain good time resolution, whereas low frequency parts use a larger window to achieve good frequency resolution [9].
Wavelet analysis uses a template function known as the âmother waveletâ. The image is then correlated with shifted and scaled versions of this mother wavelet to obtain the wavelet representation.

The continuous wavelet transform (CWT) of a function of time $\chi$ is defined as follows:

$$\int_{-\inf}^{+\inf} f(t).\chi(scale, position)dx$$

A very similar definition holds in the discrete domain.

Different mother wavelets can be used to bring out different properties of the image in the wavelet representation. In our case, one level-separation using Haar wavelets is performed and frequency components in three directions ( horizontal, vertical and diagonal) are extracted.

#### *4.2.5. Feature visualization*

A total of 113 feature vectors are extracted from each of the Fundus images from the dataset as explained in section Generic features 4.2. These features are visualized in a one dimensional space using Linear discriminant analysis [7](LDA). Linear separability is observed, which means that the selected features are relevant.

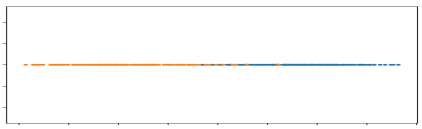

**Fig. 3**: 1-D LDA visualization of the features

#### 4.2.6. *Support Vector Machine*

Support Vector Machines are based on the concept of maximum margin decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. The features extracted are used to train a SVM classifier, which learns a decision boundary, with radial basis function(rbf) kernel which is defined as: $K(x, x') = e^{-\gamma||x-x'||^2}$
After posing the maximum-margin problem as an optimization problem and solving the dual problem, we obtain an expression of the following form for the optimal separating hyperplane:
$\Sigma_i \alpha_i K(x_i, x)$
$\alpha_i$ is zero for most of the datapoints. Therefore, the separating hyperplane is determined by a small subset of the datapoints which are called the support vectors.

#### 4.2.7. *Partial least squares(PLS) followed by LDA weights*

Partial Least Squares is a supervised dimensionality reduction that works by projecting the data and the targets to lower dimensional spaces while maintaining class separability. This technique is necessary to train a Linear Discriminant Analysis classifier because the high collinearity among the extracted features cause the within-class and between-class covariance matrices to be non-invertible, making the computation of the LDA projection matrix problematic.

#### 4.2.8. *AdaBoost Classifier using Decision Trees*

AdaBoost or Adaptive Boosting is an ensemble learning model which combines many weak classifiers to construct a powerful classifier. It works by assigning weights to the set of weak classifier such that each classifier is more sensitive to the datapoints misclassified by the previous classifiers.
The weak classifiers used in our algorithm are Decision Tree Classifiers. AdaBoost was trained with 1000 estimators, i.e the model was an ensemble of 1000 decision tree classifiers.
Let the training data be $n$ $(x_i, y_i)$ pairs where each $y_i \in \{-1, 1\}$. Further, let us assume we have L weak classifiers $\{k_1 ....k_L\}$.
Then, we define our boosted classifier recursively as follows:
The output of the boosted classifier after $m$ iterations of learning on the datapoint $x_j$ is defined as $C_m(x_j) = C_m(x_j) + \alpha_m k_m(x_j)$
If the total loss is defined as the sum of the exponential loss at each datapoint, then the optimal value of $\alpha_m$ is given by $\alpha_m = \frac{1}{2} \ln \frac{\Sigma_{y_i = k_m(x_i)} \exp(-y_i C_{m-1}(x_i))}{\Sigma_{y_i \neq k_m(x_i)} \exp(-y_i C_{m-1}(x_i))}$

## 5. RESULTS

Features extracted as mentioned in section Generic Features [4.2] are used to train multiple classifiers and detailed experimentation and analysis is performed. Accuracies mentioned in later sections are the average accuracy of k-fold cross validation with $k = 5$ for 100 times. Mean RoC plot is also shown which is plotted over the testing part of the dataset. All the confusion matrices reported are obtained from taking average of all the confusion matrices generated in every iteration of the k-fold cross validation. 1-sigma band is also drawn around the RoC curve and the standard deviation of the accuracy is mentioned for every classifier trained.

### 5.1. Results from SVM

Average accuracy of $88.99\% \pm 0.03\%$ is observed. Confusion matrix shown in the Fig 4 and one representative ROC out of 5 different trained classifier is shown in Fig 5.

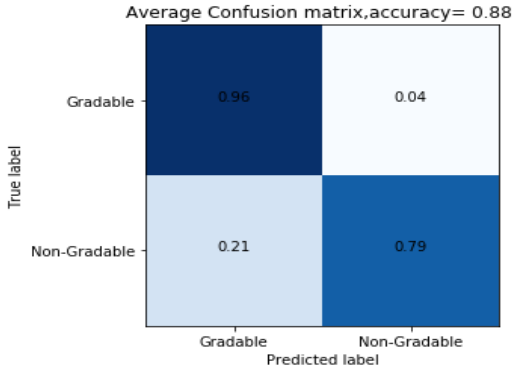


**Fig. 4**: Normlaized confusion matrix SVM

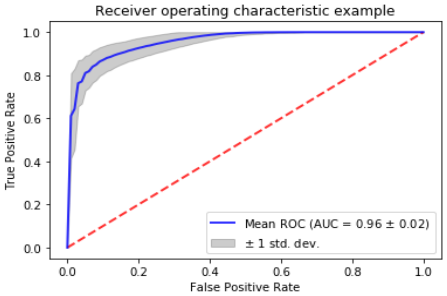


**Fig. 5**: ROC Curve from SVM

### 5.2. Results from Partial Least Squares dimensionality reduction followed by LDA

PLS was used to project the 113 features into a 10 dimensional space. An LDA classifier was trained in the projected space.
Accuracy was improved from that of SVM classifier. Average accuracy of $89.06 \pm 0.04\%$ is observed. Average confusion matrix is shown in Fig 6 and one representative ROC out of 5 different trained classifier is shown in Fig 7.

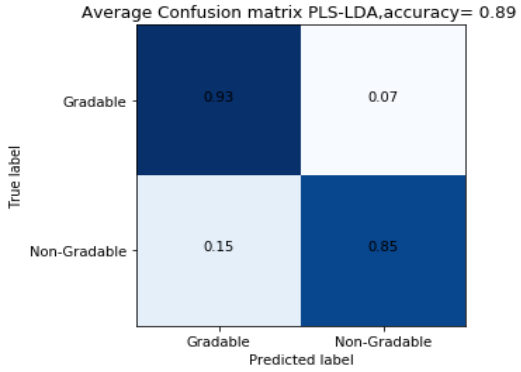


**Fig. 6**: Normalized Confusion Matrix PLS-LDA

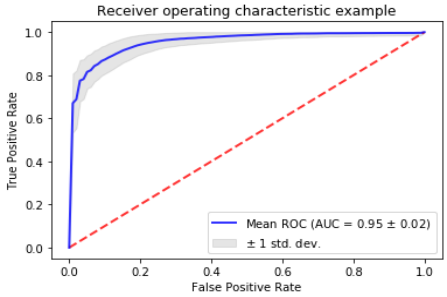


**Fig. 7**: Mean ROC Curve from PLS-LDA

### 5.3. Results from Adaboost

Average accuracy of $92 \pm 0.0018\%$ is observed. Confusion matrix shown in the Fig 8 and and one representative ROC out of 5 different trained classifier is shown in Fig 9. Significant improvement in accuracy and standard deviation is observed.

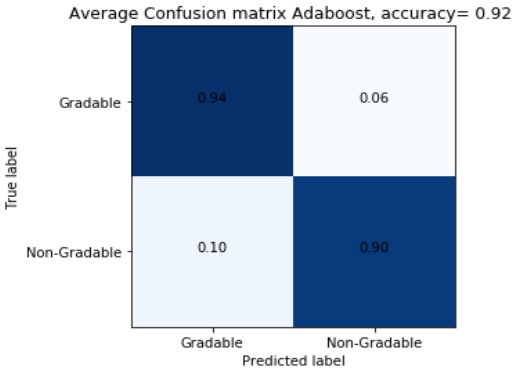


**Fig. 8**: Normalized Confusion Matrix AdaBoost

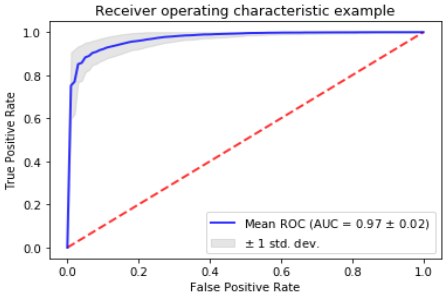


**Fig. 9**: Mean ROC Curve from AdaBoost

| Classifier | Accuracy(%) | Std(%) |
|---|---|---|
| SVM | 88.99 | 0.03 |
| PLS-LDA | 89.06 | 0.04 |
| AdaBoost | 92.00 | 0.0018 |

## 6. CONCLUSION

In this report we developed a technique that can detect the quality of a Fundus image automatically. The proposed algorithm is fast and accurate, and comparable to the existing state of the art solutions. One drawback of our method is that it does not include any prior medical knowledge about the anatomical structure of the eye, except in a very limited sense by giving more importance to the central region. This is clearly at odds with the way a human expert approaches this problem by looking for some key anatomical features.
Future attempts that include some anatomical knowledge as priors along with the features extracted here are expected to perform much closer to the level of human experts.

## 7. REFERENCES


[1] Samuel C Lee and Yiming Wang, "Automatic retinal image quality assessment and enhancement," in *Medical Imaging 1999: Image Processing*. International Society for Optics and Photonics, 1999, vol. 3661, pp. 1581–1591.

[2] Shaoze Wang, Kai Jin, Haitong Lu, Chuming Cheng, Juan Ye, and Dahong Qian, "Human visual system-based fundus image quality assessment of portable fundus camera photographs," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1046–1055, 2016.

[3] Thomas Köhler, Attila Budai, Martin F Kraus, Jan Odstrčilik, Georg Michelson, and Joachim Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*. IEEE, 2013, pp. 95–100.

[4] Zhenjie Yao, Zhipeng Zhang, Li-Qun Xu, Qingxia Fan, and Ling Xu, "Generic features for fundus image quality evaluation," in *e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*. IEEE, 2016, pp. 1–6.

[5] Herbert Davis, Stephen Russell, Eduardo Barriga, Michael Abramoff, and Peter Soliz, "Vision-based, real-time retinal image quality assessment," in *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*. IEEE, 2009, pp. 1–6.

[6] Attila Budai, Rüdiger Bock, Andreas Maier, Joachim Hornegger, and Georg Michelson, "Robust vessel segmentation in fundus images," *International journal of biomedical imaging*, vol. 2013, 2013.

[7] Zhihua Qiao, Lan Zhou, and Jianhua Z Huang, "Effective linear discriminant analysis for high dimensional, low sample size data," in *Proceeding of the world congress on engineering*. Citeseer, 2008, vol. 2, pp. 2–4.