IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 4, 2020, accepted March 19, 2020, date of publication March 23, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982588

Multivariate Regression-Based Convolutional Neural Network Model for Fundus Image Quality Assessment

ADITYA RAJ[1], (Student Member, IEEE), NISARG A. SHAH[1], ANIL KUMAR TIWARI[1], (Member, IEEE), AND MARIA G. MARTINI[2], (Senior Member, IEEE)

[1]Department of Electrical Engineering, IIT Jodhpur, Jodhpur 342037, India
[2]Wireless Multimedia and Networking Research Group (WMN), Kingston University London, Kingston Upon Thames KT1 2EE, U.K.

Corresponding author: Aditya Raj (raj.2@iitj.ac.in)

This work was supported in part by the Visvesvaraya Ph.D. Scheme for Electronics and IT Research Fellowship (MeitY, India), and in part by the Newton-Bhabha Fund sponsored by the Department of Biotechnology (DBT), India, and British Council, U.K.

ABSTRACT Objectively assessing the perceptual quality of an ocular fundus image is essential for the reliable diagnosis of various ocular diseases. A fair amount of work has been done in this field to date. However, the generalizability of the current work is limited, as the existing quality models were developed and evaluated with data-sets built with limited subjective inputs. This paper aims at addressing this limitation with the following two contributions. First, a new fundus image quality assessment (FIQuA) data-set is presented, containing *1500* fundus images with three classes of quality: Good, Fair, and Poor. Also, for each image, subjective scores (in the range [0-10]) were collected for *six* quality parameters, including structural and generic properties of the fundus images. Second, a new multivariate regression based convolutional neural network (CNN) model is proposed to predict the fundus image quality. The proposed model consists of two individually trained blocks. The first block consists of four pre-trained models, trained against the subjective scores for the six quality parameters, and aims at deriving the optimized features for classification. Next, the optimized features from each of the four models are ensembled together and transferred to the second block for final classification. The proposed model achieves a strong correlation with the subjective scores, with the values *0.941*, *0.954*, *0.853*, and *0.401* obtained for SROCC, LCC, KCC, and RMSE respectively. Its classification accuracy is *95.66%* over the FIQuA data-set, and *98.96%* and *88.43%* respectively over the two publicly available data-sets DRIMDB and EyeQ.

INDEX TERMS Fundus image quality assessment, diabetic retinopathy, multivariate regression, convolutional neural network.

I. INTRODUCTION

In the field of Ophthalmology, digital fundus photography is used for the diagnosis of various ocular disorders like Cataract [1], Diabetic Retinopathy (DR) [2], Glaucoma [3], Age-related macular degeneration (AMD) [4]. Among all, DR is one of the primary causes of vision loss worldwide. For effective medical assistance to a huge number of patients, the current number of eye specialists is inadequate [5]. To address the lack of the required ophthalmologists, telemedicine [6], and computer-aided diagnosis (CAD) systems [7] are the potential solutions. Also, today we are heading towards mobile application based diagnosis systems [8] for ocular diseases. However, for a reliable diagnosis, the quality of a fundus image must be ensured. Therefore, fundus image quality assessment (IQA) becomes an essential process, especially in the case of automated diagnosis systems. There are two types of methods available for assessing the quality of fundus images: (i) Subjective, and (ii) Objective. Subjective evaluation is carried out by the ophthalmologists who grade the fundus images into different quality classes, based on their previous experience (e.g. the experience ophthalmologists have gained from their ophthalmic diagnostic training). Subjective quality evaluation is assumed

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

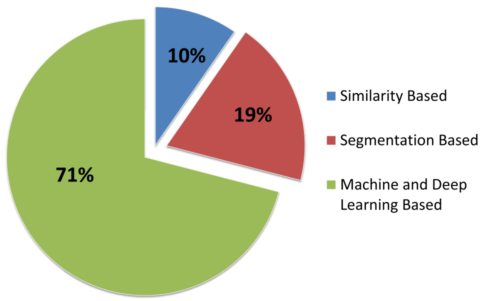


FIGURE 1. Pie chart summarizing the analysis of the fundus IQA algorithms in the literature.

to be the most reliable method, since ophthalmologists are the ultimate users of fundus images. However, it is an expensive method in terms of time, effort, and money. In contrast, objective quality assessment methods are mathematical models that classify the fundus images into categories of quality with the intention to estimate what obtained from subjective methods. For the objective quality assessment of natural images, much work is done, and various IQA metrics have been proposed to date [9]. In few works, IQA metrics developed for natural images have been adopted for medical images [10]. Specific IQA algorithms have been recently been proposed for different medical images, like magnetic resonance imaging (MRI), ultrasound imaging, and also fundus images [9]. The next subsection contains an overview of the previous fundus IQA works with theirs limitations. For further details, the reader can refer to [11].

A. PREVIOUS WORKS

Based on the study of the literature, fundus IQA algorithms can be classified into three categories: (i) Similarity-based, (ii) Segmentation based, (iii) Machine and Deep Learning based. Fig. 1 shows the percentage of the research works that adopted each type of methodology.

1) SIMILARITY BASED METHODS

These compare the features of the target fundus image with those of a set of good quality fundus images. Lee and Wang [12] proposed the first work on fundus IQA in 1999. A similarity measure was calculated between the intensity histogram of the reference template and the target image. The fundus images were graded into two classes of quality: good and poor. A set of good quality fundus images was used to form the reference template. Later, in 2001, Lalonde *et al.* [13] presented a work that uses the distribution of local intensity and edge magnitudes to derive the similarity between reference and input fundus image. The major limitation of the methods under this category is the assumption of a universal reference template of a good quality fundus image. Also, these methods are vulnerable to different types of distortions.

2) SEGMENTATION BASED METHODS

These involve the analysis of segmented objects from fundus images, like the optic disc, or blood vessels. Usher *et al.* [14] proposed a method that performs blood vessel segmentation and calculates the vessel density as a quality indicator. Kohler *et al.* [15] analyzed blur by segmenting the blood vessels and counting the pixels in the segmented area with the inclusion of generic image features. Inspired with a similar idea, the authors in [16] used the contrast property of the area corresponding to the blood vessels as a quality parameter. The methods under this category perform well over distortions like blur and uneven illumination. The limitations of segmentation based methods are the fix assumptions in terms of field of view, shape, and location of the structures visible in the image. These assumptions lead to low performance accuracy over cross data-set evaluation.

3) MACHINE AND DEEP LEARNING BASED METHODS

These classify the fundus images into two classes of quality, i.e., Good or Poor, by extracting some meaningful features from the image. Most of the works published for fundus IQA are based on this category of methods. We mention here a few recent works. Wang *et al.* [17] presented a fundus IQA algorithm that uses human visual system (HVS) based feature extraction methods. It is one of the few works where the authors have used a data-set of fundus images with subjective quality scores. It is important to mention here that subjective ratings were collected for three generic quality parameters on a scale of two (i.e., 0 or 1): (i) uneven illumination, (ii) blur, and (iii) contrast. All three parameters were extracted using multichannel sensation, just noticeable blur (JNB), and contrast sensitivity function (CSF) methods, respectively. Finally, the extracted features were used to divide the fundus images into the two categories of quality. Next, with a similar idea to [17], Shao *et al.* [18] presented a retinal IQA method. Illumination, naturalness, and structural parameters were quantified to classify the fundus images into two classes with a reported accuracy of 94.5%. Dias *et al.* [19] presented a fundus IQA method that uses four generic properties of an image: illumination, color, focus, and contrast, to classify the fundus images using a feed forward neural network. Recently, Abdel-Hamid *et al.* [20] analyzed five fundus image properties related to content and clarity: sharpness, illumination, homogeneity, field definition, and content. These quality indicators are derived using a wavelet-based feature extraction method.

Recently deep learning methods have achieved a stand out performance accuracy in IQA problems [21]–[23]. Using the advantages of CNNs, Yu *et al.* [24] presented a deep learning-based architecture that fuses the features extracted from convolution neural networks (CNN) and saliency map to classify the fundus images into two categories of quality. Similarly, Tennakoon *et al.* [25] also presented a shallow CNN network with four convolution and two fully connected layers for two-class retinal quality classification. Recently, Zago *et al.* [26] and Chalakkal *et al.* [27] have used the virtues of pretrained model architectures (GoogLeNet [28], AlexNet [29], and ResNet [30]) to classify fundus images into two categories.

B. LIMITATIONS

A careful study of the literature leads to the conclusion that the following limitations exist in the state-of-the-art of fundus IQA research.

1) FEW SUBJECTIVE INPUTS

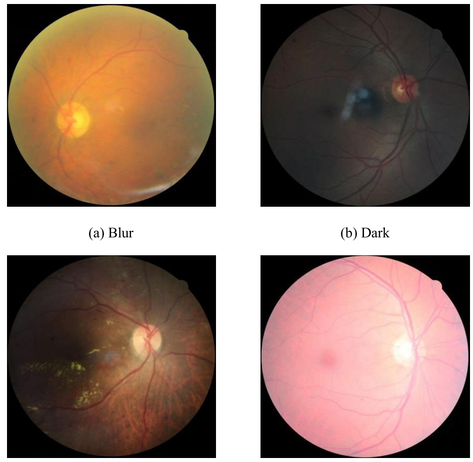
According to a study [19] held at the University of Wisconsin-Madison, the quality of a fundus image can be assessed using the following quality parameters: focus and clarity, field definition, visibility of the structures (i.e., macula, optical disc, and blood vessels). However, there exist only a few fundus IQA works that included a subjective opinion of a medical doctor about these quality parameters. As mentioned above, in [17] the authors have included the subjective evaluation of the fundus images using three generic quality parameters. However, the assessment of structural properties is not included and generic parameters give global quality information. To get the information about the local quality of an image, the evaluation of structural parameters is essential. Also, the ratings were collected on a scale of only two numbers (0 and 1), which is too small to identify the erroneous subjective inputs. Further, only three medical doctors participated in the subjective assessment, which also limits the generalizability of the data-set. In order to get a better understanding of the perceptual quality of a fundus image, it is essential to collect subjective opinions for both generic and structural quality parameters.

2) CATEGORIES OF QUALITY AND SCOPE OF ENHANCEMENT

In the case of medical images, the IQA process aims to find out their diagnostic usefulness. Hence, fundus IQA methods are used to classify the images into different categories of quality. As shown in Fig.1, most of the fundus IQA algorithms are developed using machine learning-based classification algorithms, with the aim to classify them into two categories of quality: Good and Poor. However, in real-time imaging scenarios, there also exists a type of fundus images that neither fall into good nor in the poor category. For example, the fundus images shown in Fig. 2 do contain visible artifacts, but still can be used for the diagnosis by the medical doctors. Hence, it cannot be put into ''Poor'' category of quality. At the same time, these images might lead to wrong diagnostic results from an automated diagnosis system; hence also should not be labeled as ''Good''.

Recently many methods aiming at enhancing the visual quality of fundus images [31]–[38] were published. A fully automated diagnosis system requires an effective fundus IQA algorithm that can also determine the requirement of enhancement. A binary classification based IQA method may not be able to provide such information. Hence, there must exist one more category of quality indicating an ''average'' or ''fair'' quality fundus image.

In order to address the above mentioned limitations, our contributions in this paper are as follows:



(a) Blur (b) Dark

(c) Uneven Illumination (d) Bright

FIGURE 2. Examples of average quality fundus images: (a) Blur, (b) Dark, (c) Uneven Illumination, and (d) Bright.

- A Fundus Image Quality Assessment (FIQuA) data-set of 1500 macula centered fundus images has been created, with three categories of quality: Good, Fair, and Poor. To get a clearer understanding of the ophthalmologists' perception, for each image in the data-set, subjective ratings in range of [0, 10] have been collected for six quality parameters, both structural and generic. To increase the generalizability of the data-set, subjective assessment is carried out by fifteen accomplished ophthalmologists.
- A multivariate linear regression-based convolutional neural network (CNN) model is proposed for the objective quality assessment of fundus images. The proposed model, trained with the help of the six subjective inputs, leads to achieving high classification accuracy.

To the best of our knowledge, the two contributions stated above have never been proposed earlier. The structure of the rest of the paper is as follows. Section II contains the detailed introduction of the proposed FIQuA data-set, including an analysis of the collected data. Section III explains in detail the proposed CNN model for fundus IQA. Section IV contains a detailed analysis of the experimental results. Finally, Section V discusses the conclusions and future work.

II. THE FUNDUS IMAGE QUALITY ASSESSMENT (FIQuA) DATA-SET

A. DESCRIPTION AND PECULIARITIES OF THE PROPOSED FIQuA DATA-SET

A total of 1500 fundus images were taken from the data-set provided by EyePACS at Kaggle.com [39] for the DR detection challenge. 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.*

TABLE 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)

- ***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.*

It is to mention that there is an equal number of images in each category of quality, i.e., 500 images per category. As mentioned earlier in Section I-B, 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 and generic fundus image properties. Also, a careful study of the previous work leads us to conclude that two types of quality parameters are used by the researchers are: (i) Structural, and (ii) Generic. Therefore, the identified six quality parameters are classified under two categories: (i) Structural, and (ii) Generic. Table 1 provides the classification of the quality parameters under the two categories. Throughout the paper F1-F6, as mentioned in Table 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 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 [40]. 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 and Table 2 the output of the SQA process is presented. Fig. 3 contains samples of the fundus images from the FIQuA dataset and Table 2 the respective subjective ratings.

B. 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:

FIGURE 3. **Samples of the fundus images from FIQuA dataset.**

TABLE 2. **Sample of the subjective scores and corresponding quality class graded by the ophthalmologists for the respective images shown in Fig. 3.**

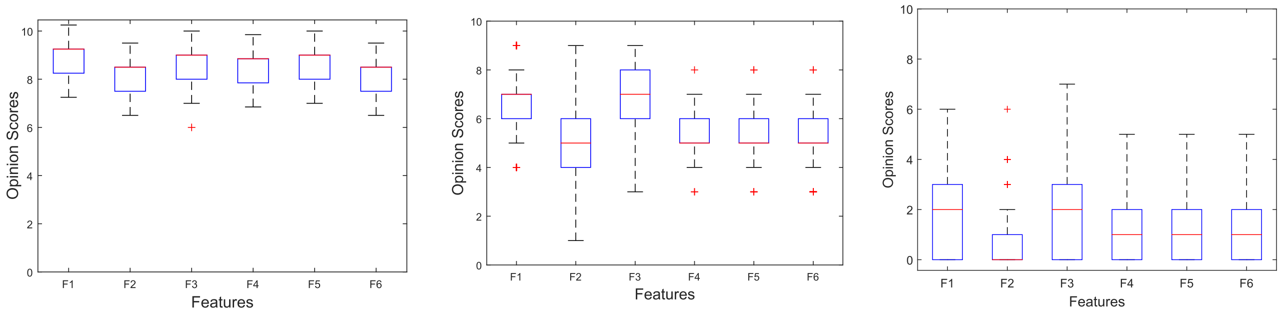
Image	F1	F2	F3	F4	F5	F6	Class
Image 1	10	9	10	9	9	10	Good
Image 2	8	6	7	5	6	6	Fair
Image 3	1	0	1	1	1	1	Poor

- validating the collected data inputs;
- 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 (MAD), given in the equation below:

$$MADN = c \; median(|X_i - median(X)|) \tag{1}$$

where c = 1.483 and X_i is the score provided by the medical expert i, with $i = 1, 2, \ldots, N$ 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 [41]. 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. 4 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 $10 \geq SV > 7$, $7 \geq SV \geq 5$, and $5 > SV \geq 1$ for the good, fair, and poor classes, respectively. Here, SV 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 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



(a) (b) (c)

FIGURE 4. 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.

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 5. Confusion Matrices for the each of the four classification results shown in Table 3.

Good & Fair and Fair & Poor classes. The confusion matrices for all the four classification algorithms mentioned in Table 3 is shown in Fig.5. 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 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 4. The obtained coefficient 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)".

III. FUNDUS IQA MODEL

The proposed fundus IQA model is a two-step process: *Block-1:* Multivariate linear regression-based CNN model that extracts optimized features against training for the subjective scores of F1-F6, and *Block-2:* Fusion of the optimized features obtained from step Block-1 for the classification. The

TABLE 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 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

comparison between the previous fundus IQA work and the proposed model is illustrated in Fig. 6. CNNs have proved to give extraordinary results not only in case of image classification [28], [29] and object detection tasks [42]–[44] but also for quality assessment [45]–[48]. The motivation for using CNNs is the reported performance of CNN based IQA models [21]–[23] for natural images. These reported works proved that CNN models are very effective for IQA and outperform the state-of-the-art methods. The architecture of the proposed fundus IQA model is shown in Fig. 7. The next subsections provide the description of the aforementioned steps.

A. MODEL DESCRIPTION

The proposed model is built leveraging on two popular concepts of learning based algorithms: (i) Transfer learning [49], and (ii) Ensemble learning [50]. As anticipated above and

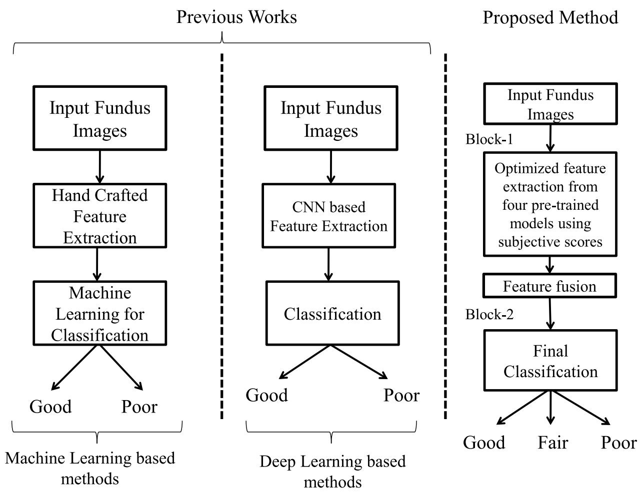


FIGURE 6. Comparison Flow Chart of the state of the art fundus IQA methods and the proposed method.

illustrated in Fig. 7, the model is divided into two blocks. A detailed description of each block is given below:

Block-1: The objective of this block is to derive the optimized features for the final classification. Transfer learning has been used to achieve the objective. Transfer learning is a popular machine learning strategy where weights obtained from popular pre-trained networks on ImageNet [51] alike large data-sets, are used as initial parameters to train another network. These pre-trained CNN models, like AlexNet [29], GoogLeNet [28], ResNet [30], DenseNet [52], Xception [53], etc., are used to solve other object detection and classification problems, not only in the domain of natural images but also for other image domains. The reason for adopting the transfer learning methodology is the limited number of fundus images available for the training phase. Training a network from scratch requires a sufficiently large number of images to get the optimal values for the network weights. Recently, transfer learning methods are also used to address the challenges of fundus image quality assessment [26], [27], [54].

As an initial setting for the training, we have used the weights of the following four pre-trained models: ResNet [30], DenseNet [52], Inception-V3 [55], and Xception [53]. ResNet is a deep residual learning based CNN architecture proposed by He *et al.* [30]. ResNet50, ResNet101, and ResNet152 are its variants, where 50, 101, and 152 indicate the number of layers present in the architecture, respectively. DenseNet{121, 169, 210} was proposed by Huang *et al.* [52] in 2017. The "dense" term indicates that each layer of this CNN model is connected to every layer of the architecture. Here 121, 169, and 201 indicate the depth of the model. Next, Inception-v3 is a successor version of GoogLeNet that also named Inception-v1. Each inception layer is built with six convolution layers, followed by one pooling layer. Finally, the Xception architecture is a linear stack of depthwise separable convolution layers with residual connections [53]. Each model is trained individually on the subjective scores of F1-F6, by adding *five* fully connected (FC) layers at the end of the each network. The details of the FC layers are as follows: FC1: 1024×1, FC2: 512×1, FC3: 120×1, FC4: 24×1, FC5: 12×1. Here the first four FC layers are followed by the rectified linear unit (ReLu) [56] activation function. The mathematical representation of the ReLu is given below:

$$y = max(0, x). \tag{2}$$

It produces the output y as x if the value of input x is positive and 0 otherwise. The ReLu activation is used because of its advantage over sigmoid and hyperbolic tangent activation functions as it avoids the vanishing gradient problem. The last FC5 layer, with the inclusion of sigmoid function, performs multivariate regression to derive the six numerical values corresponding to the F1-F6 quality parameters.

Fig. 7 shows that the CNN model takes the input image of size $512 \times 512 \times 3$ and in the fifth FC layer transforms it into a feature vector of size 12×1. In the last FC layer the model performs the multivariate linear regression onto the desired feature vector of size 6×1. Let $X_{(i)12\times1}$ be the input feature vector obtained at the fourth FC layer and $Y_{(i)6\times1}$ is the associated score vector for the i^{th} image. Then, the multivariate linear regression model can be represented as:

$$\hat{Y}_i = W_i X_i + E_i \tag{3}$$

where

- $\hat{Y}_i = [\hat{y}_{i1}, \hat{y}_{i2}, \hat{y}_{i3}, \hat{y}_{i4}, \hat{y}_{i5}, \hat{y}_{i6}]$ is the 6×1 predicted score vector for the i^{th} image.
- $X_i = [x_{i1}, x_{i2}, x_{i3} \ldots, x_{i12}]$ is the 12×1 input feature vector for i^{th} image.
- $W_i = [W_{i1}, W_{i2}, W_{i3} \ldots, W_{i6}]$ is the 6×12 weight matrix for the i^{th} image.
- $W_{ij} = [w_{ij_1}, w_{ij_2}, w_{ij_3}, \ldots, w_{ij_{12}}]$ is the 1×12 weight vector for j^{th} feature. Here, $j = 1, 2, 3, \ldots 6$.
- Finally, $E_i = [e_{i1}, e_{i2}, e_{i3}, e_{i4}, e_{i5}, e_{i6}]$ is the corresponding error matrix of size similar to Y.

It is important to mention that the batch normalization [57] method is used for the regularization of the model to avoid the over-fitting problem. Batch normalization is preferred over the dropout [58] method as empirical results were better than in the case of batch normalization. All four models were trained to achieve the maximum correlation with the subjective scores of F1-F6. Furthermore, once each of the models was trained for the maximum correlation, the values of FC3 layers from each model were assembled and transferred to Block-2. The accuracy of the correlation results is discussed in Section IV.

Block-2: This block uses the concepts of both transfer learning and ensemble learning. The objective of ensemble learning is to collect the predictions from different models to conclude with better prediction results [50]. The optimized features of the FC3 (120×1) layer from each of the four models of Block-1 are combined to form the FC6: 480×1 layer and transferred to Block-2. Block-2 consists of 5 fully connected layers: FC6: 480×1, FC7: 120×1, FC8: 24×1, FC9: 12×1, FC10: 6×1 and finally the classification results.

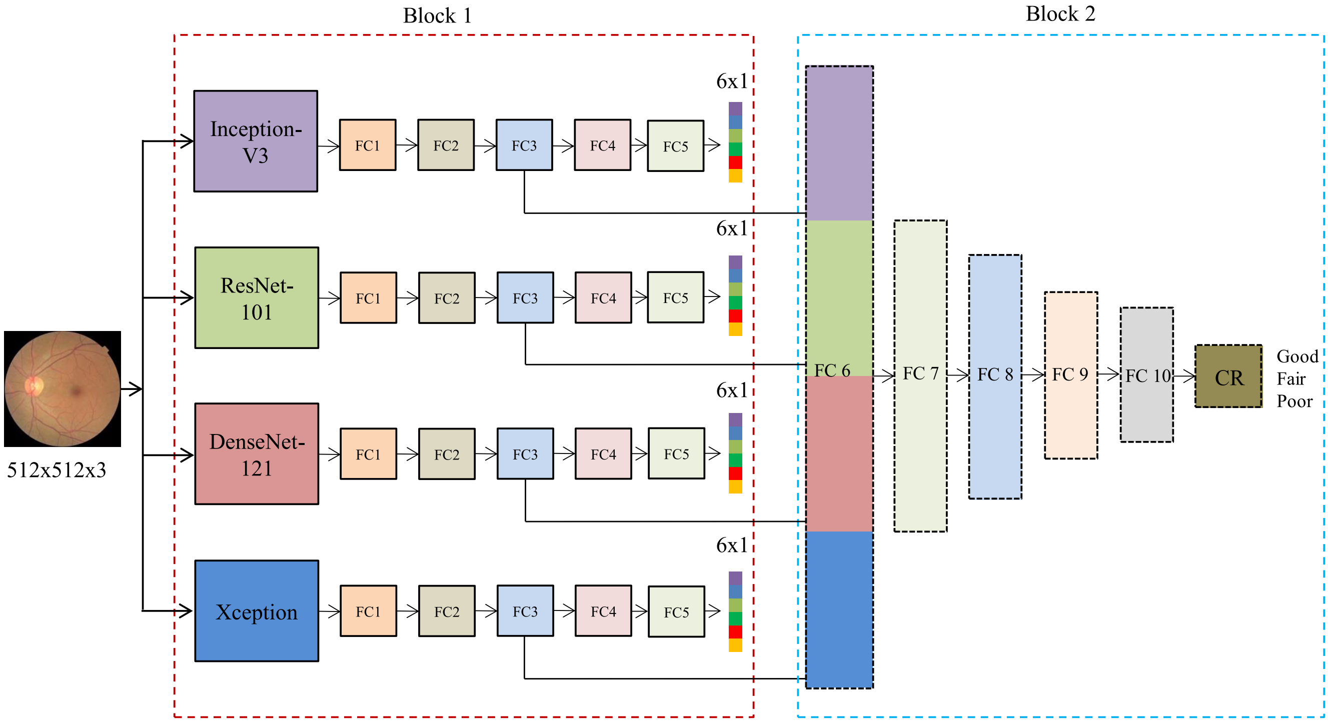


FIGURE 7. **Proposed CNN Model. FC: Fully Connected Layer, FC1: 1024 × 1, FC2: 512 × 1, FC3: 120 × 1, FC4: 24 × 1, FC5: 12 × 1, FC6: 480 × 1, FC7: 120 × 1, FC8: 24 × 1, FC9: 12 × 1, FC10: 6 × 1, CR: Classification Result.**

It is important to mention that the training of each block presented here is done individually. Block-1 was trained until the optimized features were derived. Afterwards, Block-2 was trained to get the optimized classification results. Similar to the previous block, the ReLu activation function follows each FC layer in Block-2 after the FC10 layer *softmax* function is applied to get the desired classification results.

B. IMPLEMENTATION DETAILS

- *Pre-processing:* Fundus images carry a large area of black background that might affect the training accuracy. Therefore, all the images were cropped to the boundary of the fundus area in order to reduce the area of black background. It is achieved by traversing the nearest pixel values that are close to zero to the center co-ordinates of the images. In addition, the fundus images provided on Kaggle are of high resolution. Hence, each image is further resized to the dimension of 512 × 512.
- *Loss Function:* In Block-1, the mean square error (MSE) function is used as the loss function, and can be represented as:

$$L_{B1} = \frac{1}{N}\sum_{i=1}^{N} ||(Y - \hat{Y})||^2 \tag{4}$$

where L_{B1} represents the loss computed for the Block-1, Y and $\hat{Y}$ represent the actual value and predicted value respectively, and N represents the number of samples. Moreover, in Block-2 the categorical cross entropy loss function is used. Its mathematical representation is as follows:

$$L_{B2} = -\sum_{i=1}^{C} P_i log(\hat{P}_i) \tag{5}$$

Here, L_{B2} represents the loss computed for the Block-2, C represents the total number of classes, P and $\hat{P}$ represent the actual and predicted output respectively. It is important to mention that the *softmax* activation function should be applied to the target before computing the categorical loss.
- The back-propagation and adaptive moment estimation (ADAM) [59] optimization methods are used for error minimization with learning rate of 10^{-4}. ADAM has been performed for *1000* epochs with the mentioned batch size of *8* images during the training process.
- Out of 1500 images, 1200 were used for the training and 300 for testing purpose. Here, 400 images were taken from each class for training and similarly 100 images from each category for testing.
- All the experiments were carried out on a computer system of 2.0 GHz CPU and GTX1080 Ti GPU and the CNN model is implemented using the Python programming language with Keras library.

IV. RESULTS AND ANALYSIS

1) EVALUATION METHODOLOGY

Four commonly used standard measures recommended by the Video Quality Experts Group [60] have been used to evaluate the performance of Block-1. These are the Spearman rank-order correlation coefficient (SROCC), the Kendall

TABLE 5. Correlation coefficients for the predicted values of F1-F6.

	Feature	SROCC	PLCC	KCC	RMSE
Inception-V3	F1	0.9299	0.9544	0.8463	0.4163
	F2	0.9309	0.9392	0.8319	0.4294
	F3	0.9005	0.9346	0.7932	0.4453
	F4	0.9281	0.9413	0.8281	0.4187
	F5	0.9413	0.9512	0.8517	0.4014
	F6	0.9388	0.9477	0.8454	0.4004
ResNet-101	F1	0.8758	0.8906	0.7260	0.7646
	F2	0.8920	0.8644	0.7336	0.7731
	F3	0.8378	0.8588	0.6724	0.8531
	F4	0.8815	0.8916	0.7224	0.6887
	F5	0.8981	0.9032	0.7478	0.5871
	F6	0.9025	0.9066	0.7517	0.7571
DenseNet-121	F1	0.9011	0.9008	0.7477	0.7176
	F2	0.8906	0.8981	0.7404	0.7362
	F3	0.8706	0.8782	0.7072	0.8116
	F4	0.9032	0.9053	0.7463	0.5966
	F5	0.9133	0.9148	0.7647	0.4947
	F6	0.9079	0.9113	0.7581	0.6835
Xception	F1	0.9293	0.9469	0.8532	0.4262
	F2	0.9230	0.9348	0.8369	0.4294
	F3	0.9007	0.9363	0.7898	0.4406
	F4	0.9225	0.9324	0.8225	0.4220
	F5	0.9326	0.9398	0.8482	0.4106
	F6	0.9269	0.9333	0.8402	0.4262

rank-order correlation coefficient (KCC), the Pearson Linear correlation coefficient (PLCC), and the root-mean-square error (RMSE). For the performance measurement of an IQA metric, SROCC and KCC evaluate the prediction monotonicity. The other two, PLCC and RMSE, measure the prediction accuracy. Higher values obtained in SROCC, KCC, and PLCC for an IQA metric indicate higher performance, whereas lower values of RMSE are associated with better performance. Furthermore, to evaluate the performance of Block-2 the following statistical parameters are used:

$$A = \frac{T}{N} * 100 \tag{6}$$

$$P = \frac{T_p}{T_p + F_p} \tag{7}$$

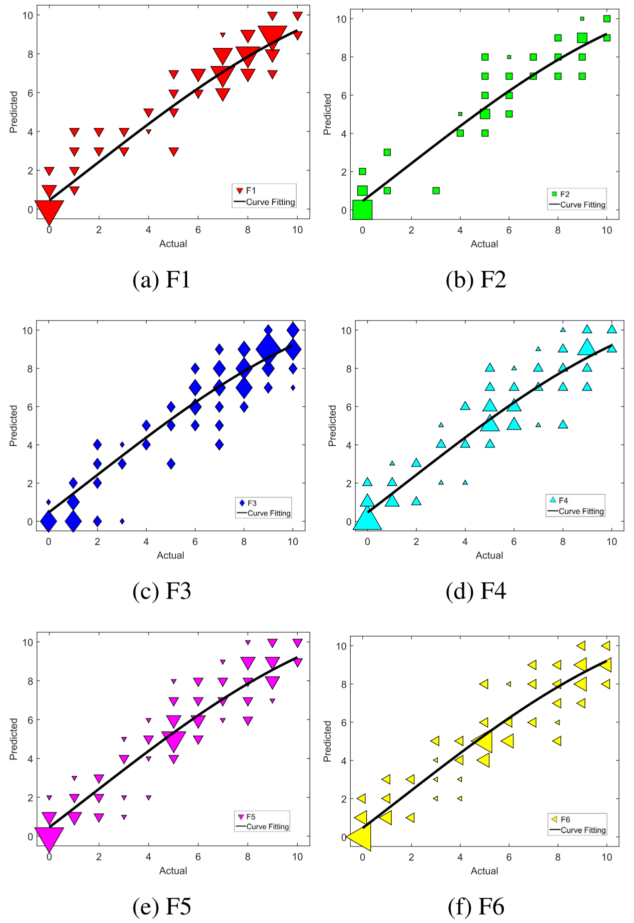
$$R = \frac{T_p}{T_p + F_n} \tag{8}$$

$$F_m = 2 * \left(\frac{PR}{P+R}\right). \tag{9}$$

Here A = Classification accuracy, P = Precision, R = Recall, T = Total number of correct classifications, N = Total number of samples, T_p = true positive, F_p = false positives, F_n = false negatives, and F_m = F-measure.

2) PERFORMANCE EVALUATION OF BLOCK-1

The feature-wise performance of each of the four models is shown in Table 5, reporting the correlation values calculated between the derived scores and the subjective score values for each quality parameter F1-F6. In addition, Table 5 shows that the highest results obtained for the SRCC, PLCC, and KCC are *0.94*, *0.95*, and *0.85* respectively and for RMSE the lowest result is *0.40*. It validates that the proposed model achieves a significantly high correlation between the subjective and derived scores. Furthermore, scatter plots with curve fitting of the mean of predicted values from each of the four models are shown in Fig. 8. These plots are obtained after performing logistic regression between predicted values and subjective OS values. These curves are obtained after non-linear fitting, as suggested in [61]. It can be observed from Fig. 8 that the consistency between the predicted and subjective values is very high. Here, the size of the object represents the frequency of the predicted values corresponding to the actual value, whereas larger size objects correspond to a higher frequency. It can also be observed that all the larger size objects lie in the close vicinity of the curve, indicating a high correlation between actual and predicted values. These high correlation results validate that the features obtained in previous FC layers are optimized. Now, the optimized features of FC-3 are ensembled together and transferred to Block-2 for the final classification of images.



(a) F1 (b) F2 (c) F3 (d) F4 (e) F5 (f) F6

FIGURE 8. Feature-wise plot of the predicted scores versus actual opinion scores.

3) PERFORMANCE EVALUATION OF BLOCK-2

Initially, the individual classification performance of different variants of each of the four models has been analyzed. Here, individual performance indicates that the 240 × 1 feature vector derived from Block-1 is used only to train Block-2 for

TABLE 6. Performance evaluation of different models for classification results on FIQuA data-set.

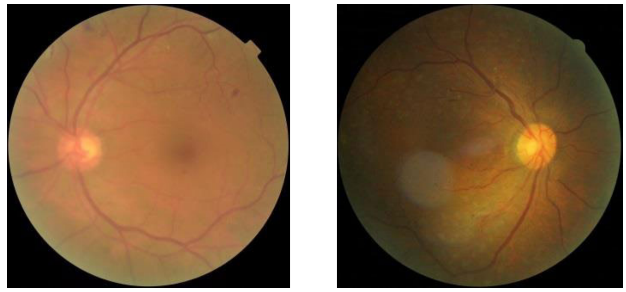
Model	Accuracy (in%)	F-measure	Precision	Recall
ResNet50 [30]	90.33	0.9032	0.9031	0.9033
ResNet101 [30]	91.66	0.9165	0.9164	0.9166
ResNet152 [30]	90.33	0.9032	0.9031	0.9033
DenseNet121 [52]	92.33	0.9233	0.9233	0.9233
DenseNet169 [52]	89.66	0.8963	0.8960	0.8966
DenseNet201 [52]	90.66	0.9065	0.9064	0.9066
Inception-V3 [55]	93.00	0.9300	0.9300	0.9300
Xception [53]	93.33	0.9334	0.9335	0.9333
Proposed	**95.66**	**0.9566**	**0.9565**	**0.9566**

Predicted Class

Actual Class

Class	Good	Fair	Poor
Good	96	4	0
Fair	4	93	3
Poor	0	2	98

FIGURE 9. Confusion matrix of the prediction results obtained on FIQuA data-set from the proposed fundus IQA model.



(a) (b)

FIGURE 10. Sample images with different distortions from the Fair category of the FIQuA data-set that are correctly classified by the proposed model. Here (a) and (b) represent the images distorted with Blur and Uneven Illumination distortion, respectively.

TABLE 7. Performance evaluation of proposed method over DRIMDB and Eye-Quality (EyeQ) data-set.

Data-set	Accuracy (in %)	Precision	Recall	F-measure
DRIMDB	98.96	0.9889	0.9859	0.9920
EyeQ	88.43	0.8697	0.8700	0.8694

TABLE 8. Performance summary of recent fundus IQA works over DRIMDB and EyeQ data-set. Here (+) indicates that the work also includes fundus images from other proprietary data-sets.

Work	Year	Data-set	No. of Images	Accuracy (in %)
[17]	2016	DRIMDB (+)	536	94.52
[18]	2018	DRIMDB (+)	4372	92.39
[26]	2018	DRIMDB (+)	1036	98.56
[27]	2019	DRIMDB (+)	7007	97.70
[54]	2019	EyeQ	28792	91.75

final classification. Table 6 contains the performance results of Block-2 with three variants of both ResNet and DenseNet. It indicates that the Xception model achieves the highest individual accuracy (93.33%). However, the performance of the proposed ensemble model after the fusion of features got approximately 2% jump on overall accuracy with 95.66%. The confusion matrix of the prediction results of the proposed method is shown in Fig. 9. It can be observed from Fig. 5 and Fig. 9 that the accuracy of the proposed fundus IQA model is closely similar to the results of the classification using subjective scores. It indicates that the inclusion of subjective scores greatly helps to train the model to derive the optimized features for the classification. Also, for illustration purposes, two example images from the Fair category of the FIQuA data-set are shown in Fig. 10. Here, (a) and (b) are the sample images distorted with blur and uneven illumination distortions, respectively. It can be observed from the Fig. 10 that all the structural information is quite visible, yet due to the presence of a small proportion of distortions, ophthalmologists labeled them as a fair quality image. The proposed model also correctly classified these images as fair quality. It indicates the robustness of the model as it efficiently mimics the visual perception of ophthalmologists by detecting these distortions in the image.

4) CROSS DATA-SET EVALUATION

The proposed fundus IQA model trained over the FIQuA data-set was also evaluated over two publicly available data-sets: DRIMDB [62] and EyeQ [54], specifically developed for fundus IQA. The DRIMDB [62] data-set was presented by U. Sevik. It contains 216 fundus images with three classes: Good (125), Poor (69), and Outlier (22). Next, Fu, *et al.* made a commendable effort and recently presented a large scale EyeQ data-set. The EyeQ data-set consists of 28,792 fundus images divided (with analogy to our approach) into three categories: Good, Usable, and Reject. Table 7 contains the classification results over the above mentioned data-sets. The results indicate that the proposed fundus IQA model achieves high classification accuracy over an unknown and large scale data-set given it was trained on a comparatively small data-set. Also, for comparison purposes, a performance summary of recent fundus IQA works that are developed and evaluated over DRIMDB and EyeQ data-set is presented in Table 8. It can be observed from both Table 7 and 8 that the performance of the proposed model outperforms the recent fundus IQA methods over the mentioned data-sets. It is essential to mention that despite being trained over a comparatively too small data-set (FIQuA), the performance of the proposed model is very close to the model proposed in [54]. It shows that the inclusion of adequate subjective inputs not only increases the performance of the model but also its generalizability over unknown image inputs. In our future work, we are planning to use reinforcement learning methods to achieve higher accuracy over the EyeQ data-set.

V. CONCLUSION

Ophthalmologists assess the quality of fundus images based on two quality parameters: Structural and Generic. This paper aims at assessing the quality of fundus images on similar grounds. First, a new data-set of 1500 images (FIQuA) has been prepared with a total of *seven* subjective inputs from ophthalmologists. Out of the seven inputs, the first six are subjective scores for six quality parameters (F1-F6) and the last one is the class of quality (good, fair, or poor). Second, a new multivariate linear regression based CNN model for fundus image quality assessment is presented. The peculiarity of the model is that it derives the optimized features for classification, using the subjective inputs provided by the ophthalmologists. It consists of two blocks: Block-1 derives the optimized features from four pre-trained CNN models: Inception-V3, ResNet-151, DenseNet-121, and Xception that are trained through transfer learning against the six subjective scores provided by the ophthalmologists. Further, these optimized features are ensembled together and forwarded to Block-2 to classify the fundus images into three classes: Good, Fair, and Poor. The results show that the proposed CNN model achieves a high correlation with subjective values. The correlation values obtained from CNN Block-1 for SROCC, LCC, and KCC for each quality parameter (F1-F6) are approximately *0.941*, *0.954*, and *0.853* respectively, and for RMSE the result is *0.401*. It indicates that for each of the six features, the derived quality scores from the proposed model are closely similar to the subjective quality scores provided by the medical doctors. Further, using the derived ensembled features, the classification accuracy achieved by the CNN Block-2 is *95.66%*. It proves that the inclusion of the subjective scores helps achieving a high classification accuracy. In the future, we are planning to increase the performance accuracy of the proposed model over unknown image inputs by applying domain adaptation techniques as a preprocessing step to bring unknown test images into the training image domain.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Meena from All India Institute of Medical Sciences (AIIMS) Jodhpur, and Dr. A. Chouhan from Mathura Das Mathur (MDM) Hospital, Jodhpur, and their colleagues for their valuable inputs.

REFERENCES

[1] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Comput. Ind.*, vol. 69, pp. 72–80, May 2015.

[2] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: A review," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2136–2155, 2013.

[3] G. D. Joshi, J. Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE Trans. Med. Imag.*, vol. 30, no. 6, pp. 1192–1205, Jun. 2011.

[4] M. U. Akram, A. Tariq, S. A. Khan, and M. Y. Javed, "Automated detection of exudates and macula for grading of diabetic macular edema," *Comput. Methods Programs Biomed.*, vol. 114, no. 2, pp. 141–152, Apr. 2014.

[5] *Number of Ophthalmologists in Practice and Training Worldwide*. Accessed: Jan. 10, 2019. [Online]. Available: http://www.icoph.org/ophthalmologists-worldwide.html

[6] L. Shi, H. Wu, J. Dong, K. Jiang, X. Lu, and J. Shi, "Telemedicine for detecting diabetic retinopathy: A systematic review and meta-analysis," *Brit. J. Ophthalmol.*, vol. 99, no. 6, pp. 823–831, Jun. 2015.

[7] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Med.*, vol. 19, no. 2, pp. 105–112, 2002.

[8] A. Bourouis, M. Feham, M. A. Hossain, and L. Zhang, "An intelligent mobile based decision support system for retinal disease diagnosis," *Decis. Support Syst.*, vol. 59, pp. 341–350, Mar. 2014.

[9] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Process.*, vol. 2013, pp. 1–53, Feb. 2013.

[10] M. Razaak, M. G. Martini, and K. Savino, "A study on quality assessment for medical ultrasound video compressed via HEVC," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 5, pp. 1552–1559, Sep. 2014.

[11] A. Raj, A. K. Tiwari, and M. G. Martini, "Fundus image quality assessment: Survey, challenges, and future scope," *IET Image Process.*, vol. 13, no. 8, pp. 1211–1224, Jun. 2019.

[12] S. C. Lee and Y. Wang, "Automatic retinal image quality assessment and enhancement," *Med. Imag., Image Process., Int. Soc. Opt. Photon.*, vol. 3661, May 1999, pp. 1581–1591.

[13] M. Lalonde, L. Gagnon, and M.-C. Boucher, "Automatic visual quality assessment in optical fundus images," in *Proc. Vis. Interface*, Ottawa, ON, Canada, vol. 32, 2001, pp. 259–264.

[14] D. Usher, M. Himaga, M. J. Dumskyj, and J. Boyce, "Automated assessment of digital fundus image quality using detected vessel area," in *Proc. Med. Image Understand. Anal.*, 2003, pp. 81–84.

[15] T. Kohler, A. Budai, M. F. Kraus, J. Odstrcilik, G. Michelson, and J. Hornegger, "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation," in *Proc. 26th IEEE Int. Symp. Comput. Med. Syst. (CBMS)*, Jun. 2013, pp. 95–100.

[16] H. A. Nugroho, T. Yulianti, N. A. Setiawan, and D. A. Dharmawan, "Contrast measurement for no-reference retinal image quality assessment," in *Proc. 6th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2014, pp. 1–4.

[17] S. Wang, K. Jin, H. Lu, C. Cheng, J. Ye, and D. Qian, "Human visual system-based fundus image quality assessment of portable fundus camera photographs," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1046–1055, Apr. 2016.

[18] F. Shao, Y. Yang, Q. Jiang, G. Jiang, and Y.-S. Ho, "Automated quality assessment of fundus images via analysis of illumination, naturalness and structure," *IEEE Access*, vol. 6, pp. 806–817, 2018.

[19] J. M. Pires Dias, C. M. Oliveira, and L. A. da S. Cruz, "Retinal image quality assessment using generic image quality indicators," *Inf. Fusion*, vol. 19, pp. 73–90, Sep. 2014.

[20] L. Abdel-Hamid, A. El-Rafei, S. El-Ramly, G. Michelson, and J. Hornegger, "Retinal image quality assessment based on image clarity and content," *J. Biomed. Opt.*, vol. 21, no. 9, Sep. 2016, Art. no. 096007.

[21] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.

[22] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.

[23] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.

[24] F. Yu, J. Sun, A. Li, J. Cheng, C. Wan, and J. Liu, "Image quality classification for DR screening using deep learning," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 664–667.

[25] R. Tennakoon, D. Mahapatra, P. Roy, S. Sedai, and R. Garnavi, "Image quality classification for DR screening using convolutional neural networks," in *Proc. Ophthalmic Med. Image Anal. 3rd Int. Workshop*, Oct. 2016, pp. 1–9.

[26] G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles, "Retinal image quality assessment using deep learning," *Comput. Biol. Med.*, vol. 103, pp. 64–70, Dec. 2018.

[27] R. J. Chalakkal, W. H. Abdulla, and S. S. Thulaseedharan, "Quality and content analysis of fundus images using deep learning," *Comput. Biol. Med.*, vol. 108, pp. 317–331, May 2019.

[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions,", in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 1–9. [Online]. Available: http://arxiv.org/abs/1409.4842

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[31] K. Noronha, J. Nayak, and S. N. Bhat, "Enhancement of retinal fundus image to highlight the features for detection of abnormal eyes," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2006, pp. 1–4.
[32] G. D. Joshi and J. Sivaswamy, "Colour retinal image enhancement based on domain knowledge," in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, Dec. 2008, pp. 591–598.
[33] M. H. A. Fadzil, H. A. Nugroho, H. Nugroho, and I. L. Iznita, "Contrast enhancement of retinal vasculature in digital fundus image," in *Proc. Int. Conf. Digit. Image Process.*, Mar. 2009, pp. 137–141.
[34] T. Shimahara, T. Okatani, and K. Deguchi, "Contrast enhancement of fundus images using regional histograms for medical diagnosis," in *Proc. SICE Annu. Conf.*, vol. 1, 2004, pp. 650–653.
[35] C.-Y. Lu, B.-Z. Jing, P. P. K. Chan, D. Xiang, W. Xie, J. Wang, and D. S. Yeung, "Vessel enhancement of low quality fundus image using mathematical morphology and combination of Gabor and matched filter," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, Jul. 2016, pp. 168–173.
[36] U. Farooq and N. Y. Sattar, "Improved automatic localization of optic disc in retinal fundus using image enhancement techniques and SVM," in *Proc. IEEE Int. Conf. Control Syst., Comput. Eng. (ICCSCE)*, Nov. 2015, pp. 532–537.
[37] M. Zhou, K. Jin, S. Wang, J. Ye, and D. Qian, "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 3, pp. 521–527, Mar. 2018.
[38] H. A. Rahim, A. S. Ibrahim, W. M. D. W. Zaki, and A. Hussain, "Methods to enhance digital fundus image for diabetic retinopathy detection," in *Proc. IEEE 10th Int. Colloq. Signal Process. Appl.*, Mar. 2014, pp. 221–224.
[39] *Diabetic Retinopathy Detection*. Accessed: Apr. 16, 2019. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection
[40] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Rec. ITU-R BT.500-13, 2012. [Online]. Available: https://www.itu.int/dmspubrec/itu-r/rec/bt/R-REC-BT.500-11-200206-S!!PDF-E.pdf
[41] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *J. Experim. Social Psychol.*, vol. 49, no. 4, pp. 764–766, Jul. 2013.
[42] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.
[43] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.
[44] Z. Zhang, R. Jiang, S. Mei, S. Zhang, and Y. Zhang, "Rotation-invariant feature learning for object detection in VHR optical remote sensing images by double-net," *IEEE Access*, vol. 8, pp. 20818–20827, 2020.
[45] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3773–3777.
[46] J. Kim, A.-D. Nguyen, and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 11–24, Jan. 2019.
[47] L. Kang, P. Ye, Y. Li, and D. Doermann, "A deep learning approach to document image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2570–2574.
[48] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.
[49] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[50] Y. Liu and X. Yao, "Ensemble learning via negative correlation," *Neural Netw.*, vol. 12, no. 10, pp. 1399–1404, Dec. 1999.
[51] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[52] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[53] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[54] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu, and L. Shao, "Evaluation of retinal image quality assessment networks in different color-spaces," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Shenzhen, China: Springer, 2019, pp. 48–56.
[55] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[56] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[57] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[58] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[60] A. M. Rohaly *et al.*, "Video quality experts group: Current results and future directions," in *Proc. Vis. Commun. Image Process., Int. Soc. Opt. Photon.*, vol. 4067, 2000, pp. 742–754.
[61] E. C. Larson and D. M. Chandler, "Unveiling relationships between regions of interest and image fidelity metrics," in *Proc. Vis. Commun. Image Process., Int. Soc. Opt. Photon.*, vol. 6822, 2008, p. 68222A.
[62] U. Sevik, C. Köse, T. Berber, and H. Erdöl, "Identification of suitable fundus images using automated quality assessment methods," *J. Biomed. Opt.*, vol. 19, no. 4, Apr. 2014, Art. no. 046006.

ADITYA RAJ (Student Member, IEEE) received the B.Tech. degree in computer science and engineering from the Kanpur Institute of Technology, Kanpur, in 2010, and the M.Tech. degree in computer science and engineering from Guru Gobind Singh Indraprastha University, New Delhi, in 2013. He is currently pursuing the Ph.D. degree with IIT Jodhpur, India. From March to June 2019, he worked as a Visiting Researcher with Kingston University London, Kingston, U.K., through the Newton Bhabha Ph.D. Placement Program. His research interests are image quality assessment and medical image processing.

NISARG A. SHAH is currently pursuing the B.Tech. degree in electrical engineering with IIT Jodhpur, Jodhpur, India. He has prime interest in the field of deep learning and its applications in computer vision, especially in biomedical image analysis. He had opportunity to work with the researchers from IIT Jodhpur and the Indian Statistical Institute, Kolkata, early in his career.

ANIL KUMAR TIWARI (Member, IEEE) received the bachelor's degree in electronics engineering from Nagpur University, and the M.Tech. and Ph.D. degrees from IIT Kharagpur. He is an Associate Professor and the Head of the Department of Electrical Engineering, IIT Jodhpur. His current research interests include the development of algorithms for image and video compression, noise minimization, interpolation, and watermarking. He has also interest in the development of affordable healthcare technology. He has nearly 50 publications in national and internationally reputed conferences and journals.

MARIA G. MARTINI (Senior Member, IEEE) received the Laurea degree *(summa cum laude)* in electronics engineering from the University of Perugia, Italy, in 1998, and the Ph.D. degree in electronics and computer science from the University of Bologna, Italy, in 2002.

She is currently a Professor with the Faculty of Science, Engineering, and Computing, Kingston University London, where she also leads the Wireless Multimedia Networking Research Group. She has led the KU Team in a number of national and international research projects such as OPTIMIX, CONCERTO, QoE-NET, and Qualinet, funded by the European Commission, U.K. research councils, U.K. Technology Strategy Board/Innovate U.K., and international industries. She has authored more than 150 scientific articles, contributions to standardization groups (IEEE and ITU), and several patents on wireless video. Her research interests include QoE-driven wireless multimedia communications, decision theory, video quality assessment, and medical applications. She has chaired/organized a number of conferences and workshops. She is a part of international committees and expert groups, including the NetWorld 2020 European Technology Platform Expert Advisory Group, the Video Quality Expert Group (VQEG), and the IEEE Multimedia Communications Technical Committee, where she has served as a Vice-Chair, from 2014 to 2016, a Chair of the 3D Rendering, Processing, and Communications Interest Group, from 2012 to 2014, and a Key Member of the QoE and Multimedia Streaming IG. She is an Expert Evaluator of the European Commission, EPSRC, and other national and international funding bodies. She was an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, from 2014 to 2018. She is currently Associate Editor of the *IEEE Signal Processing Magazine*. She has also been the Lead Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on QoE-aware wireless multimedia systems and a Guest Editor of the IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, the IEEE MULTIMEDIA, the *International Journal of Telemedicine and Applications*, and other related journals.