Computers in Biology and Medicine 108 (2019) 317–331

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Quality and content analysis of fundus images using deep learning

Renoh Johnson Chalakkal[a,*], Waleed Habib Abdulla[a], Sinumol Sukumaran Thulaseedharan[b]

[a] Electrical, Computer and Software Engineering Department, University of Auckland, New Zealand
[b] Department of Ophthalmology, Government Medical College, Kerala, India

ARTICLE INFO

Keywords:
Retinal image quality analysis
Fundus images
Deep learning
Transfer learning

ABSTRACT

Automatic retinal image analysis has remained an important topic of research in the last ten years. Various algorithms and methods have been developed for analysing retinal images. The majority of these methods use public retinal image databases for performance evaluation without first examining the retinal image quality. Therefore, the performance metrics reported by these methods are inconsistent. In this article, we propose a deep learning-based approach to assess the quality of input retinal images. The method begins with a deep learning-based classification that identifies the image quality in terms of sharpness, illumination and homogeneity, followed by an unsupervised second stage that evaluates the field definition and content in the image. Using the inter-database cross-validation technique, our proposed method achieved overall sensitivity, specificity, positive predictive value, negative predictive value and accuracy of above 90% when tested on 7007 images collected from seven different public databases, including our own developed database—the UoA-DR database. Therefore, our proposed method is generalised and robust, making it more suitable than alternative methods for adoption in clinical practice.

1. Introduction

Image quality assessment (IQA) in image processing has remained an important area of research over the past decade and IQA techniques specific to the field of retinal image analysis have been well researched [1–13]. Recently, retinal images have been used to build many automatic pathology screening procedures. However, it has been reported that approximately 10% and 21% of the retinal images taken using fundus cameras for dilated and non-dilated pupils, respectively, are unsuitable for medical screening or diagnosis and, therefore, they cannot be used in automatic retinal screening systems (ARSS) [6]. To increase the robustness of ARSS, the IQA of the input retinal image has to be performed at the time of image capturing. If not, the performance metrics may be inconsistent [3,14,15]. Retinal image quality may be affected by various artefacts, such as dust or eyelashes on the fundus camera lens, out-of-focus imaging, bad illumination or blurring. These artefacts affect the image sharpness, illumination and homogeneity. Additionally, to use a retinal image for ARSS, it must cover appropriate parts of the retina and highlight important structures related to the pathology automatic screening. This aspect of retinal image quality is referred to as 'field-definition'.

Certain restrictions related to the field definition of the retinal image are stated by various health boards worldwide [16,17]. According to the recommendation by the Health Technology Board for Scotland [17], the retinal images used for automatic screening must be centred on the macular region and contain the optic disc (OD). If the image is to be used for diabetic retinopathy screening or grading, then the fovea (central region of the macula) should be at least two optic disc diameters (ODD) away from the edge of the image and the OD and the temporal arcades need to be completely visible. The New Zealand Ministry of Health has presented similar restrictions that should be followed while using a retinal image to screen diabetic retinopathy [16]. The retinal image quality assessment (RIQA) should thus be performed as a preliminary step in any automatic retinal image analysis (ARIA) techniques. In this article, we have developed a novel method to determine the retinal image quality and this proposed RIQA technique analyses retinal image quality and content. It follows a two-stage classification approach: exploiting the advantages of supervised classification and then exploiting the advantages of unsupervised classification. The deep learning-based first stage analyses the quality of the input image, while the unsupervised second stage analyses the content/field definition of the input image. Since the deep learning-based first stage classification follows a transfer learning approach, pre-trained deep convolutional neural networks are employed, which helps reduce the number of training samples. The content analysis ensures that sufficient information is available in the input retinal image to be classified

* Corresponding author.
E-mail addresses: renoh.chalakkal@auckland.ac.nz (R.J. Chalakkal), w.abdulla@auckland.ac.nz (W.H. Abdulla), sinuthulasi@gmail.com (S.S. Thulaseedharan).

https://doi.org/10.1016/j.compbiomed.2019.03.019
Received 22 October 2018; Received in revised form 20 March 2019; Accepted 21 March 2019

as a medically suitable retinal image (MSRI) that can then be used to triage and/or diagnose various pathologies, such as diabetic retinopathy.

The main contributions discussed in this article are as follows:

1. A new RIQA method is introduced based on image quality and field definition.
2. A combined two-level classification, based on deep learning and unsupervised classification, is introduced for analysing the quality in terms of sharpness, illumination, homogeneity, field definition and content of the input retinal image.
3. Transfer learning using pre-trained deep convolutional neural networks [18] is used in the first stage. This approach helps to reduce the number of retinal images needed for training. Transfer learning also leads to faster convergence and better classification performance.
4. New metrics based on the field definition of the retinal image that are required for diabetic retinopathy screening are introduced.

In this article we discuss the state-of-the-art IQA techniques used in retinal image analysis, the various databases used in this study, the proposed IQA method, the performance and comparison of the proposed method with other methods, and the conclusions and future directions of this work.

2. Recent advancements in retinal image quality assessment

The RIQA methods can be categorised into two classes—segmentation-based and histogram-based IQA methods. In Fleming et al. [1], the visibility of the vessels in the macular region is taken as a quality measure. The images are graded into four categories—excellent, good, fair or of inadequate clarity—based on the visibility of retinal vessels within one ODD around the macula. Also based on the field of view (FOV), there are three classes: excellent, good and inadequate. This classification is based on the positioning and visibility of the macula and OD in the image. An image is graded as inadequate quality if either the image clarity or the FOV is determined inadequate; otherwise, the image is classified as of adequate quality.

Giancardo et al. [2] discussed a method for estimating the image quality of a retinal image by using 33 features extracted from each image. Of these 33 features, 18 are extracted from the segmented retinal vessels. For vessel segmentation, they used a method proposed by Zana and Klein [19]. The vessel image is divided into 18 sections obtained by sampling the region of interest (ROI; fitted ellipse using Hough transform). The area of the vessels pixels in each window is determined and normalised. The other 15 features are extracted using histogram analysis by taking five bins per RGB colour channel. The classification is performed using support vector machine (SVM) and K-nearest neighbour (KNN)–based classifiers.

Karnowski and colleagues [3] have discussed a method based on setting two separate thresholds—quality threshold (T_Q) and optic nerve confidence (T_C). According to this method, the performance of an ARIA system may be improved by selecting input images satisfying the thresholds T_Q and T_C. The authors determined the quality threshold based on the retinal vessel density obtained using a method described by Giancardo et al. [2]. An SVM classifier is trained on images categorised into three classes: poor, fair and good quality. A hard threshold set empirically at 0.40 is used to identify images with insufficient quality to be graded into any of the other three classes. Tobin and colleagues [20] developed a previous method focusing on the OD, in which its location is identified using a feature-based likelihood ratio (FBLR) method. For each pixel in the segmented vessel image, four features are identified and a training set is formed to create a Gaussian model. The OD is located by computing the maximum *a posteriori* estimation (MAP) and the optic nerve head is detected by using a supervised classification approach and principal component analysis (PCA)-derived features. The T_C value was based on the Euclidean distance between the locations of the two OD obtained using the two methods and they compared the effect of varying T_Q between 0.4 and 1, while restricting the T_C to be above a certain value. A T_Q value of 0.6 resulted in the best screening performance [20].

A hybrid method based on the combination of the segmentation method and the global image information (in turn based on the histogram and intensity properties of the retinal image) has been discussed by Sevik et al. [4]. This article treated the RIQA problem in two steps. First, 177 features were extracted from the retinal images, including both the segmentation-based features; the inferior, superior, nasal and temporal (ISNT) quadrant features [21] obtained from the segmented vessel image and the global features; Zernike moments [22]; Haralick texture features (a gray-level co-occurrence matrix (GLCM) feature) [23]; and edge histogram and colour-related features. Sevik et al. [4] also developed a retinal image database (Diabetic Retinopathy Images Database (DRIMDB) [24]) exclusive to the IQA problem. Using these features, the performance of various machine learning–based classifiers can be compared. Second, the images that passed the quality screening are analysed for medically suitable retinal image (MSRI) criteria, which requires the OD and fovea to be within an acceptable distance from the retinal image centre-line.

Abdel-Hamid and colleagues [5] discussed a RIQA method based on multi-resolution analysis specific to diabetic retinopathy screening. In this method, they evaluated the RIQA based on five clarity and content features, namely sharpness, illumination, homogeneity, field definition and outliers. The wavelet transform was used to determine the sharpness and illumination features. To assess homogeneity, both the wavelet and specific retinal saturation channel were analysed. Field definition and outliers were assessed using the same sharpness, illumination features and colour information, respectively. Alternative RIQA methods are discussed in other studies [6–8,25].

After the success of deep learning–based methods in the field of image classification, there are a few deep learning–based methods proposed for the IQA problem [10–13]. The advantage of these methods is that there is no need for extracting or defining the features. The deep convolutional neural network automatically identifies and extracts the deep features from the input images. Additionally, the performances of these methods are superior to traditional machine learning–based IQA methods. One such deep learning–based IQA method specific to the retinal images was discussed by Yu and colleagues [9]. They used convolutional neural network (CNN)–based features, along with the unsupervised saliency map-based features. With the help of these fused features, the SVM-based classifier categorises the images into high-quality and poor-quality images. They reported an overall accuracy of 95.42% after testing on 2200 (1101 high-quality and 1099 low-quality) images randomly picked from the Kaggle database [26]. These images were classified manually by professionals to provide a ground-truth benchmark. As performance of an automatic system can significantly vary by proper selection of the training and testing datasets, the reported performance cannot be considered a generalised one. They have not tested their method on other public databases [9].

In Zago et al. [27], another deep learning–based method to assess retinal image quality was discussed. They used a pre-trained network (Inception-v3 [28]) to classify the retinal images based on the image quality completely. They utilised two databases–DRIMDB [24] and ELSA-Brasil [29]—to train and validate their method and have reported an area under the receiver operating characteristic (ROC) curve (AUC) of 99.98% and 98.56%, respectively.

Since the performance of most of the above reviewed methods are based on a specific set of parameters or features of different databases, a fair comparison cannot be achieved. Apart from the technique detailed by Zago et al. [27], other methods rely on specific datasets for their training and testing and are not generic. Our approach is different from the reviewed method [27], although it does use pre-trained deep convolutional neural networks (DCNN) for the first level of classification.

We propose a deep learning–based two-level classification approach for RIQA. We have experimented with six different pre-trained DCNN to implement three different RIQA approaches and assessed the performances of the methods after testing on images from various public databases.

3. Retinal IQA–related databases

There are many public retinal image databases available that may be used freely for research purposes. However, the number of appropriate public databases are limited for the specific RIQA problem. We have used seven public databases for the development and performance evaluation of the proposed approach.

3.1. DRIMDB database

The DRIMDB was created by Sevik et al. for developing and testing their method [4]. This database contains 216 images obtained from the Retina Department of Ophthalmology, Medical Faculty at Karadeniz Technical University in Turkey. The images were taken using a Canon CF-60UVi Fundus Camera using 60° FOV and were then stored in JPEG files at 570 × 760 pixels resolution. The images are classified into three classes: 125 good (MSRI), 69 bad (not MSRI) and 22 outlier (non-retinal) class images. The classification was conducted by an expert ophthalmologist [4].

3.2. DR1 and DR2 databases

The Diabetic Retinopathy databases (DR1–DR2) were developed by Pires et al. [7]. The DR1 dataset contains 5776 images, with an average resolution of 640 × 480 pixels taken using a TRC-50X mydriatic camera with a maximum resolution of one megapixel (MP) and a FOV of 45°. The images were collected by the ophthalmology department in the Federal University of Sao Paulo, Brazil. The images are classified into three classes: 1300 good quality (no blur and centred on macula), 1392 poor quality (contains blur) and 3084 periphery images (retinal images not centred on macula). The DR2 dataset consists of 920 12.2 MP images, with a resolution of 867 × 575. The images are also classified into three classes: 466 good quality (centred on macula), 194 poor quality (with blur centred on macula) and 260 periphery images (not centred on macula). The images were captured using a TRC-NW8 retinographer with a Nikon D90 camera.

3.3. HRF database

The high-resolution fundus image (HRF) database was developed by Kohler et al. [30] and contains 18 pairs of images. In each pair, one image is of good quality (no blur) and the other is of poor quality (decreased sharpness and presence of blur). The images were taken using a Canon CR-1 fundus camera with a FOV of 45°. The resolution of the images varies from 3888 × 2592 to 5184 × 3456 pixels.

3.4. MESSIDOR database

The methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology (MESSIDOR) database [31] consists of 1200 good-quality digital colour retinal images taken by a colour video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a FOV of 45°. The images have a resolution of 1440 × 960, 2240 × 1488 or 2304 × 1536 in TIFF compression format. Although this is the most comprehensive retinal image database used by researchers worldwide, we recently found that there were 13 pairs of duplicated images in this database (hence there are only 1187 images). There were also some grading inconsistencies in this database, which the database providers have acknowledged recently. Additionally, all of the images are graded as of good quality by the database providers; however, not all the images are of MSRI quality.

3.5. UoA-DR database

The University of Auckland diabetic retinopathy database (UoA-DR) [32] was developed through research collaboration between the University of Auckland and Al-Salama Eye Hospital, India, and other collaborating hospitals [33]. This database consists of 200 high-quality images captured using a Zeiss VISUCAM 500 fundus camera with a FOV of 45° and a resolution of 2124 × 2056 pixels in JPEG format. All of the images were de-identified and no linked patient-related information. The optic nerve head centre, optic nerve head, macula, fovea and the retinal vessels of all of the 200 images in this database were manually segmented by a specialist ophthalmologist who acted as the first observer and by an optometrist as the second observer. Additionally, all of the 200 images of this locally created database are graded according to the severity of the diabetic retinopathy based on the International Clinical Diabetic Retinopathy scale [34], by a specialist ophthalmologist from the Government Medical College, Thrissur, India [35].

The manually segmented features, such as retinal vessels, OD and fovea, for the 200 retinal images may be used to benchmark the performance of new ARIA methods in the future. This database can be downloaded for free with certain access rights mentioned in Ref. [36]. Similar to the MESSIDOR database, not all of the images marked as high quality are of MSRI quality.

3.6. Kaggle diabetic retinopathy database

The Kaggle Diabetic Retinopathy database was developed to test the performance of various algorithms developed as part of an online diabetic retinopathy (DR) detection competition [26]. The database consists of around 80,000 images that are divided into training and testing sets. We randomly selected 3000 images from this database to test the proposed algorithm's performance. The images had different resolutions and were captured by different fundus cameras under a variety of imaging conditions. Left and right retina images have been provided for each patient and the images are classified into five different classes—no DR, mild, moderate, severe and proliferative DR—according to the severity of DR. For the purpose of validating the proposed RIQA approach, we manually classified the randomly selected 3000 retinal images into two grades of quality: 1997 good (with no blur and centred on macula with OD visible) and 1003 poor quality (with blur and/or decreased sharpness, not centred on macula, OD or macula not visible, or outlier images), according to the conditions given in Ref. [16].

3.7. IDRiD database

The Indian Diabetic Retinopathy Image Dataset (IDRiD) [37] was developed recently as part of the 'Diabetic Retinopathy: Segmentation and Grading Challenge' workshop at the IEEE International Symposium on Biomedical Imaging (ISBI-2018). The database consists of 516 good-quality images that may be used to evaluate the performance of algorithms developed for automated detection and grading of diabetic retinopathy and diabetic macular oedema. Images were acquired using a Kowa VX-10 alpha digital fundus camera with a FOV of 50° and all are centred near to the macula. The images have a resolution of 4288 × 2848 pixels and are stored in JPEG file format. Not all images marked as high quality are of MSRI standards.

4. The proposed classification method

Our proposed method consists of two stages for classification: a deep learning–based first stage and an unsupervised second stage of classification, for analysing the quality and field definition of the retinal images, respectively. The block diagram of the proposed method is shown in Fig. 1. In the first level, a pre-trained DCNN is used to classify

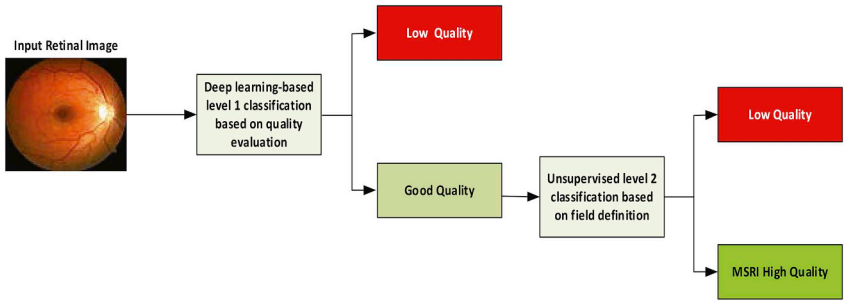

Fig. 1. Block diagram of the proposed RIQA method.

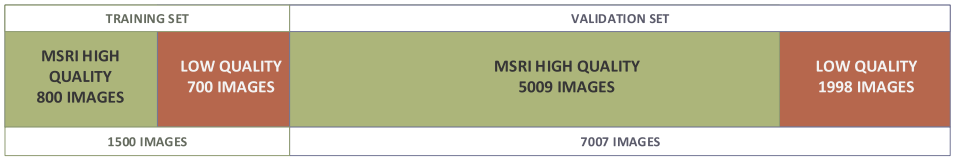

Fig. 2. Deep learning–based level 1 classification: Split-up of images between training and validation sets.

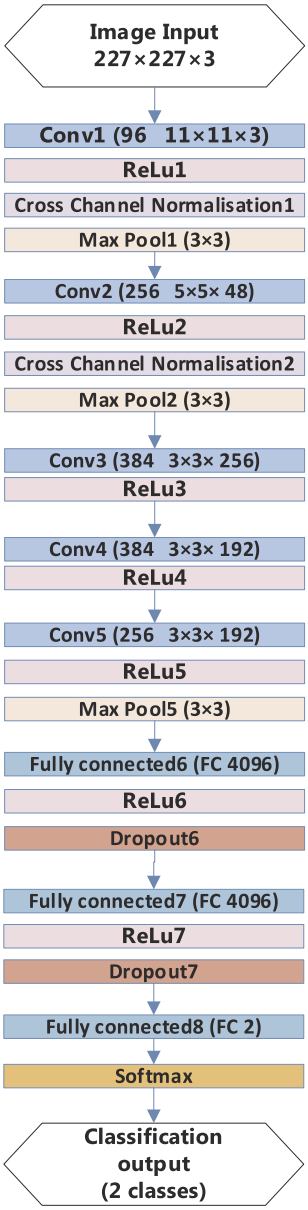

Fig. 3. AlexNet [39] architecture.

the images into two classes, low quality and good quality, using three different approaches. In the second stage, we applied an unsupervised classification technique to detect the important structures (e.g., OD, macula, fovea) in the images classified as good quality in stage 1. Using the modified metrics for identifying the MSRI, the good-quality images were further classified into high-quality MSRI and low-quality images. The high-quality MSRI can be used for ARSS, while the low-quality images cannot.

4.1. Deep learning–based stage-1 classification for quality evaluation

In stage 1 of classification, we utilised a deep learning–based approach to classify the retinal images into two classes: low-quality and good-quality images. We experimented with six different pre-trained CNNs for this task.

4.1.1. Training and validation sets

We used images from one particular database (DR1 [7]) to create the training set and the images from all the other databases were included in the validation set. This was to verify the generalised attributes of the proposed approach on different databases. From the 2692 images in the DR1 database, we selected 1500 images. These comprised of 800 good-quality images centred on the macula without any blur and 700 low-quality images with blur that were not centred on the macula and/or outlier images. Employing a selected number of images from a particular database in the training set was adapted to test the robustness of the system. This approach was also adopted by Agurto et al. [38], to verify whether the same trained classifier system could be used to classify images from different databases without the need for retraining. This scenario made our approach more likely to be endorsed for clinical practices.

The inclusion criterion of images in the training set was based on the visual inspection of images that could clearly identify the two separate classes (i.e., high-quality MSRI and low-quality images). The periphery images from the DR1 database (which are not centred on the macula) are not included in the training or validation sets, as they are not labelled according to the image quality measure (high-quality/low-quality). All other images, including the images obtained from other public databases, are included in the validation set and the outlier images (non-retinal images) from different databases are considered low-quality images for this study. The validation set consisted of 7007 images, of which 5009 images were of high-quality. The division between the training and validation sets is shown in Fig. 2. The validation set was not introduced to the system until the DCNN were trained and their hyper-parameters were tuned. The validation and training sets

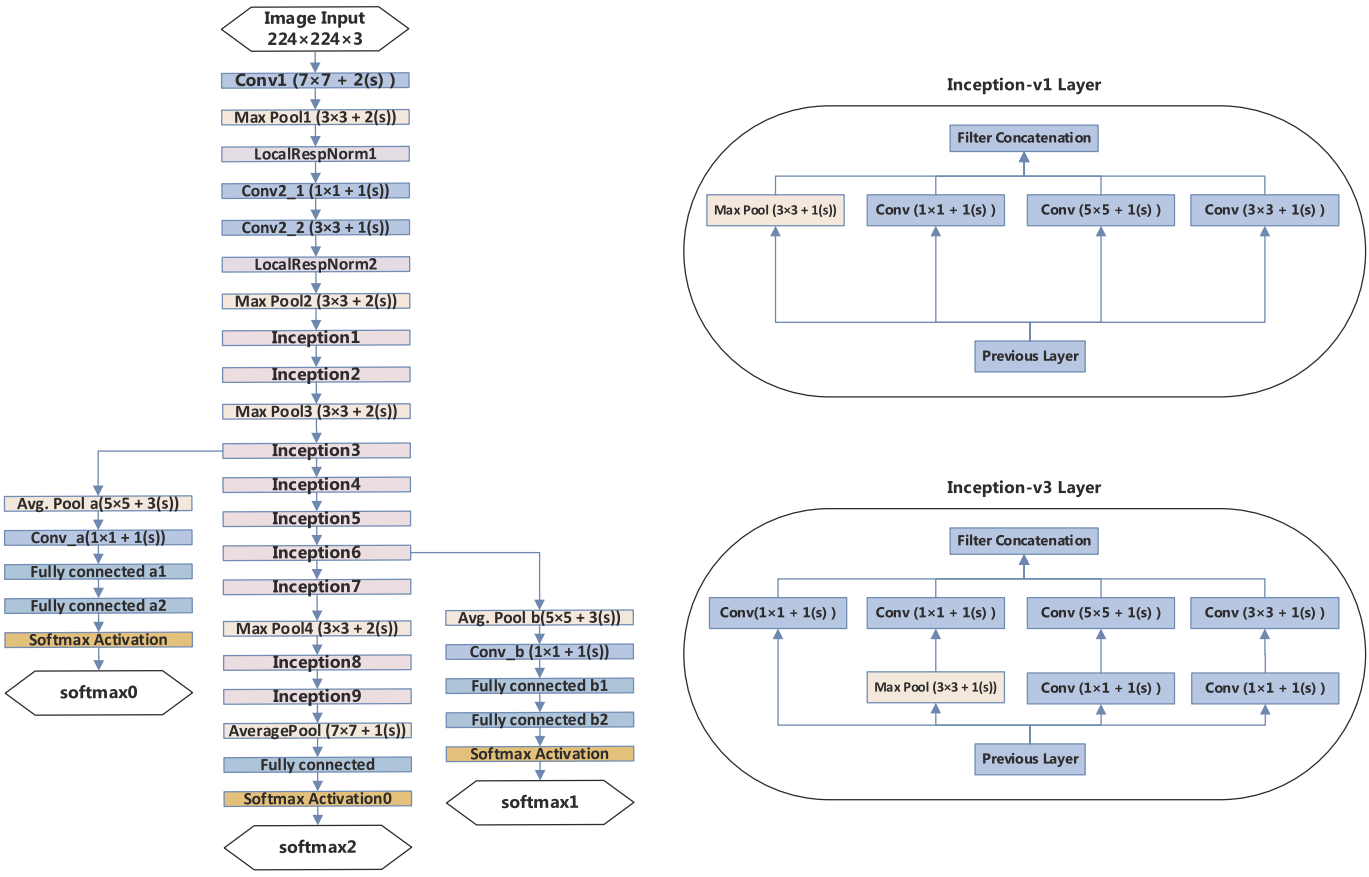

Fig. 4. GoogLeNet [40] architecture.

were mutually exclusive.

4.1.2. Architectures of the DCNN used

We experimented with six different pre-trained DCNNs—AlexNet [39], GoogLeNet [40], ResNet50 [41], ResNet101 [41], Inception-v3 [28] and SqueezeNet [42]—to create stage 1 classifiers. The DCNN architectures used in this article are shown in Fig. 3–Fig. 6.

AlexNet [39] is the first reported successful CNN architecture that won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. AlexNet has five convolution layers, three pooling layers and two fully-connected layers with approximately 60 million free parameters.

GoogLeNet (Inception-v1) [40] is a more complex DCNN architecture and it was the winner of the ILSVRC in 2014. It has two convolution layers, two pooling layers and nine inception layers; each of the inception layers consists of six convolution layers and one pooling layer. The inception layer used in GoogLeNet was Inception-v1 (Fig. 4).

ResNet [41] is a CNN architecture based on residual models. ResNet50 has 50 layers and ResNet101 has 101 layers (Fig. 5). ResNet architecture can solve the vanishing gradient problem that is common when increasing the number of layers in a CNN architecture. ResNet architecture was the winner of the ILSVRC in 2015.

Inception-v3 [28] is a modified version of its predecessor, GoogLeNet (Inception-v1). Inception-v3 has a newly defined and structured inception architecture compared to the original inception architecture used in GoogLeNet (Fig. 4). By using additional 1×1 convolutional layers before the large convolutions in the Inception-v3 architecture, the number of computations is reduced by approximately a factor of ten compared to the original inception architecture [28].

SqueezeNet [42] is a small CNN architecture that achieved AlexNet-level accuracy on ImageNet with fewer parameters. It was also 500 times smaller than AlexNet and introduced a new module, the Fire module. This module consisted of a 'squeeze' convolution layer and an 'expand' convolution layer (Fig. 6).

4.1.3. Training process based on transfer learning

Transfer learning [18] is common in the image processing domain. Pre-trained DCNN, which are trained on large datasets like ImageNet [43], are used in other classification and object detection problems. In these methods, the weights of the pre-trained DCNN are used as initial settings for the related classification or object detection task. Recently, CNN trained on natural images from the ImageNet are used effectively in medical image recognition applications using transfer learning [44–50]. A transfer learning–based approach [18] is followed to retrain a DCNN. The DCNN used in this step are pre-trained on 1000 common image classes available in the ImageNet dataset [43]. We fine-tuned the pre-trained DCNN by modifying the last few layers of the DCNN and reduced the number of output classes from 1000 to two (a binary classification). The initial weights for our specific task (RIQA level 1 classification) are copies from pre-trained DCNN layers previously trained on the ImageNet dataset with more than one million images.

There are three options to train DCNN such as AlexNet [39], GoogLeNet [40], ResNet50 [41], ResNet101 [41], Inception-v3 [28] and SqueezeNet [42]. The DCNN can be either 'trained from scratch', 'fine-tuned from a pre-trained model' or using 'off-the-shelf' CNN features of a DCNN pre-trained on ImageNet, and we only train the final layers to fit the new classification problem. We have used the 'fine-tuning' approach in the proposed level 1 deep learning–based classification stage. The main reason for selecting the 'fine-tuning' option is due to the limited number of images (approximately 8500 images) available for training/validation in the RIQA specific problem. Another reason is the larger distance or variation between the source (the type of images on

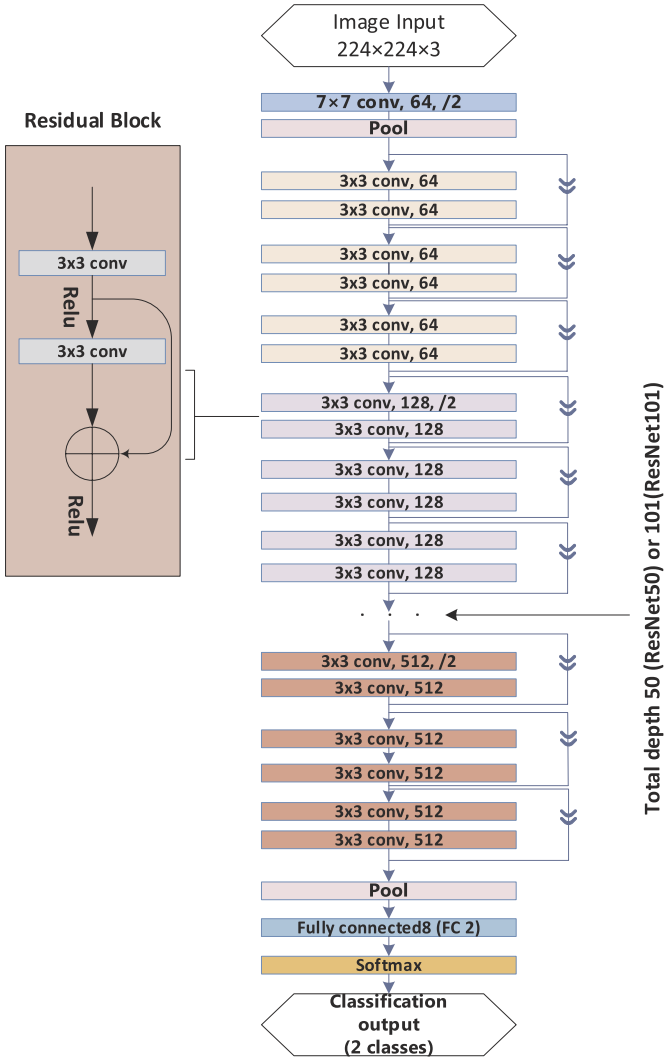

Fig. 5. ResNet [41] architecture.

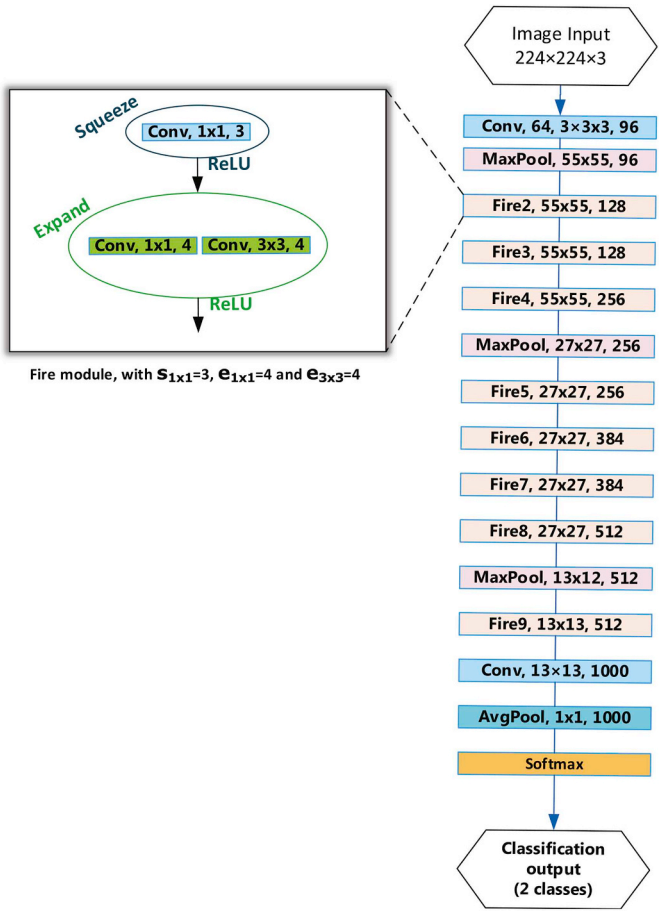

Fig. 6. SqueezeNet [42] architecture.

which the CNNs are pre-trained) and the target (fundus retinal images). Tajbakhsh et al. [44] demonstrated that, the transfer learning in these scenarios using 'fine-tuning' is the best option. Compared to the 'train from scratch' option, which requires a sufficiently large number of labelled retinal images to train millions of free parameters in the DCNN, the 'fine-tuning' option is optimal for this specific research problem. The option using the 'off the shelf' features and training only the last classification layer has its limitations, as this needs other machine-learning models, such as SVM and random forest classifiers, trained on the CNN extracted features to perform the final classification [51–53]. This further increases the system complexity and the results are inferior compared to the end-to-end fine-tuning approach in the image classification tasks, as in Refs. [44,50]. All models are trained for 100 epochs with early stopping criteria. Other hyper-parameters are solver: stochastic gradient descent; minibatchsize: 10; momentum: 0.9; and step learning rate schedule with a base learning rate of 0.0001 decreased by a factor of 10 when validation error reaches the minimum and remains constant. The initial learning rate η and momentum γ are tuned through an extensive set of trial and error experiments.

An $\eta = 0.0001$ provided a proper convergence for the specific task for all the CNN used in this study. Further, $\gamma = 0.9$ helped in increasing the speed of convergence. The grid search–based hyper-parameter tuning process is explained in Fig. 7 and it is clear that the combination of $\eta = 0.0001$ and $\gamma = 0.9$ provides a better validation accuracy for the classifier.

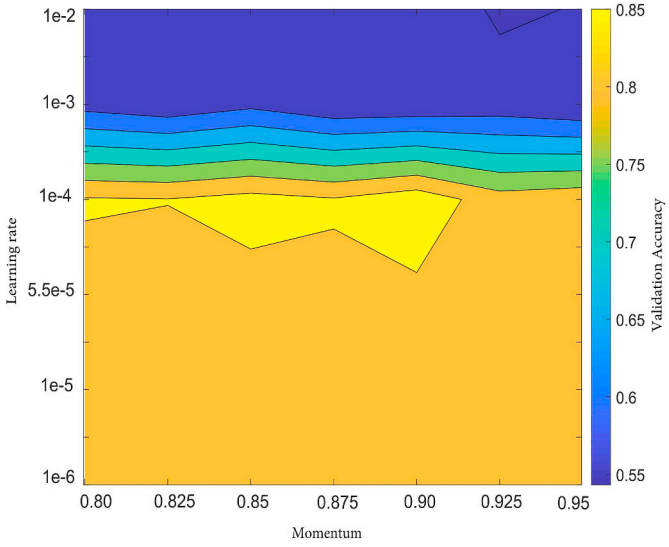

Fig. 7. Hyper-parameter tuning: Contour plot showing validation accuracy for various η and γ values.

4.1.4. Deep learning approaches

Three approaches were followed using each DCNN. In the first approach, the DCNN—AlexNet (AlexNet-T), GoogLeNet (GoogLeNet-T), ResNet50 (ResNet50-T), ResNet101 (ResNet101-T), Inception-v3 (Inceptionv3-T) and SqueezeNet (SqueezeNet-T)—were trained on all 1500 images in the training dataset and were then used to classify the entire set of images in the validation set into two classes (good quality and low quality).

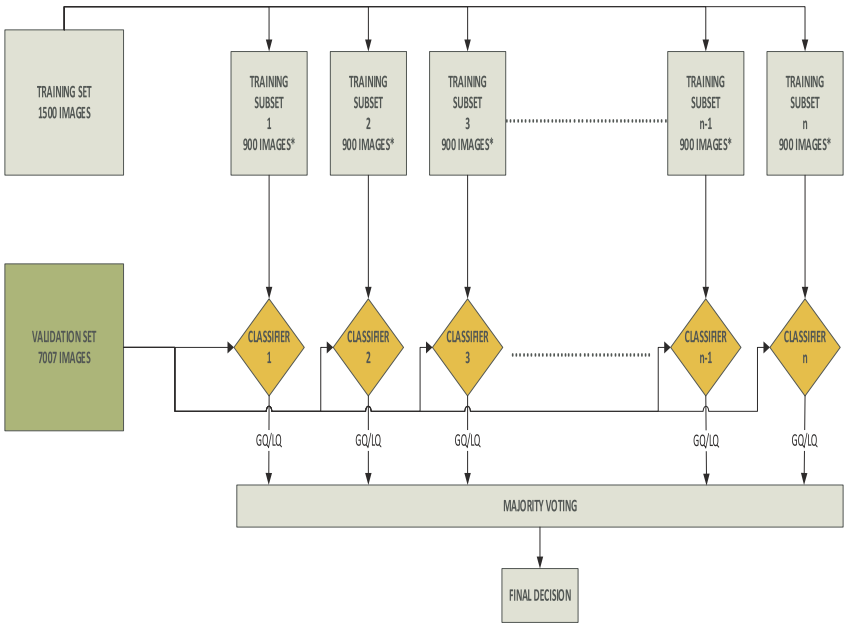

Fig. 8. Deep learning–based second approach (EB1).

In the second approach, 'n' subsets of training images from the full training dataset were created. In each of the training subsets, 60% (900 images) of the full dataset (1500 images) were selected randomly with replacement. Those 'n' subsets were used to independently train 'n' classifiers and AlexNet (AlexNet-EB1), GoogLeNet (GoogLeNet-EB1), ResNet50 (ResNet50-EB1), ResNet101 (ResNet101-EB1), Inception-v3 (Inceptionv3-EB1) and SqueezeNet (SqueezeNet-EB1)–based classifiers were built. After the training, the images from the validation set (7007 images) were classified by the 'n' classifiers independently to introduce 'n' decisions and then the majority voting procedure was used to decide the final classification result (Fig. 8).

In the third approach, we implemented an image subsampling method in which the training dataset was split into 'n' number of training subsets. The first subset was formed by randomly selecting 900 images (60%) from the full training dataset. To create the second training subset, we first classified the entire 1500 images of the original training set by the classifier trained on the first training subset. Then, the misclassified samples were selected first and then the remaining

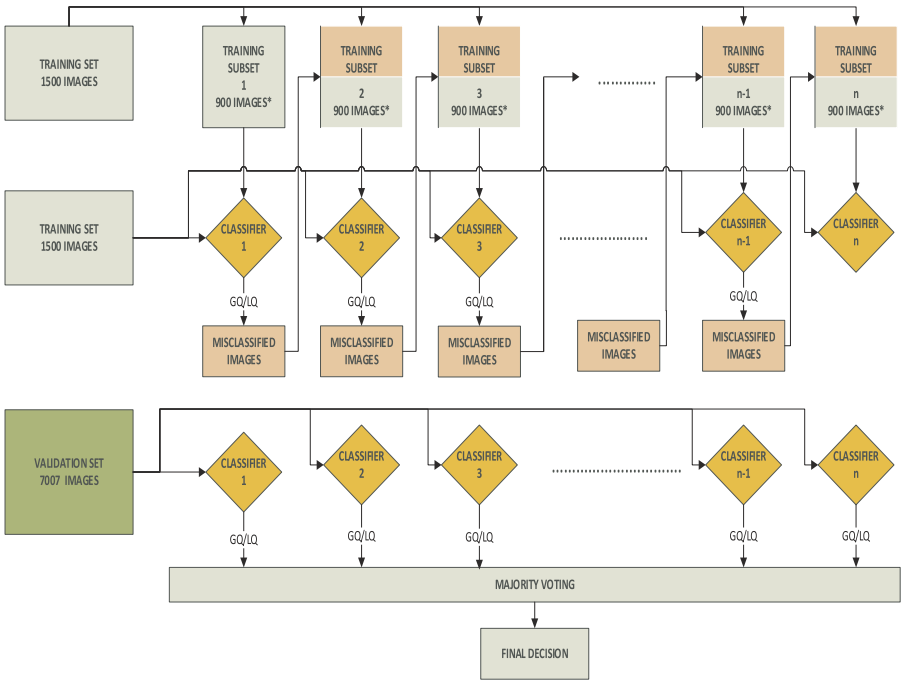

Fig. 9. Deep learning–based third approach (EB2).

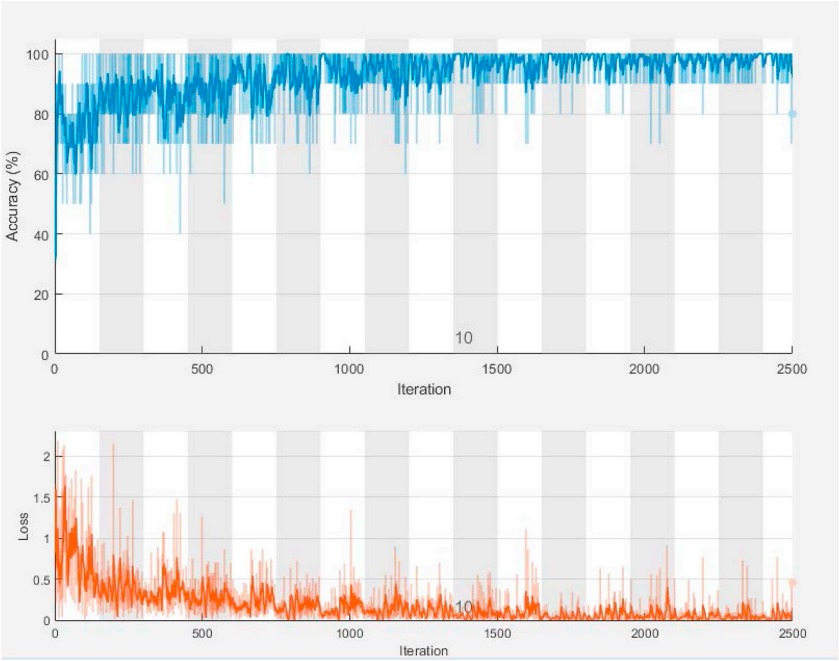

Fig. 10. Training accuracies and loss for GoogLeNet-EB1–based method using the first training subset (plotted for the first 17 epochs).

images randomly selected from the full training dataset to create the next subset. This procedure was repeated to create the subsequent subsets until the 'n' number of subsets were obtained. AlexNet (AlexNet-EB2), GoogLeNet (GoogLeNet-EB2), ResNet50 (ResNet50-EB2), ResNet101 (ResNet101-EB2), Inception-v3 (Inceptionv3-EB2) and SqueezeNet (SqueezeNet-EB2) were built using this approach. After creating 'n' different independent classifiers, the images from the validation set (7007 images) were classified and a majority voting was performed to reach the final decision (Fig. 9). For the second and third ensemble-based approaches, 'n' was 50.

The training accuracy and loss were monitored and the trained classifiers were used to classify the images in the validation set. Fig. 10 shows the training accuracy and loss for GoogLeNet-EB1.

4.2. Unsupervised level 2 classification for field definition

In the level 2 stage of classification, we used an unsupervised method as discussed in our recently published work [54] to automatically detect and segment the OD and macula from the retinal images (Fig. 11).

This method can accurately detect the OD and fovea in the retinal images. The method automatically detects the OD using histogram-based template matching and maximum sum of vessel information in the retinal image. OD detection is performed in two steps: coarse OD and fine-tuned OD detection. For detecting the fovea, each retinal image is split into three horizontal strips and the strip containing the detected OD is selected as the preferred region of interest; the contrast of this horizontal strip is enhanced using a series of image processing steps. The macula region is detected in this strip using various morphological operations and connected component analysis. For the detection of the OD, we have considered the vesselness information in a retinal image. The vessel segmentation is performed using a method we have previously discussed [55]. Most state-of-the-art methods (e.g. Refs. [1,2]), have considered the vessel visibility in the retinal image as

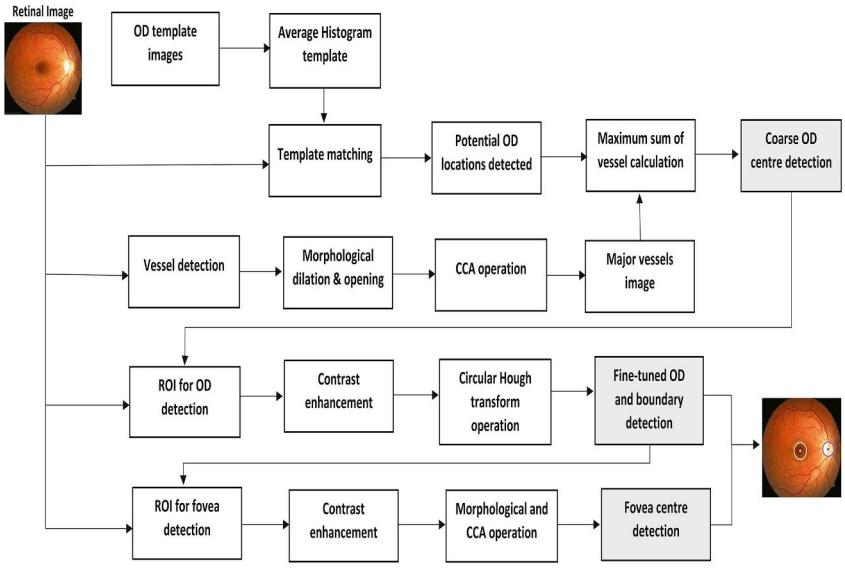

Fig. 11. Block diagram of the automatic detection and segmentation of the optic disc and macula [54].

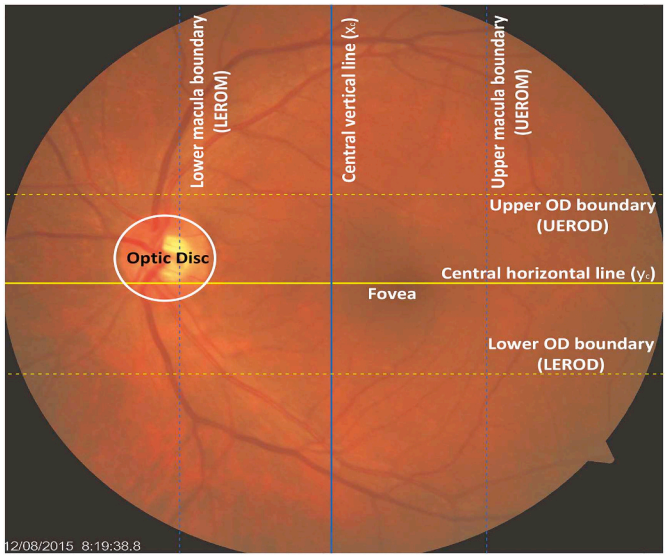

Fig. 12. Optimal boundaries of ERODs and EROMs for level two classification.

the most important criterion related to the retinal image quality. Since we have used retinal vessel segmentation as an important step in the OD centre detection, in the images with less vessel visibility, there is a higher chance of classifying these images as low quality.

The OD, macula and fovea are located in the images classified as good quality. The relative position of OD and fovea is then used to create an MSRI metric that is capable of classifying the good-quality images into MSRI high-quality images and low-quality images. The MSRI metric used in this study is a modified version of the method by Sevik et al. [4] and is defined in terms of an upper and lower bound for the expected region of OD (EROD). These bounds are defined as follows:

$$L_{EROD} = y_c + 1\ ODD \tag{1}$$

$$U_{EROD} = y_c - 1\ ODD \tag{2}$$

Where y_c is the retinal image centre-line along the horizontal x-axis and 1 ODD is the length of one OD diameter. In addition to this, the EROD is further divided into three vertical sections corresponding to the three expected regions of the macula (EROM), in which each section has a width of 3 ODD. These bounds are defined as follows:

$$L_{EROM} = x_c + 1.5\ ODD \tag{3}$$

$$U_{EROM} = x_c - 1.5\ ODD \tag{4}$$

where x_c is the retinal image centre-line along the vertical y-axis. Based on the relative location of the detected OD and fovea concerning the EROD and EROM, respectively, the MSRI metric is defined as:

$$MSRI_{OD} = |y_c - y_{OD}| \tag{5}$$

$$MSRI_{macula} = |x_c - x_{macula}| \tag{6}$$

where y_{OD} is the y component of the detected OD location and x_{macula} is the x component of the detected macula location.

Once the detected OD is within the defined L_{EROD} and U_{EROD}, and the detected fovea is within the defined L_{EROM} and U_{EROM} regions, the respective image is classified as a MSRI high-quality image. These upper and lower bounds are presented in Fig. 12. Ideal values for $MSRI_{OD}$ and $MSRI_{macula}$ for an MSRI high-quality image is near zero, which indicates that the detected OD is located near the horizontal centre-line y_c and the detected macula is near the vertical centre-line x_c.

5. Experimental results

The programming and computations were performed using a computer with Intel Core i7, a 3.60 GHz processor with 16 GB RAM and NVIDIA Quadro K620 GPU support. The proposed method was tested on images from seven different public databases. The performance of the proposed approach is reported based on the sensitivity (SE), specificity (SP), positive prediction value (PPV), negative prediction value (NPV), overall accuracy (Acc) and AUC.

The performance of the level 1 classifier was analysed using ROC (Fig. 13) and the performance of the proposed approach is reported in Table 1. The values for SE, Acc, PPV, NPV and AUC for various approaches are reported for a SP of 0.60. A low SP is selected because with the level 1 classifier, we are analysing the quality measure of the retinal image. The main goal of the level 1 classifier is thus to identify the high-quality MSRI images and, therefore, a high SE value is desirable.

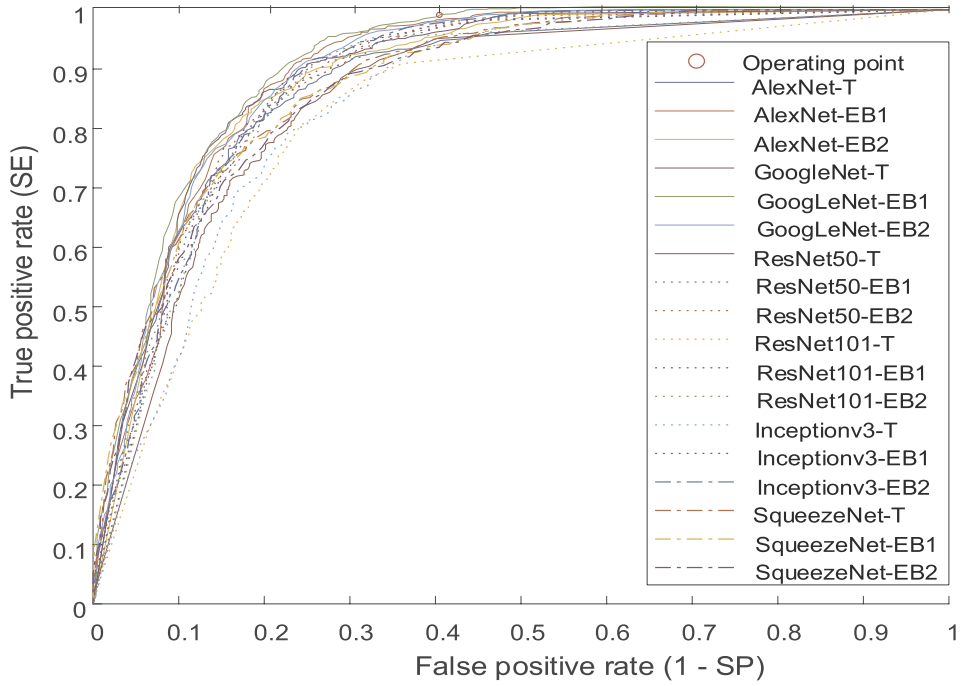

Fig. 13. ROC curves for various DCNN based classifiers.

Table 1
Performance of different deep learning–based approaches.

Method	SE	SP	PPV	NPV	Acc	AUC
AlexNet-T	0.9482	0.60	0.8106	0.7862	0.8106	0.8891
GoogLeNet-T	0.9831	0.60	0.8436	0.9410	0.8529	0.8971
ResNet50-T	0.9514	0.60	0.7888	0.8867	0.8145	0.8644
ResNet101-T	0.9091	0.60	0.7985	0.8170	0.8041	0.8415
Inception-v3-T	0.9408	0.60	0.7862	0.8651	0.8074	0.8495
SqueezeNet-T	0.9503	0.60	0.7893	0.8848	0.8145	0.8825
AlexNet-EB1	0.9535	0.60	0.7919	0.8922	0.8184	0.8891
GoogLeNet-EB1	**0.9956**	**0.60**	**0.8619**	**0.9819**	**0.8828**	**0.9144**
ResNet50-EB1	0.9619	0.60	0.7968	0.9111	0.8268	0.8787
ResNet101-EB1	0.9683	0.60	0.8000	0.9254	0.8326	0.8787
Inception-v3-EB1	0.9693	0.60	0.7912	0.9253	0.8248	0.8860
SqueezeNet-EB1	0.9397	0.60	0.7867	0.8633	0.8074	0.8799
AlexNet-EB2	0.9619	0.60	0.7913	0.9093	0.8216	0.8945
GoogLeNet-EB2	0.9789	0.60	0.7969	0.9481	0.8345	0.8981
ResNet50-EB2	0.9704	0.60	0.7976	0.9293	0.8313	0.8829
ResNet101-EB2	0.9524	0.60	0.8221	0.9002	0.8449	0.8701
Inception-v3-EB2	0.9789	0.60	0.7901	0.9467	0.8281	0.8864
SqueezeNet-EB2	0.9366	0.60	0.7848	0.8565	0.8041	0.8737

GoogLeNet-T and GoogLeNet-EB1–based methods had better performance metrics compared to the other approaches (Table 1). However, the GoogLeNet-EB1–based method with high SE values (0.995) and AUC (0.9144) was selected as the preferred level 1 classifier. The confusion matrix for GoogLeNet-EB1 is shown in Fig. 14, in which the total number of misclassified images are 821. The misclassified high-quality MSRI images—false negatives (FN)—are only 22, while the low-quality images wrongly classified as high-quality MSRI images—false positives (FP)—are 799. A higher number of FP results in a reduced specificity measure of 60%. The level 1 classifier was successful in identifying the retinal images with insufficient clarity and imaging artefacts. Additionally, the level 1 classifier identified the outlier images (non-retinal images) with 100% accuracy. The majority of the FP images had sufficient clarity to be classified as a high-quality retinal image; however, the image either lacked adequate field definition to include the necessary retinal structures or the retinal structures present were not within the specific bounds defined in Eqs. (1)–(4).

Level 2 classification handles this issue successfully. The images classified as good quality were analysed by an unsupervised level 2 classifier. On this level, specific structures inside the images and their relative locations were specified. Out of 5786 images classified as good quality by the level 1 classifier, only 762 images were classified as low quality by the level 2 classifier. The confusion matrix for the level 2 classifier is shown in Fig. 15.

Among those 762 images classified as low quality, 703 images were true negative (TN), while 59 images were FN (Fig. 16). A majority of these FN images were from the UoA-DR (28 images), MESSIDOR (8 images) and IDRiD (7 images) databases. Even though these images were labelled as high quality by the database providers, they are not labelled as MSRI high quality according to the guidelines given by various public sector health agencies like the Health Technology Board of Scotland [17] and the New Zealand Ministry of Health [16]. The OD and macula in these images are either out of the pre-defined EROD and EROM regions or they are not visible in the fundus image; this is why these images were classified as low quality by the proposed unsupervised level 2 classification stage. Out of the 96 FP images obtained after level 2 classification, most of these images contained OD and fovea within the $L_{EROD} - U_{EROD}$ and $L_{EROM} - U_{EROM}$ limits.

However, the reason for the misclassification is due to the low SP value of the level 1 classification stage. There is a trade-off between the SP value of the level 1 classifier and the number of FP images after level 2 classification. The number of FP after level 2 is only 1.9% (96 images)

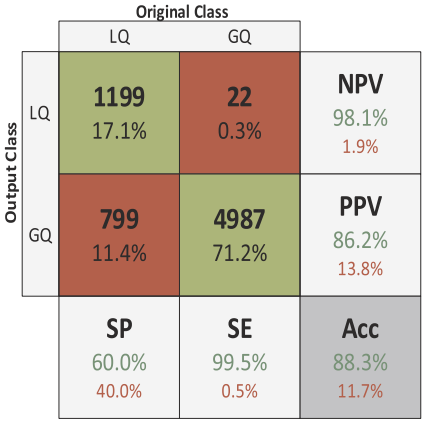

Fig. 14. Confusion matrix for GoogleNet-EB1–based level 1 classifier.

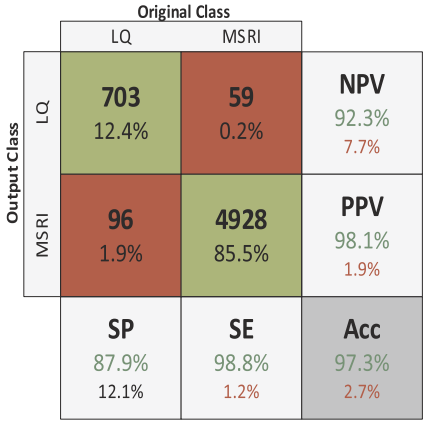

Fig. 15. Confusion matrix for level 2 classifier.

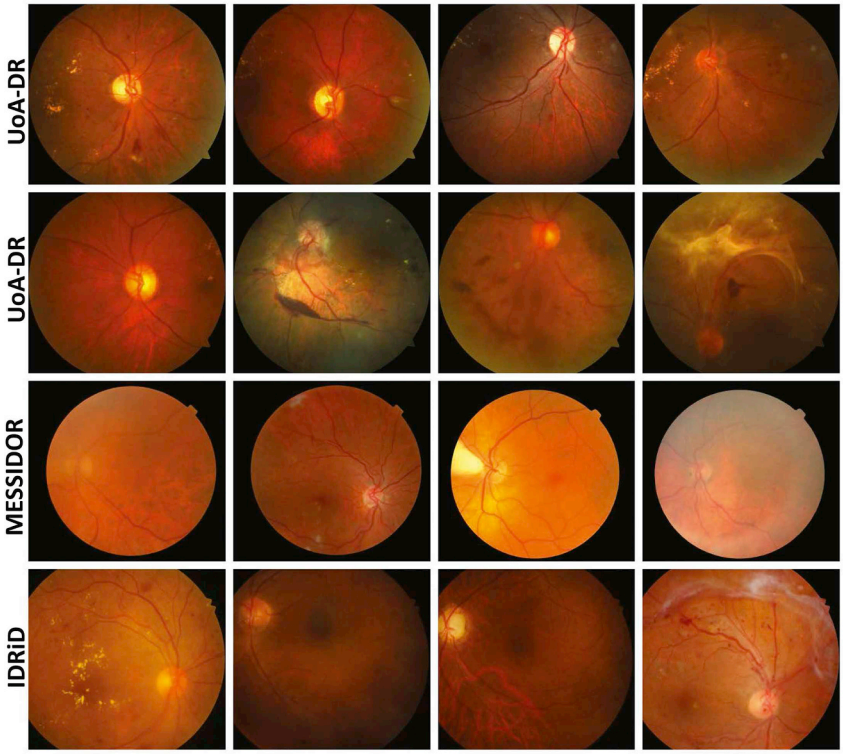

Fig. 16. False negative images: Misclassified high-quality images by level 2 classification stage. The first two rows show images from UoA-DR database; the third row shows images from MESSIDOR database and the fourth row shows images from IDRiD database.

of the total number of images (Fig. 15). The majority of FP images are from DR2 and Kaggle databases and the OD and fovea are visible within the defined EROD and EROM regions (Fig. 17). Accordingly, they are classified as high quality.

The images are classified either MSRI high-quality or low-quality images. Out of the 7007 images in the full testing set, 4928 out of 5009 (labelled MSRI high quality) are classified as MSRI high-quality images, and 1902 out of 1998 (labelled low quality) are classified as low quality (Table 2). The confusion matrix in Fig. 18 displays the overall classification performance of the proposed method.

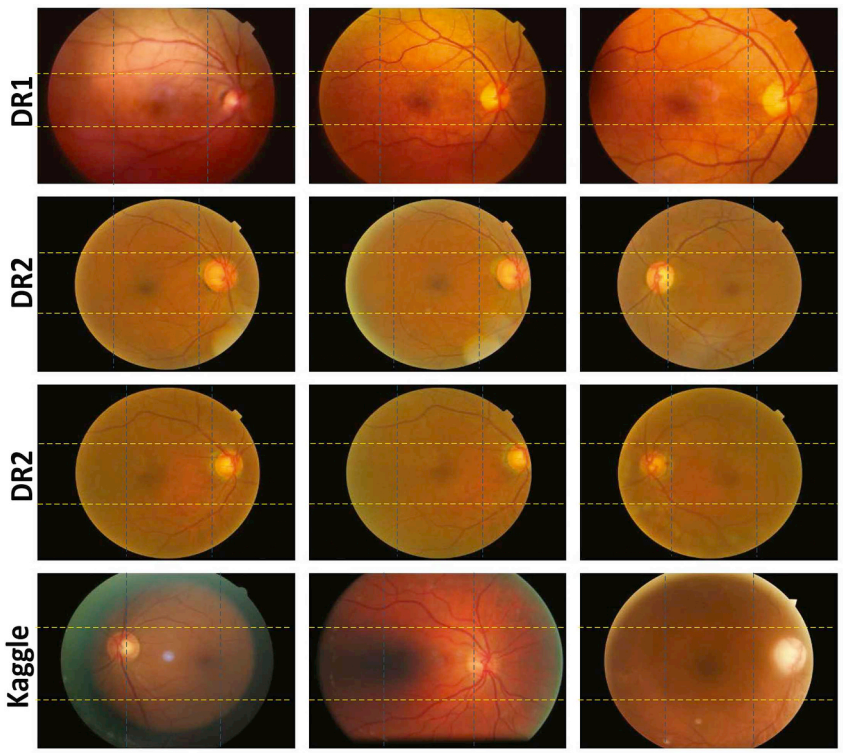

Fig. 17. False positive images: Misclassified low-quality images by level 2 classification stage. The first row shows images from DR1 database. The second and third rows show images from DR2 database, while the fourth row shows images from Kaggle database. The region between the two blue dotted lines represents EROM, and the region between the two yellow dotted lines represents EROD.

Table 2
Performance comparison with state-of-the-art methods.

Method	Year	Database used	No: Images	Performance
Fleming *et. al.* [1]	2006	Local	1039	SE: 99.1% SP: 89.4%
Hunter *et. al.* [25]	2011	Local	200	SE: 100% SP: 93% Acc: 94%
Yu *et. al.* [6]	2012	Local	1884	SE: 99% SP: 80% AUC: 0.981
Pires *et. al.* [7]	2012	DR2	854	AUC: 0.955
Sevik *et. al.* [4]	2014	DRIMDB, DIARETDB1 DRIVE	254	SE: 98%
Abdel-Hamid *et. al.* [5]	2016	DR1	2692	AUC: 0.927
Welikala *et. al.* [8]	2016	UK Biobank	800	SE: 95.33% SP: 91.13% AUC: 0.9828
FengLi *et. al.* [9]	2017	Kaggle	2200	SE: 96.63% SP: 93.10% Acc: 95.42% AUC: 0.9819
Gabriel *et. al.* [27]	2018	ELSA-Brasil, DRIMDB	1058	SE: 94.55% SP: 98% AUC: 0.9927% κ: 0.88
Proposed	2019	MESSIDOR, DR1-DR2 DRIMDB, HRF, UoA-DR Kaggle, IDRiD	7007	SE: 98.38% SP: 95.19% Acc: 97.47%

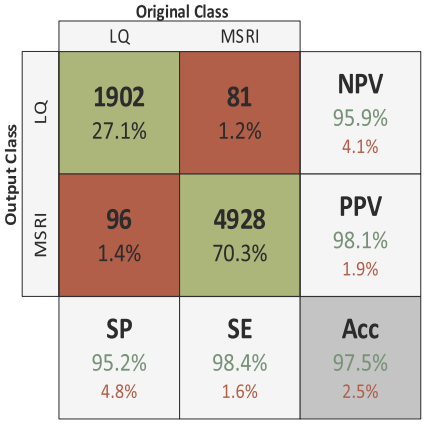

Fig. 18. Confusion matrix for the proposed method.

6. Discussion

Our proposed method has the second highest value for SE next to the method by Fleming et al. [1] (Table 2). However, the method reported in Ref. [1] has a lower SP value and has been tested with only 1039 images. Conversely, our proposed method has a higher SP value and has been tested with 7007 images. The recently reported RIQA method based on deep learning [27] reported a specificity of 98%, which is the maximum SP value among the reported methods. They used only 1058 images for the training and validation of their method. Using the (GoogLeNet-EB1 + unsupervised level 2) classification approach, we achieved a SE, SP and Acc of 97.6%, 97.8%, and 97.68%, respectively, on the 216 images from DRIMDB database (Table 3). Five images (two low quality and three high quality) were misclassified. We were unable to test our method on the ELSA-Brasil database, as the images were unavailable. Together, the proposed method achieved SE, SP and Acc values of over 95%, which outperforms the other methods on average. The proposed method performs well upon testing on retinal images from seven different public databases and it can be considered as the preferable generalised method.

Table 3
Performance of the proposed method on different retinal image databases.

Database	SE	SP	Acc
DR1	0.998	0.981	0.9883
DR2	0.991	0.912	0.968
HRF	0.888	1	0.944
UoA-DR	0.85	–	0.85
MESSIDOR	0.989	–	0.989
Kaggle	0.991	0.936	0.973
IDRiD	0.981	–	0.981
DRIMDB	0.976	0.978	0.977

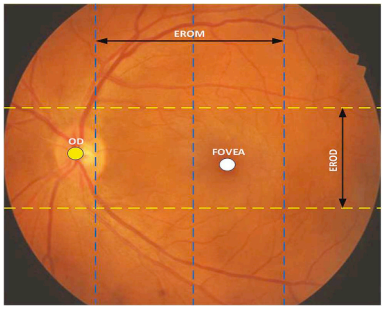

Fig. 19. An example image from IDRiD database which is classified as MSRI high-quality by the proposed method. The detected OD and fovea are marked by yellow and white circles respectively.

The performance of the proposed method was the best when tested on the DR1 database (Table 3). This is understandable, as the system is trained on images from the DR1 database—this is thus an intra-database performance. When we compare the inter-database performances, the proposed method achieves significant classification success. It achieves over 90% in SE, SP and Acc when tested on five public databases (excluding UoA-DR and HRF databases). After testing on UoA-DR and HRF databases, the SE achieved was less than 90%. In the UoA-DR database, even though the images are labelled as high quality by the

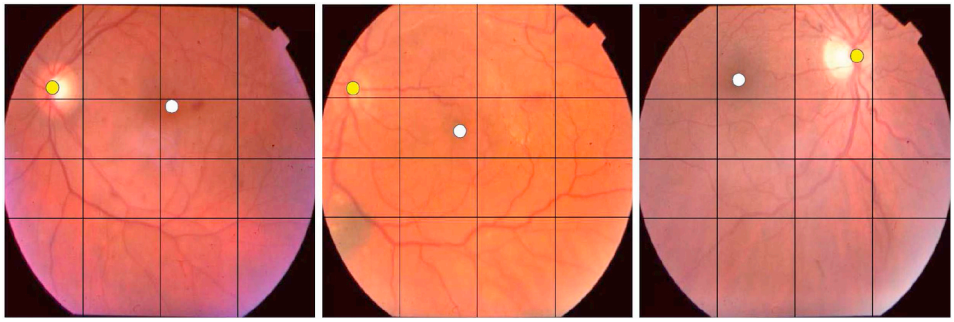

Fig. 20. Images classified as low quality by level 2 classification stage, while classified as good quality by level 1 classification stage. The yellow circle represents the detected OD centre and, the white circle represents the detected fovea. ERODs and EROMs are shown as black lines.

database provider, they do not conform with the MSRI standard. Hence, the reported low SE value is due to the level 2 classification stage, which analyses the content of such incorrectly labelled ground-truth images (location of OD and fovea). In the HRF database, two out of 18 high-quality images were misclassified. All low quality images in the HRF database were correctly classified; hence the SP is 1.

As mentioned in Ref. [27], most works [1,6,25] in the field of RIQA have used the same database images for both training and testing their algorithms. This makes their results biased and less generalised and, consequently, unreliable for practical clinical settings. Our proposed method has combined the advantages of both supervised and unsupervised classifications to obtain superior performance across various databases. The method also takes advantage of using pre-trained DCNN under transfer learning in medical imaging [27,44]. The transfer learning assists the proposed method to obtain superior classification performance over the 'training from scratch' methods.

Fig. 19 portrays an image classified as MSRI high-quality from the IDRiD database using the proposed method. Both the detected OD and fovea are inside the EROD and EROM regions, respectively. Conversely, Fig. 20 shows images that were originally classified as good quality by the deep learning–based level 1 classification, but were then classified as low quality by level 2 classification stage—either the detected OD was outside the EROD, the detected fovea was outside the EROM, or both.

7. Conclusions

This article presents a deep learning–based approach for RIQA evaluation. Level 1 classification is based on a deep learning approach, while level 2 classification is based on unsupervised learning. The level 1 classifier assesses the quality of the input retinal images, whereas the level 2 classifier analyses the field definition—the images are classified as either MSRI high-quality or low-quality images based on the two-level classification. Our proposed method with a SE, SP and Acc of 98.38%, 95.19%, and 97.47%, respectively, outperforms the other state-of-the-art methods in RIQA evaluation. The proposed method uses a deep learning approach and eliminates the need for classical image-specific learning features, which is the main concern of the traditional RIQA classifiers. The transfer learning approach using pre-trained DCNN also helps the system to be less dependent on the number of retinal images required in the training phase. Since we can use the pre-trained models (GoogLeNet, trained on millions of images) for the classification problem, we have the additional advantage of faster convergence and better classification performance. Additionally, our method is more generalised and robust and does not require training with different databases. This ensures that our proposed method is more suitable to adopt in clinical practice. The unsupervised level 2 classification stage helps to analyse the field definition of the input retinal images, which is significant in DR screening.

The proposed RIQA method can be used alongside other retinal image–based pathology detection algorithms. This method will help in screening MSRI images to build automatic screening systems for different retinal image–based pathologies, such as diabetic retinopathy, hypertension and glaucoma. The proposed system could also be used at the time of image capturing by the specialist ophthalmologist/optometrist or by any technician or photographer. The system may automatically guide the clinician in deciding whether image recapturing is required or not. This attribute can be employed in the field of telemedicine, in which the person capturing the retinal images might not be an imaging expert. The system can also be integrated with many ARSS systems as a pre-processing unit that could reject images with low quality for further processing. In low-quality images, many vital but subtle signs of DR and glaucoma are hidden due to poor contrast and blur; these images might be labelled as normal by the ARSS. Therefore, our proposed system, if used at the pre-processing stage, can reject such images and request a repeat of the imaging procedure of the same subject, thereby reducing the risk of misclassification.

Conflicts of interest

None declared.

Acknowledgement

This work was supported by Education New Zealand through the New Zealand International Doctoral Research Scholarship (NZIDRS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2019.03.019.

References

[1] A.D. Fleming, K.A. Goatman, P.F. Sharp, S. Philip, J.A. Olson, Automated assessment of diabetic retinal image quality based on clarity and field definition, Investig. Ophthalmol. Vis. Sci. 47 (3) (2006) 1120–1125, https://doi.org/10.1167/iovs.05-1155.

[2] L. Giancardo, M.D. Abramoff, E. Chaum, T.P. Karnowski, F. Meriaudeau, K.W. Tobin, Elliptical local vessel density: a fast and robust quality metric for retinal images, IEEE Eng. Med. Biol. Soc. Conf. (2008) 3534, https://doi.org/10.1109/IEMBS.2008.4649968.

[3] T.P. Karnowski, D. Aykac, L. Giancardo, Y. Li, T. Nichols, K.W. Tobin, E. Chaum, Automatic detection of retina disease: robustness to image quality and localization of anatomy structure, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 5959–5964, , https://doi.org/10.1109/IEMBS.2011.6091473.

[4] U. Sevik, C. Kose, T. Berber, H. Erdol, Identification of suitable fundus images using automated quality assessment methods, J. Biomed. Opt. 19 (4). doi:10.1117/1.JBO.

19.4.046006.
[5] L. Abdel-Hamid, A. El-Rafei, S. El-Ramly, G. Michelson, J. Hornegger, Retinal image quality assessment based on image clarity and content, J. Biomed. Opt. 21 (9). doi:10.1117/1.JBO.21.9.096007.
[6] H. Yu, C. Agurto, S. Barriga, S.C. Nemeth, P. Soliz, G. Zamora, Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening, 2012 IEEE Southwest Symposium on Image Analysis and Interpretation, 2012, pp. 125–128, , https://doi.org/10.1109/SSIAI.2012.6202469.
[7] R. Pires, H.F. Jelinek, J. Wainer, A. Rocha, Retinal image quality analysis for automatic diabetic retinopathy detection, 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, 2012, pp. 229–236, , https://doi.org/10.1109/SIBGRAPI.2012.39.
[8] R. Welikala, M. Fraz, P. Foster, P. Whincup, A. Rudnicka, C. Owen, D. Strachan, S. Barman, Automated retinal image quality assessment on the UK biobank dataset for epidemiological studies, Comput. Biol. Med. 71 (2016) 67–76 https://doi.org/10.1016/j.compbiomed.2016.01.027.
[9] F. Yu, J. Sun, A. Li, J. Cheng, C. Wan, J. Liu, Image quality classification for dr screening using deep learning, 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC), 2017, pp. 664–667, , https://doi.org/10.1109/EMBC.2017.8036912.
[10] B. Bare, K. Li, B. Yan, An accurate deep convolutional neural networks model for no-reference image quality assessment, 2017 IEEE International Conference on Multimedia and Expo, ICME), 2017, pp. 1356–1361, , https://doi.org/10.1109/ICME.2017.8019508.
[11] S. Jia, Y. Zhang, D. Agrafiotis, D. Bull, Blind high dynamic range image quality assessment using deep learning, 2017 IEEE International Conference on Image Processing, ICIP), 2017, pp. 765–769, , https://doi.org/10.1109/ICIP.2017.8296384.
[12] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. 27 (1) (2018) 206–219, https://doi.org/10.1109/TIP.2017.2760518.
[13] R. Yang, J. Su, W. Yu, No-reference image quality assessment based on deep learning method, 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference, ITOEC), 2017, pp. 476–479, , https://doi.org/10.1109/ITOEC.2017.8122340.
[14] K.W. Tobin, M.D. Abramoff, E. Chaum, L. Giancardo, V.P. Govindasamy, T.P. Karnowski, M.T.S. Tennant, S. Swainson, Using a patient image archive to diagnose retinopathy, 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, pp. 5441–5444, , https://doi.org/10.1109/IEMBS.2008.4650445.
[15] N. Patton, T.M. Aslam, T. MacGillivray, I.J. Deary, B. Dhillon, R.H. Eikelboom, K. Yogesan, I.J. Constable, Retinal image analysis: concepts, applications and potential, Prog. Retin. Eye Res. 25 (1) (2006) 99–127 https://doi.org/10.1016/j.preteyeres.2005.07.001.
[16] Ministry of Health, Diabetic retinal screening, grading, monitoring and referral guidance, Wellington: Ministry of Health, https://www.health.govt.nz/system/files/documents/publications/diabetic-retinal-screening-grading-monitoring-referral-guidance-mar16.pdf.
[17] K. Facey, E. Cummins, K. Macpherson, A. Morris, L. Reay, J. Slattery, Organisation of services for diabetic retinopathy screening, health technology assessment report 1, Glasgow, Scotland, UK: Health Technology Board for Scotland, http://www.healthcareimprovementscotland.org/previous_resources/hta_report/hta_1.aspx.
[18] J.P. Sinno, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, https://doi.org/10.1109/TKDE.2009.191.
[19] F. Zana, J. Klein, Segmentation of vessel- like patterns using mathematical morphology and curvature evaluation, IEEE Trans. Image Process. 10 (7) (2001) 1010–1019, https://doi.org/10.1109/83.931095.
[20] K.W. Tobin, E. Chaum, V.P. Govindasamy, T.P. Karnowski, Detection of anatomic structures in human retinal imagery, IEEE Trans. Med. Imaging 26 (12) (2007) 1729–1739, https://doi.org/10.1109/TMI.2007.902801.
[21] J. Nayak, A.U. Rajendra, P.S. Bhat, N. Shetty, T.-C. Lim, Automated diagnosis of glaucoma using digital fundus images, J. Med. Syst. 33 (5) (2009) 337, https://doi.org/10.1007/s10916-008-9195-z.
[22] H. Sun-Kyoo, K. Whoi-Yul, A novel approach to the fast computation of zernike moments, Pattern Recogn. 39 (11) (2006) 2065–2076, https://doi.org/10.1016/j.patcog.2006.03.004.
[23] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Sys. Man Cybernetics SMC- 3 (6) (1973) 610–621, https://doi.org/10.1109/TSMC.1973.4309314.
[24] Diabetic retinopathy image database. [link], http://isbb.ktu.edu.tr/multimedia/drimdb.
[25] A. Hunter, J.A. Lowell, M. Habib, B. Ryder, A. Basu, D. Steel, An automated retinal image quality grading algorithm, 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 5955–5958, , https://doi.org/10.1109/IEMBS.2011.6091472.
[26] Kaggle - diabetic retinopathy detection database. [link], https://www.kaggle.com/c/diabetic-retinopathy-detection/data.
[27] G.T. Zago, R.V. Andreo, B. Dorizzi, E.O.T. Salles, Retinal image quality assessment using deep learning, Comput. Biol. Med. 103 (2018) 64–70, https://doi.org/10.1016/j.compbiomed.2018.10.004.
[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, CoRR abs/1512.00567, arXiv:1512.00567 http://arxiv.org/abs/1512.00567.
[29] E.M.L. Aquino, S.M. Barreto, I.M. Bensenor, M.S. Carvalho, D. ra Chor, B.B. Duncan, P.A. Lotufo, J.G. Mill, M.D.C. Molina, E.L.A. Mota, V.M.A. Passos, M.I. Schmidt, M. Szklo, Brazilian longitudinal study of adult health (elsa-brasil): objectives and design, Am. J. Epidemiol. 175 (4) (2012) 315–324, https://doi.org/10.1093/aje/kwr294.
[30] T. Kohler, A. Budai, M.F. Kraus, J. Odstrilik, G. Michelson, J. Hornegger, Automatic no-reference quality assessment for retinal fundus images using vessel segmentation, Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, 2013, pp. 95–100, , https://doi.org/10.1109/CBMS.2013.6627771.
[31] E. Decencire, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, J.-C. Klein, Feedback on a publicly distributed image database: the messidor database, Image Anal. Stereol. 33 (3) (2014) 231–234, https://doi.org/10.5566/ias.1155.
[32] R.J. Chalakkal, W.H. Abdulla, S. Sinumol, Comparative analysis of university of auckland diabetic retinopathy database, Proceedings of the 9th International Conference on Signal Processing Systems, ICSPS, ACM, New York, NY, USA, 2017, pp. 235–239, , https://doi.org/10.1145/3163080.3163087 http://doi.acm.org/10.1145/3163080.3163087.
[33] Al-salama Eye hospital. [link], http://alsalamahospital.in/.
[34] American Academy of Ophthalmology, International clinical diabetic retinopathy disease severity scale, detailed table, http://www.icoph.org/resources/45/International-Clinical-Diabetic-Retinopathy-Disease-ΩSeverity-Scale-Detailed-Table-.html.
[35] Government medical College, Thrissur, India. [link], http://gmctsr.org.
[36] UoA-DR database info. [link], https://figshare.com/articles/UoA-DR_Database_Info/5985208.
[37] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian Diabetic Retinopathy Image Dataset (Idrid), IEEE Dataport, 2018, https://doi.org/10.21227/H25W98.
[38] C. Agurto, V. Murray, H. Yu, J. Wigdahl, M. Pattichis, S. Nemeth, E.S. Barriga, P. Soliz, A multiscale optimization approach to detect exudates in the macula, IEEE J. Biomed. Health Info. 18 (4) (2014) 1328–1336, https://doi.org/10.1109/JBHI.2013.2296399.
[39] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90, https://doi.org/10.1145/3065386.
[40] C. Szegedy, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2015, pp. 1–9, , https://doi.org/10.1109/CVPR.2015.7298594.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2016, pp. 770–778, , https://doi.org/10.1109/CVPR.2016.90.
[42] F.N. Iandola, M.W. Moskewicz, K. Ashraf, S. Han, W.J. Dally, K. Keutzer, Squeezenet: Alexnet-Level Accuracy with 50x Fewer Parameters and < 1MB Model Size, CoRR abs/1602.07360, http://arxiv.org/abs/1602.07360.
[43] J. Deng, W. Dong, R. Socher, L. Li, Imagenet, A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255, , https://doi.org/10.1109/CVPR.2009.5206848.
[44] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312, https://doi.org/10.1109/TMI.2016.2535302.
[45] J. Arevalo, F.A. Gonzlez, R. Ramos-Polln, J.L. Oliveira, M.A.G. Lopez, Representation learning for mammography mass lesion classification with convolutional neural networks, Comput. Methods Progr. Biomed. 127 (2016) 248–257, https://doi.org/10.1016/j.cmpb.2015.12.014.
[46] T. Schlegl, J. Ofner, G. Langs, Unsupervised Pre-training across Image Domains Improves Lung Tissue Classification, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8848, (2014), pp. 82–93 https://doi.org/10.1007/978-3-319-13972-2_8.
[47] I. Banerjee, A. Crawley, M. Bhethanabotla, H.E. Daldrup-Link, D.L. Rubin, Transfer learning on fused multiparametric mr images for classifying histopathological subtypes of rhabdomyosarcoma, Comput. Med. Imag. Graph. 65 (2018) 167–175, https://doi.org/10.1016/j.compmedimag.2017.05.002.
[48] H. Chougrad, H. Zouaki, O. Alheyane, Deep convolutional neural networks for breast cancer screening, Comput. Methods Progr. Biomed. 157 (2018) 19–30, https://doi.org/10.1016/j.cmpb.2018.01.011.
[49] S. Pang, Z. Yu, M.A. Orgun, A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images, Comput. Methods Progr. Biomed. 140 (2017) 283–293, https://doi.org/10.1016/j.cmpb.2016.12.019.
[50] H. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298, https://doi.org/10.1109/TMI.2016.2528162.
[51] B. van Ginneken, A.A.A. Setio, C. Jacobs, F. Ciompi, Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans, 2015 IEEE 12th International Symposium on Biomedical Imaging, ISBI), 2015, pp. 286–289, , https://doi.org/10.1109/ISBI.2015.7163869.
[52] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology detection using deep learning with non-medical training, IEEE 12th International Symposium on Biomedical Imaging (ISBI) 2015, 2015, pp. 294–297, , https://doi.org/10.1109/ISBI.2015.7163871.
[53] F. Ciompi, B. de Hoop, S.J. van Riel, K. Chung, E.T. Scholten, M. Oudkerk, P.A. de Jong, M. Prokop, B. van Ginneken, Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a

convolutional neural network out-of-the-box, Med. Image Anal. 26 (1) (2015) 195–202, https://doi.org/10.1016/j.media.2015.08.001.
[54] R.J. Chalakkal, W.H. Abdulla, S.S. Thulaseedharan, Automatic detection and segmentation of optic disc and fovea in retinal images, IET Image Process. 12 (11) (2018) 2100–2110, https://doi.org/10.1049/iet-ipr.2018.5666.
[55] R.J. Chalakkal, W. Abdulla, Automatic segmentation of retinal vasculature, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP), 2017, pp. 886–890, , https://doi.org/10.1109/ICASSP.2017.7952283.

Renoh Johnson Chalakkal received his Bachelor of Technology degree from University of Calicut, India with distinction in 2010 and Master of Technology degree in Signal Processing from Mahatma Gandhi University, India in 2012 with distinction. He worked as an Assistant Professor at Mahatma Gandhi University, India from 2012 to 2015 prior to joining the University of Auckland in 2015 to pursue his PhD degree in the ECE department. Currently he is a PhD student in the Electrical, Computer and Software Engineering Department at the University of Auckland, New Zealand. He is a recipient of the prestigious New Zealand International Doctoral Research Scholarship from the Education New Zealand for conducting his research at the University of Auckland. He is currently carrying out research on retinal image processing. His areas of interest include Image & Video processing, Biomedical Signal Processing, DSP Controller, Speech and Audio Processing. He is also a member of professional societies like IEEE and ISTE.

Waleed H. Abdulla holds a PhD degree from the University of Otago, New Zealand. He is currently an associate professor at the University of Auckland. He was Vice President - Member Relations and Development (APSIPA). Waleed has published more than 170 refereed publications (among them two best papers awards) with over 1000 citations, one patent, and two books. He is on the editorial boards of six journals. He has supervised over 25 postgraduate students. He has been the recipient of many awards and funded projects, such as JSPS, ETRI, and Tsinghua fellowships. He has received Excellent Teaching Awards for 2005 and 2012. Waleed is also a Senior Member of IEEE. His research interests include Signal Processing Analysis and Recognition; human biometrics; image processing; machine learning; and active noise control.

Sinumol S. received MBBS degree from Government T D Medical College, Kerala, India in 1994 and the MS (Master of Surgery) degree in Ophthalmology from the Regional Institute of Ophthalmology, Government Medical College, Trivandrum, India, in 1999. She passed DNB (Diplomate National Board) Ophthalmology in 2000. Currently she is working as Associate Professor in Ophthalmology at the Government Medical College Thrissur, India. She is a member of Kerala Ophthalmological Society and All India Ophthalmological Society. She has got a number of publications in reputed journals and conferences in the field of ophthalmology. She is also a member of Institutional Research Committee, Government Medical College, Thrissur, India from 2015. Her work is currently focused on research into Lens Epithelial cells.