Deep Learning for Image Quality Assessment of Fundus Images in Retinopathy of Prematurity

Aaron S. Coyner BS[1], Ryan Swan BS[1], James M. Brown PhD[3], Jayashree Kalpathy-Cramer PhD[3,4], Sang Jin Kim MD[2,5], J. Peter Campbell MD[2], Karyn E. Jonas MD[6], Susan Ostmo MS[2], R.V. Paul Chan MD[7], Michael F. Chiang MD, MA[1,2]

[1]Medical Informatics & Clinical Epidemiology, and [2]Ophthalmology
Oregon Health & Science University, Portland, OR, United States;
[3]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, MGH/Harvard Medical School, Charlestown, MA, United States;
[4]MGH & BWH Center for Clinical Data Science, Boston, MA
[5]Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea;
[6]Ophthalmology, University of Illinois at Chicago, Chicago, IL, United States;
[7]Ophthalmology, Illinois Eye and Ear Infirmary, Chicago, IL, United States

Abstract

Accurate image-based medical diagnosis relies upon adequate image quality and clarity. This has important implications for clinical diagnosis, and for emerging methods such as telemedicine and computer-based image analysis. In this study, we trained a convolutional neural network (CNN) to automatically assess the quality of retinal fundus images in a representative ophthalmic disease, retinopathy of prematurity (ROP). 6,043 wide-angle fundus images were collected from preterm infants during routine ROP screening examinations. Images were assessed by clinical experts for quality regarding ability to diagnose ROP accurately, and were labeled "acceptable" or "not acceptable." The CNN training, validation and test sets consisted of 2,770 images, 200 images, and 3,073 images, respectively. Test set accuracy was 89.1%, with area under the receiver operating curve equal to 0.964, and area under the precision-recall curve equal to 0.966. Taken together, our CNN shows promise as a useful prescreening method for telemedicine and computer-based image analysis applications. We feel this methodology is generalizable to all clinical domains involving image-based diagnosis.

Introduction

Several diseases and disorders can be diagnosed via image-based analysis.[1-5] This possibility has opened the doors for telemedicine and other computer-based image analysis applications. However, a significant issue surrounding these applications lies in image quality. This problem affects most imaging modalities, from those that use natural light to x-rays to MRIs.[6-11] Many studies have noted that, in telemedicine applications, poor image quality can negatively affect diagnostic accuracy and confidence.[12-18] Similarly, it has been found that this same effect is witnessed with computer-based image analysis techniques.[13, 19, 20] To begin addressing this concern, we focused our attention on the disease retinopathy of prematurity (ROP).

Owing to medical advances in the neonatal intensive care unit, the survival rate of both smaller and younger premature infants has dramatically increased.[21] However, this has not come without consequence, as both a birth weight of less than 1250 grams and a gestational age less than 31 weeks significantly increases the risk of a child developing ROP, a potentially blinding disorder.[21] Approximately 15 million premature infants weighing less than 1250 gram are born each year, and roughly 184,700 are afflicted with some form of ROP.[21, 22] In developed countries, most instances (~90%) of ROP are not treatment-requiring, whereas in developing countries the rate is much lower (~60%).[21-23] Even after treatment, some eyes show unfavorable structural or visual outcome.[21] Overall, blindness due to ROP in developed countries remains less than 10%, however ROP-induced blindness may account for up to 40% of blindness in developing regions.[21-23] Additionally, other vision-related complications may arise later in life, such as retinal detachment, myopia, strabismus, amblyopia, and glaucoma.[21]

ROP diagnosis is made by dilated ophthalmoscopic examination, or by review of fundus images.[1] Fundus imaging is of particular interest as it opens up possibilities for telemedicine and other computer-based image analysis techniques. However, the images are meaningless if their quality is not sufficient.[12, 14] For example, a poor quality retinal fundus photo acquired from a child in a developing country that was sent to a ROP expert in the United States may be useless if the expert cannot visualize the retinal vasculature or the retina itself. Similarly, we have found that this same issue involving image quality arises during computer-based image analysis of ROP images using machine learning and deep learning methods.[19, 20] For example, one such deep learning tool uses a retinal fundus photo as input into a convolutional neural network (CNN), which then provides the user with an estimated diagnosis.[24] This method performs very well overall, however it has not been validated on images of lower quality.[24] While we are specifically interested in solving this problem for ROP, this issue exists and has attempted to be addressed in many other ocular- and non-ocular-related diseases.[6-12, 15, 17, 18, 25-33]

To address this gap in knowledge, we have developed a CNN strictly for the purpose of assessing the quality of retinal images. This algorithm may be used as a pre-screening technique for our ROP diagnostic tool or could also be used for telemedicine applications, where it would benefit both patients and physicians alike.[12-16, 34] Data for this study were obtained through a multi-center, NIH-funded, Imaging and Informatics in ROP (i-ROP) study centered at Oregon Health & Science University (OHSU).

Methods

Institutional Review Board

This study was approved by the Institutional Review Board at the coordinating center (OHSU) and at each of 8 study centers (Columbia University, University of Illinois at Chicago, William Beaumont Hospital, Children's Hospital Los Angeles, Cedars-Sinai Medical Center, University of Miami, Weill Cornell Medical Center, Asociacion para Evitar la Ceguera en Mexico [APEC]). This study was conducted in accordance with the Declaration of Helsinki. Written informed consent for the study was obtained from parents of all infants enrolled.

Retinal Fundus Image Dataset

For this study, 6,043 wide-angle fundus images were collected from preterm infants during routine ROP screening examinations as part of the i-ROP study, a multicenter ROP cohort study. Images of Stage 4 and Stage 5 ROP (partial retinal detachment and total retinal detachment, respectively) were excluded. The breakdown of the training, validation, test, and rank sets for the CNN are provided in **Table 1**. Images were randomly selected for each dataset.

Table 1. Image sets and corresponding number of images.

Image Set	Number of Images
Training	2,770
Validation	200
Test	3,073
Rank	30
Total	6,043

Ground Truth Labeling

A group of six independent experts from various study centers (abovementioned) evaluated each image for ROP classification and image quality, labeled as "Acceptable for diagnosis of ROP," "Possibly acceptable for diagnosis of ROP," or "Not acceptable for diagnosis of ROP" (**Figure 1**). Because there were few images labeled "Not acceptable for diagnosis of ROP" (<500), images labeled "Possibly acceptable for diagnosis of ROP" and images labeled "Not acceptable for diagnosis of ROP" were binned together. For these analyses, images labeled "Acceptable for diagnosis of ROP" will be referred to as "acceptable quality" and the set of binned images will be referred to as "possibly acceptable quality"

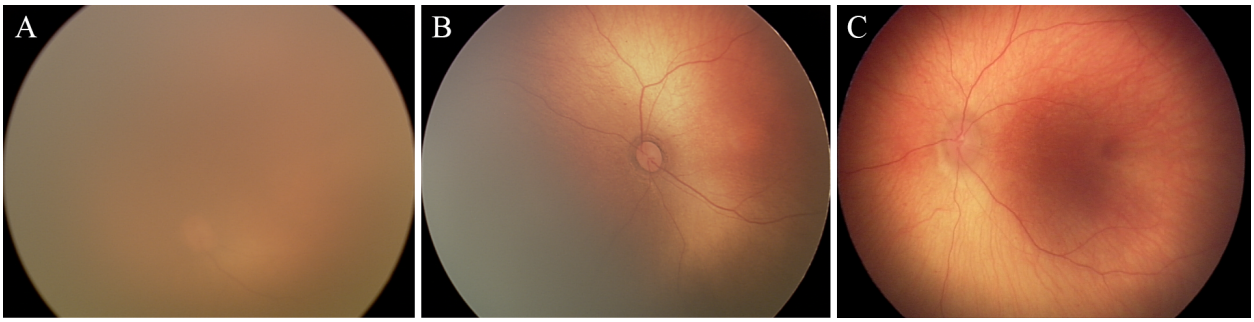

Figure 1. Representative images from the (A) Not acceptable for diagnosis of ROP, (B) Possibly acceptable for diagnosis of ROP, and (C) Acceptable for diagnosis of ROP image sets.

Pairwise Rankings

To further evaluate the CNN, we provided it a set of 30 images that were ranked from worst quality to best quality by six independent experts. The same six independent experts were also asked to rank a smaller set of 30 images from worst quality (1) to best quality (30) for diagnosis of ROP. To do this, we implemented a web-based interface that presented each expert with two separate images from the set of 30 along with the prompt "select the higher quality image for the diagnosis of plus disease." The output of the CNN for this test was the raw score for each image (i.e. a score between 0 and 1). This allowed for us to evaluate the ability of the CNN to rank images and compare it to individual experts as well as their consensus ranking. After multiple comparisons, Elo rating values for each image were determined for each individual expert. From these, a consensus rank of the images was formed.

Model Architecture and Training

The model was built and trained with the Keras package in Python using the TensorFlow backend. The convolutional model of the network was built using the pretrained VGG19 CNN and weights were initialized using the values after training on the ImageNet database.[35] Five dense, fully connected layers were built on top of the convolutional layers. Layer sizes, from convolutional to output layers, were as follows: 2048, 2048, 1024, 512, 1. Each layer used a ReLu activation, except for the last layer which used a sigmoid activation. There are 26 layers in total. The network was trained using the parameters presented in **Table 2**.

Table 2. Parameters used to train the CNN.

Parameter	Value
Optimizer	Stochastic Gradient Descent
Learning Rate	0.0001
Momentum	0.9
Decay	0.0
Nesterov	False
Loss	Binary Cross Entropy
Training Epochs	100
Input Image Size	256 x 256 x 3

During training, the validation set accuracy and loss were monitored to prevent overfitting of the model.[35, 36] After every training epoch, the validation set was used to assess what the CNN had learned. Each time the validation set accuracy increased or the result of the loss function decreased, a checkpoint saved the model architecture and corresponding weight matrix. The model with the highest validation set accuracy and lowest validation set loss was selected for use on the test set.

Data Analysis

Using the CNN, test set predictions for each image were determined. Briefly, batches of images were fed through the CNN, and a score between 0 and 1 was determined for each image. A score less than 0.5 placed the image into the "possibly acceptable quality" category, and a score greater than or equal to 0.5 placed the image into the "acceptable

quality" category. The overall accuracy of the CNN was evaluated, as was the area under the receiver operating characteristics curve (AUROC), and the area under the precision-recall curve (AUPR).

Additionally, the rank set of images was assessed using the CNN. As mentioned above, the output of the CNN for any given image is a score between 0 and 1. These values were used to rank the set of 30 images from worst (1) to best (30). Spearman rank correlations were used to assess the similarity of the CNN to rank the dataset of 30 images to the performance to human graders to rank the 30 images in terms of image quality. Additionally, a consensus rank of the images was determined for all six human graders, and the CNN was compared to that as well. A correlation matrix and scatter plot with a line of best fit were used to describe the findings.

Results

Classification Performance

The final CNN model was able to accurately classify 91% of images in the validation set as acceptable versus not acceptable, with loss equal to ~0.258. When this model was evaluated on the test set, accuracy was ~89.1% (**Table 3**). The AUROC was equal to 0.96 (**Table 3** and **Figure 2**) and the AUPR was equal to 0.97 (**Table 3** and **Figure 2**) demonstrating nearly perfect discrimination between acceptable and less than acceptable images. **Table 4** demonstrates the confusion matrix of the performance of the CNN compared to the ground truth. The CNN has a similar false positive rate as compared to its false negative rate (**Table 4**). Depending upon the application for which the CNN were to be implemented, the 0.5 CNN output threshold for binning images into the "possibly acceptable quality" image group or the "acceptable quality" image group could be adjusted up or down to decrease false positives or false negatives, respectively.

Table 3. Final evaluation of the CNN on the test dataset.

Statistic	Value
Accuracy	89.1%
Area Under the Receiver Operating Characteristic Curve	0.96
Area Under the Precision-Recall Curve	0.97

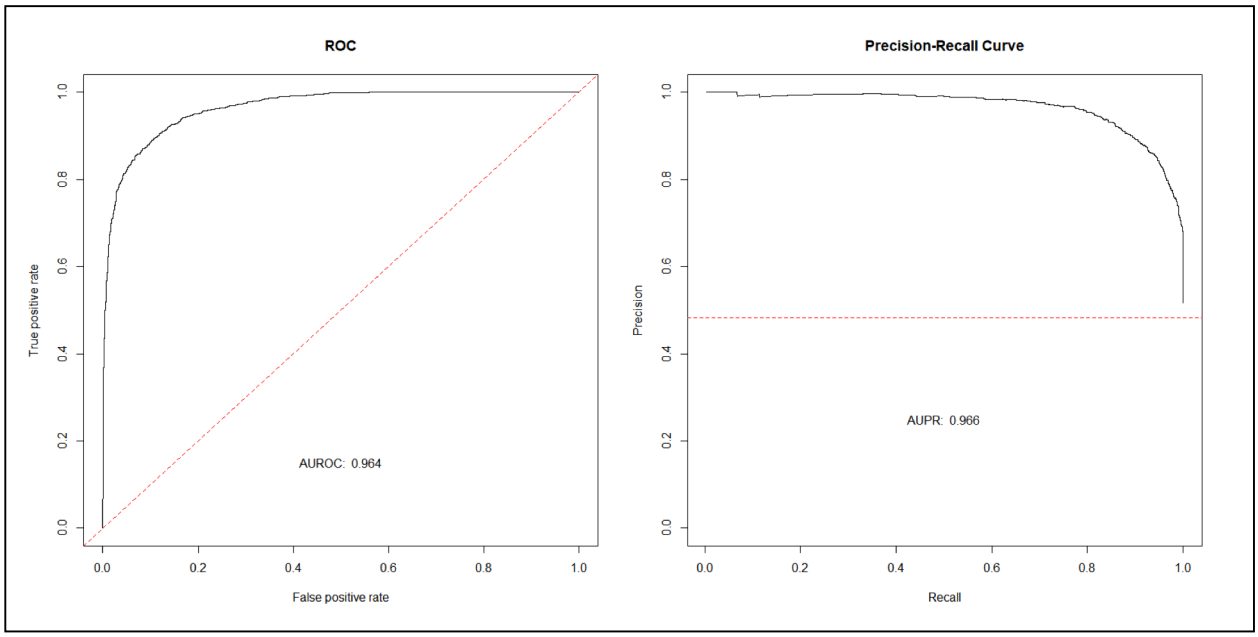

Figure 2. Receiver operating characteristics (ROC) curve for model evaluation on the test set data displayed in the left panel. Area under the ROC curve (AUROC) is ~0.96. Precision-recall curve for model evaluation on the test set

data is shown in the right panel. Area under the precision-recall curve (AUPR) is ~0.97. Red dashed lines indicate expected results of a naïve algorithm.

Table 4. Confusion matrix of the results of the test set data after evaluation by CNN.

		TRUTH LABELS	
		Not Acceptable	**Acceptable**
PREDICTED LABELS	**Not Acceptable**	1340	189
	Acceptable	145	1399

Pairwise Rankings

Evaluation of image quality rankings from the model versus the consensus ranking developed by the expert graders are described below (**Table 5** and **Figure 4**). The Spearman correlation coefficients between experts ranged from 0.86 to 0.98 demonstrating very high correlation on agreement of relative image quality. These results indicate that our model not only has high inter-group discrimination, but intra-group discrimination. That is to say, that given two images from the same image set ("acceptable quality" or "possibly acceptable quality"), our model can discriminate which of the two images is of higher quality. This insinuates that our model has not only learned the difference between an "acceptable quality" image and a "possibly acceptable quality" image, but that it has learned what features make any retinal fundus image of higher quality than another, which may be the focus of future work.

Table 5. Correlation matrix describing Spearman's correlation values between the CNN image ranking, each expert grader's image ranking, and the expert graders' consensus ranking.

	CNN	**Consensus**	**Expert 1**	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**
CNN	1.00	0.89	0.98	0.94	0.94	0.97	0.96	0.96
Consensus	0.89	1.000	0.86	0.89	0.84	0.92	0.91	0.89
Expert 1	0.86	0.98	1.000	0.890	0.89	0.94	0.93	0.93
Expert 2	0.89	0.94	0.90	1.000	0.94	0.96	0.95	0.92
Expert 3	0.84	0.94	0.89	0.94	1.000	0.91	0.92	0.91
Expert 4	0.92	0.97	0.94	0.96	0.91	1.000	0.97	0.94
Expert 5	0.91	0.96	0.93	0.95	0.92	0.97	1.000	0.97
Expert 6	0.89	0.96	0.93	0.92	0.91	0.94	0.97	1.00

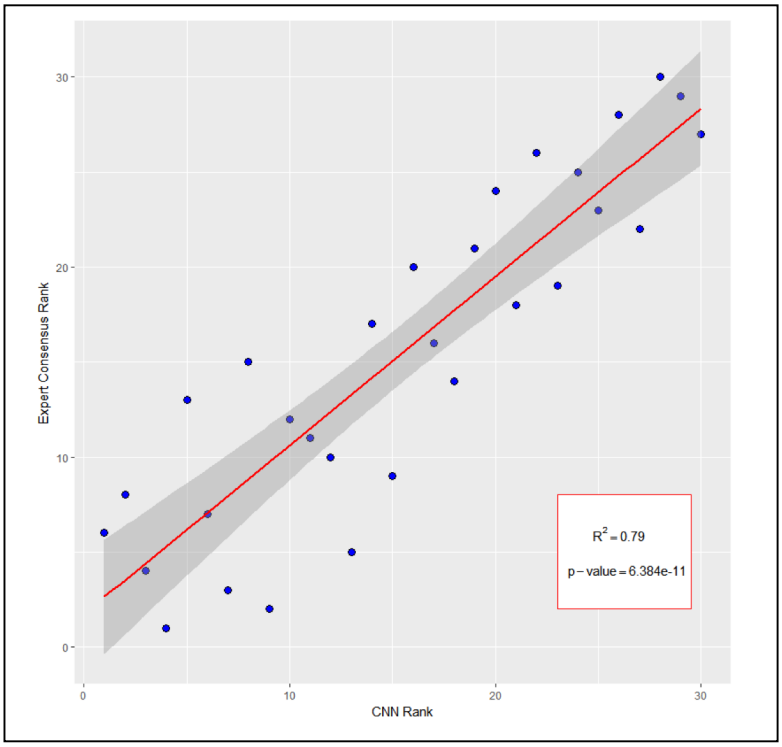

Figure 4. Scatterplot of experts' consensus rank versus CNN rank. Each point represents an image and its corresponding expert consensus rank (y-axis) and CNN rank (x-axis). Red line indicates best-fit line. Confidence interval is represented by dark gray band. Summary statistics are displayed within the plot.

Discussion

This study aimed to assess the quality of retinal fundus images for screening ROP using a CNN. There are two key findings: (1) Our CNN can reliably discriminate between images of acceptable quality from images of possibly acceptable (or worse) quality and (2) our CNN is highly generalizable to image quality detection in ROP, as it did not overfit to the data, since validation set accuracy and test set accuracy were similar (within 2%, **Table 3**). Sensitivity, specificity, and positive predictive values were also increased, and this was demonstrated by the area under the receiver operating characteristics curve (~0.964) and the area under the precision-recall curve (~0.966, **Table 3** and **Figure 2**). Taken together, these data suggest that our CNN can correctly identify what our experts consider to be an "acceptable quality" image from a "possibly acceptable quality" image. And, depending upon the application, the threshold at which the CNN labels images could be adjusted.

However, an even more compelling finding is the ability of the CNN to rank images based on their quality. The Spearman rank correlation of the CNN versus each individual expert displayed a high level of correlation, ranging from 0.84 to 0.92 (**Table 5** and **Figure 4**). The CNN versus the consensus ranking of the experts had a Spearman rank correlation of 0.89 (**Table 5** and **Figure 4**). These findings suggest that the CNN has not memorized what separates an "acceptable quality" image from a "possibly acceptable quality" image, but that it has learned what image quality is in general, as applied to retinal fundus photos. This further proves the point that the threshold at which the CNN labels images could be adjusted in an application-dependent manner without serious side-effects.

For instance, in a telemedicine application where physicians were manually reviewing images, the model would likely remain unchanged, since it was trained using the opinions of six experts in the field of ROP. However, implementation of this model as a prescreening tool in our ROP diagnostic tool may warrant some modifications. For example, it may be that the ROP diagnostic tool can still provide a reliable ROP diagnosis on some images in the "possibly acceptable quality" group. Assuming that the ROP diagnostic tool grades and ranks images (in terms of ROP severity) similarly to clinicians and given that the CNN can rank images with high correlation to experts in the field, the threshold at which images are binned into different categories could simply be adjusted until images were correctly binned based on the ability of the ROP diagnostic tool to accurately provide a diagnosis.

It should be noted that numerous groups have attempted to train CNNs or create other algorithms to learn the features of acceptable quality images from possibly acceptable quality images, but they all have certain limitations.[25, 27-30, 33] For example, Saha et al. built a CNN on top of AlexNet for assessing the quality of diabetic retinopathy images.[33, 36] Their CNN performed with an accuracy of 100% on a set of 3,572 images.[33] However, their image set only included images where all graders agreed on the quality of the images (i.e. images without consensus were excluded from the test set), which unfortunately left the dataset with a very bimodal distribution that was not an accurate representation of real-world images.[33] Consequently, it is likely that their CNN would not generalize well in practice.

Other groups have implemented linear algorithms for image quality assessment of retinal fundus photos, which have performed well.[11, 18, 25-30] However, it should be noted that all training and test data sets were relatively small in comparison to the dataset we used to train, validate, and test our CNN. We believe that, because our CNN was trained on nearly 3,000 images and tested on more than 3,000 images, it has proven that it generalizes well to most cases of ROP.

One notable limitation of our model is that only posterior pole images were used for training and testing. In practice, nasal, temporal, superior, and inferior images may be used in addition to posterior pole images for diagnosis of ROP.[1] Future training of this model may include those images to better suit it for clinical application. Another limitation of this CNN is its lack of ability to distinguish a retinal fundus photo from any other photo (e.g. a dog, a house, a car, etc.). Because this is to be used as a prescreening device prior to review by our ROP diagnostic tool (which only knows and can make accurate calls on retinal fundus photos), we plan to train a CNN to distinguish between images that are retinal fundus images and images that are not.

Our last future direction is to test this model on another set of retinal fundus images, however not on those acquired from premature infants. This model has shown a great ability to classify quality of retinal fundus images acquired during routine ROP screenings, but we would like to assess its ability to determine quality of images acquired from other patients (e.g. those undergoing screening for diabetic retinopathy or other retinal degenerative diseases). If the CNN performs well in those scenarios, the applications of this algorithm could range from a simple prescreening method for other programs to full implementation in a retinal fundus camera. For instance, an imaging technician could capture an image of a retina and instantly be alerted as to whether or not the image was of acceptable quality for diagnosis of disease.

Conclusion

We have successfully implemented a convolutional neural network for the assessment of retinal fundus image quality in retinopathy of prematurity. However, the implications of this study are extensive. This algorithm will be evaluated on a set of diabetic retinopathy images to test generalizability using a different disease and adult subjects. In general, we have shown that a convolutional neural network is sufficient for providing a high degree of discrimination between low- and high-quality images, and that it has learned the features that make one image better or worse than another. This type of model has potential to be implemented as a quality metric for all imaging modalities, from fundus imaging to X-rays to MRIs.

Acknowledgements

This study was supported by grants T15 LM007088, K99 LM12238 and P30 EY10572 from the National Institutes of Health (Bethesda, MD), and by unrestricted departmental funding from Research to Prevent Blindness (New York, NY).

Conflicts of Interest:

No financial conflicts of interest. Michael F. Chiang is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA), and a Consultant for Novartis (Basel, Switzerland). R.V. Paul Chan is a scientific advisory board for Visunex Medical Systems (Fremont, CA) and a Consultant for Alcon (Fort Worth, TX), Allergan (Irvine, CA), and Bausch and Lomb (St. Louis, MO). Jayashree Kalpathy-Cramer is a Consultant for INFOTECH Soft (Miami, FL).

References

1. American Academy of Pediatrics. Section on O. Screening examination of premature infants for retinopathy of prematurity. Pediatrics. 2001;108(3):809-11.
2. Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WP, 3rd, et al. American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin. 2003;53(3):141-69.
3. Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Vaituzis AC, Dickstein DP, et al. Quantitative brain magnetic resonance imaging in attention-deficit hyperactivity disorder. Arch Gen Psychiatry. 1996;53(7):607-16.
4. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44 Suppl 2:S27-72.
5. Wan KH, Chen LJ, Young AL. Screening and Referral of Diabetic Retinopathy: A Comparative Review of the Practice Guidelines. Asia Pac J Ophthalmol (Phila). 2013;2(5):310-6.
6. Bartlett E, DeLorenzo C, Parsey R, Huang C. Noise contamination from PET blood sampling pump: Effects on structural MRI image quality in simultaneous PET/MR studies. Med Phys. 2018;45(2):678-86.
7. Patel T, Peppard H, Williams MB. Effects on image quality of a 2D antiscatter grid in x-ray digital breast tomosynthesis: Initial experience using the dual modality (x-ray and molecular) breast tomosynthesis scanner. Med Phys. 2016;43(4):1720.
8. Smet MH, Breysem L, Mussen E, Bosmans H, Marshall NW, Cockmartin L. Visual grading analysis of digital neonatal chest phantom X-ray images: Impact of detector type, dose and image processing on image quality. Eur Radiol. 2018.
9. Strauss RW, Krieglstein TR, Priglinger SG, Reis W, Ulbig MW, Kampik A, et al. Image quality characteristics of a novel colour scanning digital ophthalmoscope (SDO) compared with fundus photography. Ophthalmic Physiol Opt. 2007;27(6):611-8.
10. Veiga D, Pereira C, Ferreira M, Goncalves L, Monteiro J. Quality evaluation of digital fundus images through combined measures. J Med Imaging (Bellingham). 2014;1(1):014001.
11. Teich S, Al-Rawi W, Heima M, Faddoul FF, Goldzweig G, Gutmacher Z, et al. Image quality evaluation of eight complementary metal-oxide semiconductor intraoral digital X-ray sensors. Int Dent J. 2016;66(5):264-71.
12. Briggs R, Bailey JE, Eddy C, Sun I. A methodologic issue for ophthalmic telemedicine: image quality and its effect on diagnostic accuracy and confidence. J Am Optom Assoc. 1998;69(9):601-5.
13. Chiang MF, Gelman R, Martinez-Perez ME, Du YE, Casper DS, Currie LM, et al. Image analysis for retinopathy of prematurity diagnosis. J AAPOS. 2009;13(5):438-45.
14. Chiang MF, Wang L, Busuioc M, Du YE, Chan P, Kane SA, et al. Telemedical retinopathy of prematurity diagnosis: accuracy, reliability, and image quality. Arch Ophthalmol. 2007;125(11):1531-8.
15. Lundberg T, Westman G, Hellstrom S, Sandstrom H. Digital imaging and telemedicine as a tool for studying inflammatory conditions in the middle ear--evaluation of image quality and agreement between examiners. Int J Pediatr Otorhinolaryngol. 2008;72(1):73-9.
16. Richter GM, Williams SL, Starren J, Flynn JT, Chiang MF. Telemedicine for retinopathy of prematurity diagnosis: evaluation and challenges. Surv Ophthalmol. 2009;54(6):671-85.
17. Takeda H, Minato K, Takahasi T. High quality image oriented telemedicine with multimedia technology. Int J Med Inform. 1999;55(1):23-31.
18. Wang S, Jin K, Lu H, Cheng C, Ye J, Qian D. Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs. IEEE Trans Med Imaging. 2016;35(4):1046-55.
19. Ataer-Cansizoglu E, Bolon-Canedo V, Campbell JP, Bozkurt A, Erdogmus D, Kalpathy-Cramer J, et al. Computer-Based Image Analysis for Plus Disease Diagnosis in Retinopathy of Prematurity: Performance of the "i-ROP" System and Image Features Associated With Expert Diagnosis. Transl Vis Sci Technol. 2015;4(6):5.
20. Campbell JP, Ataer-Cansizoglu E, Bolon-Canedo V, Bozkurt A, Erdogmus D, Kalpathy-Cramer J, et al. Expert Diagnosis of Plus Disease in Retinopathy of Prematurity From Computer-Based Image Analysis. JAMA Ophthalmol. 2016;134(6):651-7.
21. Institute NE. Retinopathy of Prematurity 2018 [Available from: https://nei.nih.gov/health/rop.
22. Blencowe H, Lawn JE, Vazquez T, Fielder A, Gilbert C. Preterm-associated visual impairment and estimates of retinopathy of prematurity at regional and global levels for 2010. Pediatr Res. 2013;74 Suppl 1:35-49.
23. Quinn GE. Retinopathy of prematurity blindness worldwide: phenotypes in the third epidemic. Eye Brain. 2016;8:31-6.
24. James M. Brown JK-K, J. Peter Campbell, Sang Jin Kim, Susan Ostmo, Paul Chan, Michael F. Chiang. Deep ROP. 2018.

25. Bartling H, Wanger P, Martin L. Automated quality evaluation of digital fundus photographs. Acta Ophthalmol. 2009;87(6):643-7.
26. Dietrich TJ, Ulbrich EJ, Zanetti M, Fucentese SF, Pfirrmann CW. PROPELLER technique to improve image quality of MRI of the shoulder. AJR Am J Roentgenol. 2011;197(6):W1093-100.
27. Gajendra Jung Katuwal JK, Rajeev Ramchandran, Christye Sisson, and Navalgund Rao. Automatic Fundus Image Field Detection And Quality Assessment. IEEE Xplore. 2013.
28. Giancardo L, Abramoff MD, Chaum E, Karnowski TP, Meriaudeau F, Tobin KW. Elliptical local vessel density: a fast and robust quality metric for retinal images. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:3534-7.
29. Jiang Y, Huo D, Wilson DL. Methods for quantitative image quality evaluation of MRI parallel reconstructions: detection and perceptual difference model. Magn Reson Imaging. 2007;25(5):712-21.
30. Li H, Hu W, Xu ZN. Automatic no-reference image quality assessment. Springerplus. 2016;5(1):1097.
31. Maberley D, Morris A, Hay D, Chang A, Hall L, Mandava N. A comparison of digital retinal image quality among photographers with different levels of training using a non-mydriatic fundus camera. Ophthalmic Epidemiol. 2004;11(3):191-7.
32. Niemeijer M, Abramoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Med Image Anal. 2006;10(6):888-98.
33. Sajib Kumar Saha BF, Jorge Cuadros, Di Xiao, Yogesan Kanagasingam. Deep Learning for Automated Quality Assessment of Color Fundus Images in Diabetic Retinopathy Screening. 2017.
34. Chiang MF, Starren J, Du YE, Keenan JD, Schiff WM, Barile GR, et al. Remote image based retinopathy of prematurity diagnosis: a receiver operating characteristic analysis of accuracy. Br J Ophthalmol. 2006;90(10):1292-6.
35. Simonyan KZ, Andrew. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR 20152014.
36. Krizhevsky A SI, and Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems. 2012(25):1097-105.