4-1-2023

Methods for Exploiting High-Resolution Imagery for Deep Learning-Based Diabetic Retinopathy Detection and Grading

Adam M. Saunders
University of Dayton

Methods for Exploiting High-Resolution Imagery for Deep Learning-Based Diabetic Retinopathy Detection and Grading

Honors Thesis

Adam Saunders

Department: Electrical and Computer Engineering

Advisor: Russell Hardie, Ph.D.

April 2023

Methods for Exploiting High-Resolution Imagery for Deep Learning-Based Diabetic Retinopathy Detection and Grading

Honors Thesis

Adam Saunders

Department: Electrical and Computer Engineering

Advisor: Russell Hardie, Ph.D.

April 2023

Abstract

Diabetic retinopathy is a disease that affects the eyes of people with diabetes, and it can cause blindness. To diagnose diabetic retinopathy, ophthalmologists image the back surface of the inside of the eye, a process referred to as fundus photography. Ophthalmologists must then diagnose and grade the severity of diabetic retinopathy by analyzing details in the image, which can be difficult and time-consuming. Alternatively, due to the availability of labeled datasets containing fundus images with diabetic retinopathy, AI methods like deep learning can provide automated detection and grading algorithms. We show that the resolution of an image has a large effect on the accuracy of grading algorithms. So, we study several techniques to increase the accuracy of the algorithm by taking advantage of higher-resolution data, including using a region of interest as the input and applying an image transformation to make the circular fundus image square. While none of our proposed methods result in an increase in performance for grading diabetic retinopathy, the circle to square transformation results in an increase in accuracy and AUC for detection of diabetic retinopathy. This work provides a useful starting point for future research aimed at increasing the resolution content in a fundus image.

University *of* Dayton

Table of Contents

1. Introduction

Diabetic retinopathy (DR) is a disease that can cause blindness in people with diabetes. It is the leading cause of vision loss among the elderly. High blood sugar and high blood pressure among people with diabetes can lead to damage of the blood vessels in the back of the eye. This damage impacts vision negatively. DR is progressive, meaning vision gets worse as time goes on. People with severe DR can experience growth of new blood vessels and retinal detachment. It is important that healthcare workers screen people with diabetes for DR early and often, as there are little to no symptoms until the late stages of DR [1].

The process for diagnosis and monitoring DR can be tedious for doctors. Doctors have to look at fine details on the back of a patient's eye, called the fundus. Due to a lack of medical resources, there are often delays in the diagnosis and treatment of DR [1]. Unfortunately, developing nations like India suffer from severe shortages of ophthalmologists, especially in rural areas [2]. As a result, there is a demand for imaging techniques to monitor DR remotely in a telehealth setting. For these remote screening programs, patients can use retinal cameras designed for use with a smartphone [1,2].

There are two imaging modalities commonly used to monitor DR: optical coherence tomography (OCT) and color fundus photography (CFP) [2]. OCT involves constructing an image from reflected light, while CFP is a much simpler process that takes a color image of the fundus. The retina can be enhanced in CFP images using fluorescein angiography to dye the blood vessels [3]. Image processing algorithms are well-suited to studying DR, as OCT and CFP images of the eye can help diagnose and grade the severity of DR.

Due to the availability of datasets with labeled and graded fundus images, deep learning algorithms are a popular technique used to detect and classify DR severity [2-5]. In fact, the United States Food and Drug Association approved the first AI method for DR detection in 2018 [1]. While DR detection is a simple binary classification, DR grading uses multiple classes to describe the staging of DR. Many clinicians grade DR on a discrete scale from 0 to 4, where 0 indicates no DR and 4 indicates proliferative DR. However, some datasets use a scale from 0 to 5, and others simply partition the classes into non-proliferative DR and proliferative DR, where non-proliferative DR can include mild DR cases alongside cases without DR [3,6].

The datasets we use for DR detection and grading are often filled with images that have many different sizes and resolutions. The images contain varying levels of light, and the optic nerve may be in different areas. Often, deep learning methods trained on these datasets use data augmentation techniques like rotation, scaling, and translation in order to increase the limited data found in fundus image datasets [2,7]. Additionally, image preprocessing techniques like cropping, padding, and histogram modification methods can help make fundus image datasets more uniform [8]. One preprocessing method used by several researchers is to apply some sort of color normalization or histogram equalization like contrast-limited adaptive histogram equalization followed by some sort of denoising filter like a Gaussian blur or a median filter [6-10].

Due to the limited amount of data in some DR datasets, many researchers have taken advantage of pretrained networks and used transfer learning to detect and grade DR [3,6-10,11,12]. Transfer learning is particularly useful due to the limited data in many DR datasets. Researchers have shown that transfer learning for DR detection requires many fewer images than training a new model with only a slight decrease in performance [6].

One popular convolutional neural network used is Inception-v3, which uses inception modules that perform several levels of convolution and concatenate the results. Inception-v3 has outperformed other deep learning models for DR detection and grading [7,11,12]. This method results in deep learning models that are able to learn several levels of detail within a local set of layers. Inception-v3 replaces the large 5×5 convolutional layers in inception modules with multiple smaller 3×3 convolutional layers, resulting in a network that can achieve high performance using fewer parameters and less computational complexity [13]. Figure 1 shows an example of an inception module used in Inception-v3.

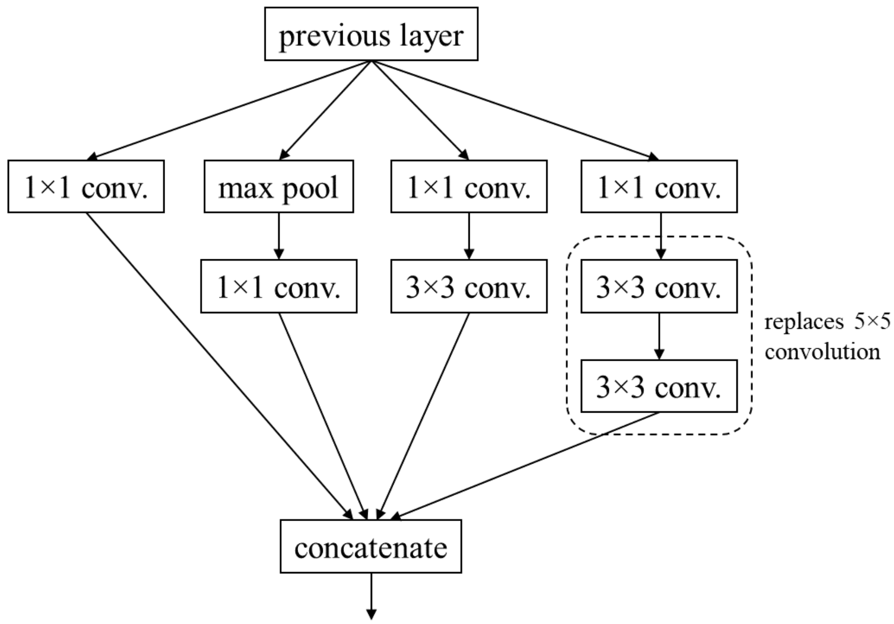

Figure 1: Example of an inception module in Inception-v3 (based on Ref. 13)

Several researchers have used deep learning models to extract features and then classified these features using a support vector machine [7,8,12]. We can use a single model like Inception-v3 to extract features, or we can extract features from several different deep learning models to create a large set of features. We can reduce this set of features by using statistical methods to select statistically significant features [7]. Alternatively, we can perform dimensionality reduction on the set of features using principal component analysis. Researchers have achieved a high accuracy using this method, especially for DR grading, which can be difficult with limited data for each of the classes [8].

The image preprocessing methods for DR detection and grading often resize the image to the input size of the model used for transfer learning. The input size can be a fraction of the size of the original image, meaning deep learning algorithms for DR detection and grading are not taking advantage of the available resolution content. Notably, one study tested their DR detection network using high-resolution fundus images alongside low-resolution fundus images and found a large increase in performance compared to using the low-resolution images alone [14]. The downsizing process may remove valuable information from the image. Additionally, since the images are not usually square and the input size of deep learning models is typically square, the resizing process can corrupt the spatial relationships present in the image.

There are two goals for this thesis. First, we wish to study how image resolution affects the accuracy of DR grading techniques. Next, we wish to study the effect of techniques that take advantage of the available high resolution of fundus images to see if they can increase the accuracy of DR detection and grading algorithms. In particular, we attempt to use a region of interest (ROI) as the input to a deep learning model. We also try using a spatial transformation to convert the circular fundus image to a square. Section 2 describes the methods used for preprocessing the images and the experiments performed, as well as details on the training and testing process for the deep learning models. Section 3 presents results from the experiments. Finally, Section 4 provides some conclusions about what we learned from this study and recommendations for further research.

2. Methods

2.1 Image Preprocessing

We use the APTOS 2019 dataset. The dataset consists of 2,930 CFP images graded on a scale from 0 to 4, with 0 indicating no DR and 4 indicating proliferative DR. The images are taken from a variety of locations using different cameras [4]. The classes are extremely unbalanced, as demonstrated in Table 1. The images are also extremely varied, as shown in Figure 2. The images feature different resolutions, aspect ratios, and lighting conditions. Even the size of the fundus is different in the images, with some containing the entire circular fundus and others containing only a portion of the fundus.

Since the images in the dataset have varying sizes and colors, it is important to preprocess the images to create a more uniform dataset. We compare several different preprocessing techniques for DR grading to show ours provides a higher accuracy at the standard input resolution of Inception-v3, the deep learning model used here for classification. The results of this comparison can be found in Sec. 3.1. Here, we describe the full preprocessing used for all other experiments.

Table 1: Distribution of images across classes in the APTOS 2019 dataset

Class	Number of Images
0	1434
1	300
2	808
3	154
4	234

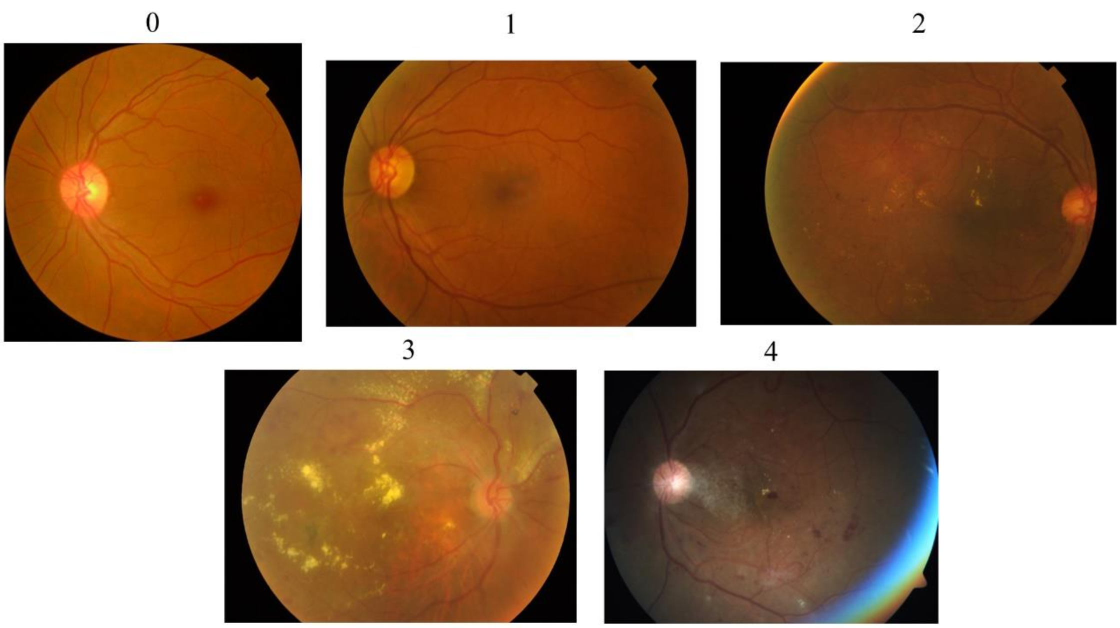

Figure 2: Examples from the APTOS 2019 dataset: 0 is no DR, 1 is mild DR, 2 is moderate DR, 3 is severe DR, and 4 is proliferative DR

First, we shift the image by its mean RGB value so that the mean value is zero. Then, by thresholding with a value of 10, we create a mask to cover the background of the image. We fill in any holes in the mask so that only the eye is showing. We normalize each RGB color channel so that they each have 2.5 standard deviations within 0 and 1. We calculate the mean and standard deviation using only the fundus portion of the image, not the background. We set the background of the image to 0.

Next, to find the optic disc, we select the point in the image with the maximum value after applying a Gaussian low-pass filter. The filter helps to find the brightest region of the image instead of simply selecting a bright pixel from noise or other artifacts. If the optic nerve is on the left half of the image, we flip it to be on the right half. Finally, we crop out the rows and columns that contain only zero-valued pixels. We pad the image with rows and columns of zeros so that the image is square. This way, we do not change the aspect ratio of the image when we resize it. Finally, we resize the image to the appropriate input size for our given network. Figure 3 shows the steps of the preprocessing algorithm for an example fundus image.

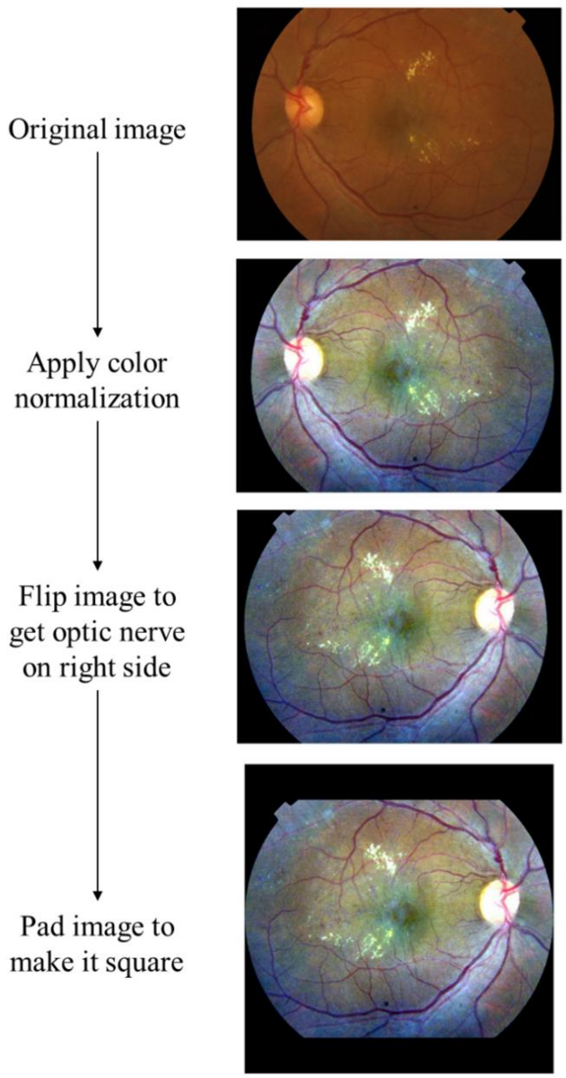

Figure 3: Fundus image preprocessing

2.2 Transfer Learning Architecture and Performance Metrics

For all of the experiments presented here, we adopt a transfer learning approach. For DR detection, we use all of the images from the APTOS 2019 dataset, with label 0 indicating no DR and labels 1-4 indicating DR. For DR grading, we only use the images that contain DR. That is, we use images with labels 1 through 4. Next, we split the data into training and testing data using k-fold cross-validation, where $k = 10$. We set aside 10% of the training data to use as validation data. We use the preprocessing described in Sec. 2.1. The selected deep learning algorithm for transfer learning is Inception-v3, as it outperformed other deep learning models in several DR detection and grading studies [7,11,12]. Since our preprocessing already involves color normalization, we use no normalization in the input layer. We replace the final fully connected layer with a fully connected layer containing two output nodes for DR detection and four output nodes for DR grading. Figure 4 summarizes the transfer learning architecture.

We train the networks using the Adam optimizer with the default parameters of $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. We use an initial learning rate of $\eta = 0.001$ and a mini-batch size of 32. We train for a maximum of 10 epochs, with early stopping enabled if the validation accuracy has not increased for three consecutive epochs. We report results from the epoch that had the best validation loss during training. We train the models in Matlab using two NVIDIA GeForce RTX 3090 GPUs.

We record two performance metrics to compare the performance of the models. First, we calculate the accuracy. We define accuracy as the number of correct classifications divided by the number of test samples. We also measure the area under curve (AUC) of the receiver operating characteristic (ROC) curve. The ROC curve is a plot of true positive rate versus false positive rate for many different thresholds values for the classification scores. Since DR grading is a multi-class problem, we select class 2 (moderate DR) as the positive class and treat all others as a negative class. We use class 2 since it has the most samples in the APTOS 2019 dataset out of the classes with DR. Figure 5 shows an example ROC curve for DR grading.

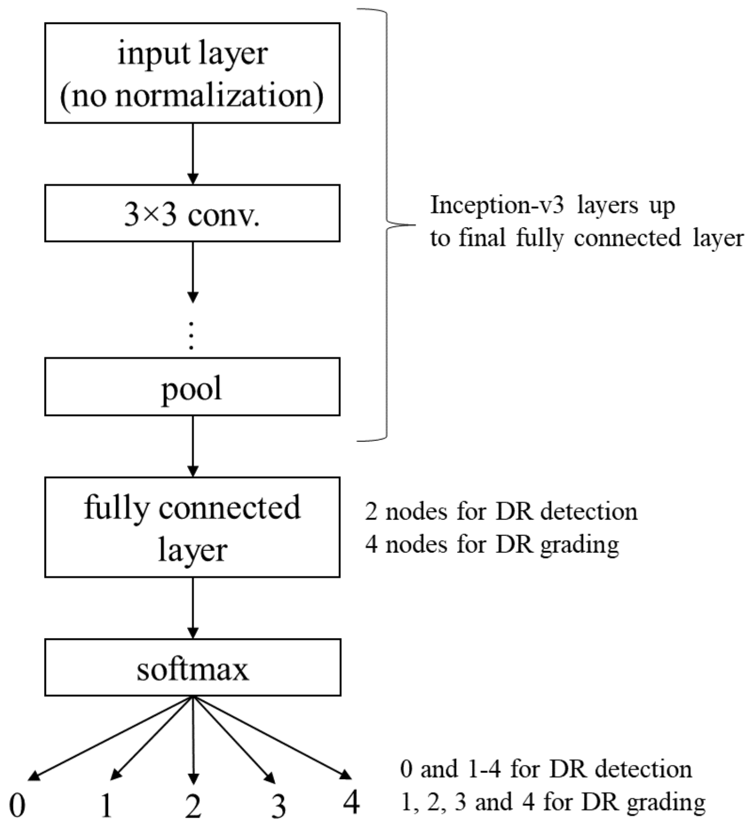

Figure 4: Transfer learning architecture for DR detection and grading

2.3 Resolution Study

We aim to study the effect of resolution on the success of grading the severity of images that have DR. That is, we wish to find if using a different resolution for the images affects the performance of the grading algorithm. After preprocessing the images using the methods described in Sec. 2.1, we train and test the network using the process described in Sec. 2.2. We train and test several models using different values of square resolutions, at the original input size of Inception-v3 (299 by 299 pixels), as well as several different larger and smaller scales. Figure 6 shows a comparison of an example fundus image at several scales, though note that the images are too large to be drawn to scale. Section 3.2 provides the results of this experiment.

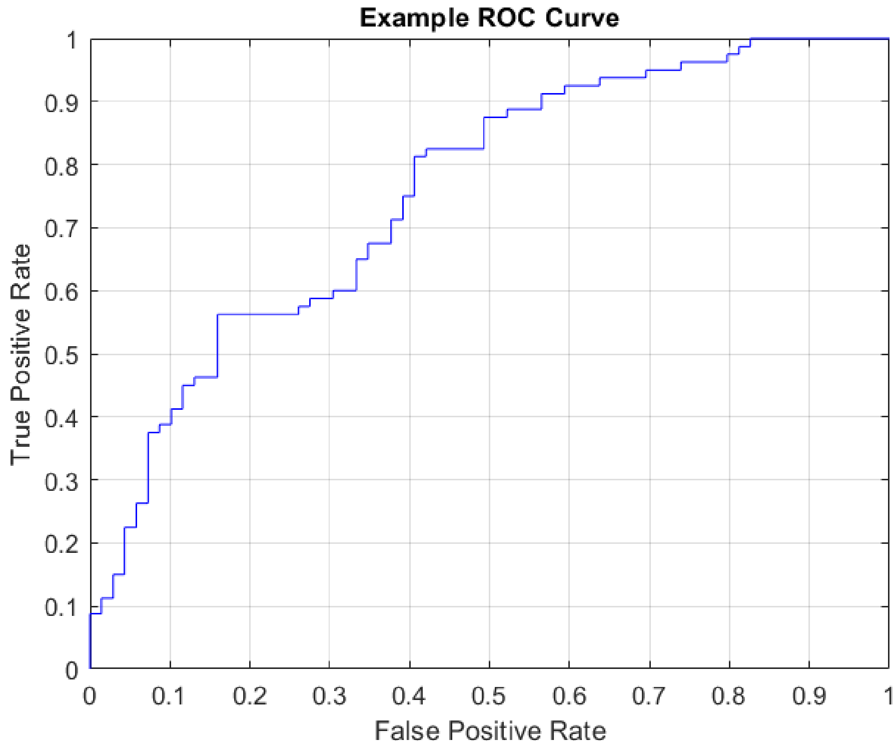

Figure 5: Example ROC curve for DR grading (AUC = 0.7551)

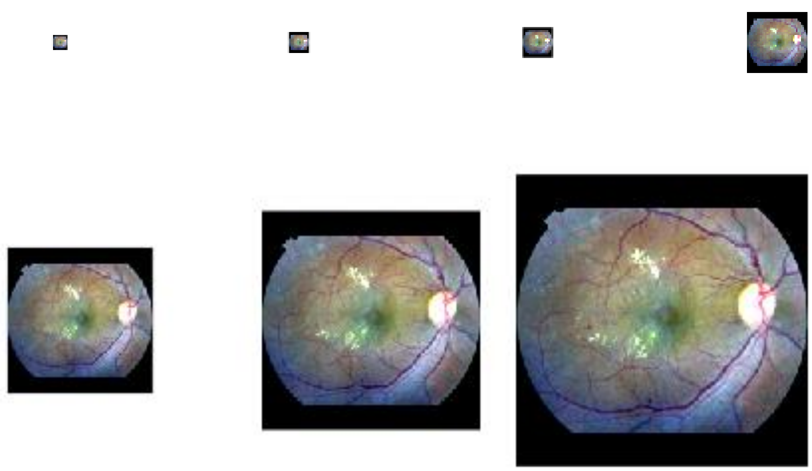

Figure 6: Example fundus image at several resolution scales (1/4, 1/3, 1/2, 1, 2, 3 and 4, not drawn to scale)

2.4 ROI Input

To make use of the high available resolution, we next introduce an ROI to input a clinically useful region to the network. We compare this method for both DR detection and grading. To select this region, we first input a downsized image at 224 by 224 pixels into ResNet50 as a preliminary classifier. We again use the preprocessing described in Sec. 2.1.

After ResNet50 classifies the image, we use gradient class-activation mapping (grad-CAM) to generate a heatmap that shows the areas most relevant to the classification of the image. We use CAM to find these areas since researchers have previously used CAM heatmaps as an interpretability measure for deep learning models [9]. We then select the pixel with the highest value from the heatmap. We select the corresponding pixel on the original high-resolution image as the center of our ROI. We crop an ROI of size 500 by 500 pixels from around the center pixel on the high-resolution image. If the ROI includes pixels beyond the boundary of the image, we shift the center of the ROI to make sure the ROI does not sample from outside of the image. After using bicubic interpolation to downsize this ROI to the input size of Inception-v3 at 299 by 299 pixels, we use this ROI as the input to our network as before. Again, we measure the accuracy and AUC using 10-fold cross-validation for this method. We train both networks using the process described in Sec. 2.2. Figure 7 shows the ROI selection process applied to an example fundus image, and Sec. 3.3 describes the results of this experiment.

2.5 Circle to Square Spatial Transformation

Fundamentally, fundus images are circular. However, deep learning models and image processing algorithms work mostly on square images. So, a portion of our fundus images consist of black background that is not useful for classification. Aiming to reduce the amount of background in the image, we next use a spatial transformation to convert the circular fundus image into a square image.

Figure 7: ROI used as input to the network: original image (top left), grad-CAM heatmap from ResNet50 (top right), cropped ROI from highest point on heatmap (bottom)

There are many transformations to convert a circle into a square. Among the transformations that preserve the location of the center as well as the points on the x-axis and y-axis, none of them preserve both angles and area [15]. However, one simple transformation is the elliptical grid mapping. The inverse transformation of the elliptical grid mapping for a circle of radius 1 centered at the origin is given by

$$u = x\sqrt{1 - \frac{y^2}{2}} \tag{1}$$

and

$$v = y\sqrt{1 - \frac{x^2}{2}}, \quad (2)$$

where (u, v) are the coordinates of the circle and (x, y) are the coordinates of the square [15].

We apply this circle to square spatial transformation to our fundus image dataset. We compare this method for both DR detection and grading. We scale this transformation by the radius of the fundus, measured along the x-axis. We apply the transformation after preprocessing the images as described in Sec. 2.1 but before resizing to the input size of Inception-v3.

Note that many of the images in the fundus image dataset are not truly circular as the tops and bottoms of the image are cut off. So, in addition to measuring the accuracy and AUC after using this transformation, we also measure the accuracy and AUC after using the transformation and cropping out the rows that contain black pixels along the vertical line of the center of the image. This results in an image that may not be square, so the downsizing process may slightly change the spatial relationships present in the image. For both experiments, we again use 10-fold cross-validation and the training process described in Sec. 2.2. Figure 8 shows the image transformation performed on an example fundus image, as well as the image transformation after cropping. Section 3.3 discusses the results for this experiment.

3. Results

3.1 Image Preprocessing

To demonstrate that our preprocessing technique results in an increased accuracy, we compare the results of our transfer learning process using different preprocessing techniques. We refer to resizing the image alone as “resize.” We call resizing and normalizing the image by the average intensity of all RGB channels together “norm.” Next, we call resizing and normalizing each color channel separately “color norm.”

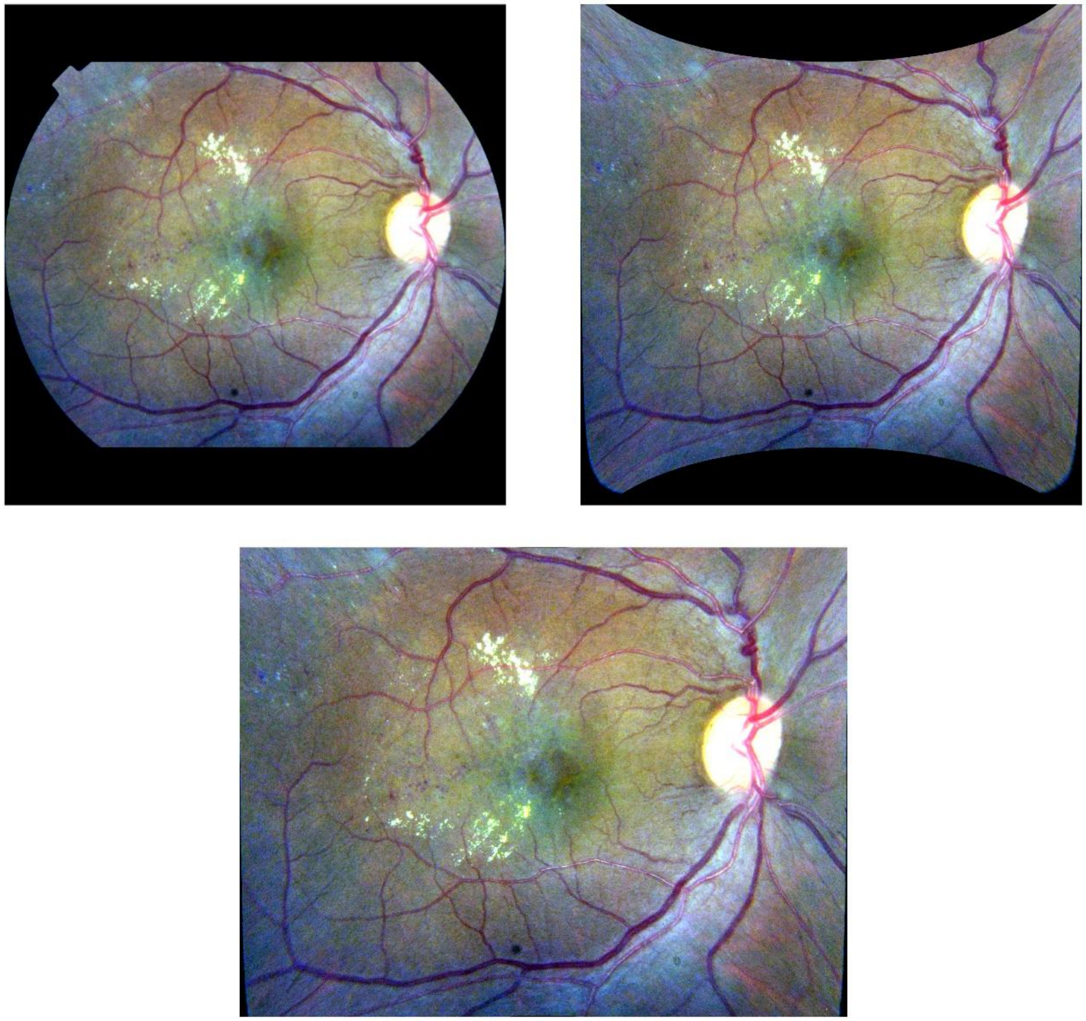

Figure 8: Circle to square image transformation: original image (top left), transformed image (top right), and transformed image after cropping (bottom)

We try zero-padding the image before resizing, and we also include flipping the image so that the optic nerve lies on the same side for each image, referred to as "norm. + pad + flip." Finally, we try this method with normalizing each color channel separately, referred to as "color norm. + pad + flip." Figure 9 shows a visual comparison of the preprocessing techniques for an example fundus image.

Figure 10 shows the results of using the preprocessing techniques with 10-fold cross-validation. We compare the results for DR grading alone. We notice that the "color norm.

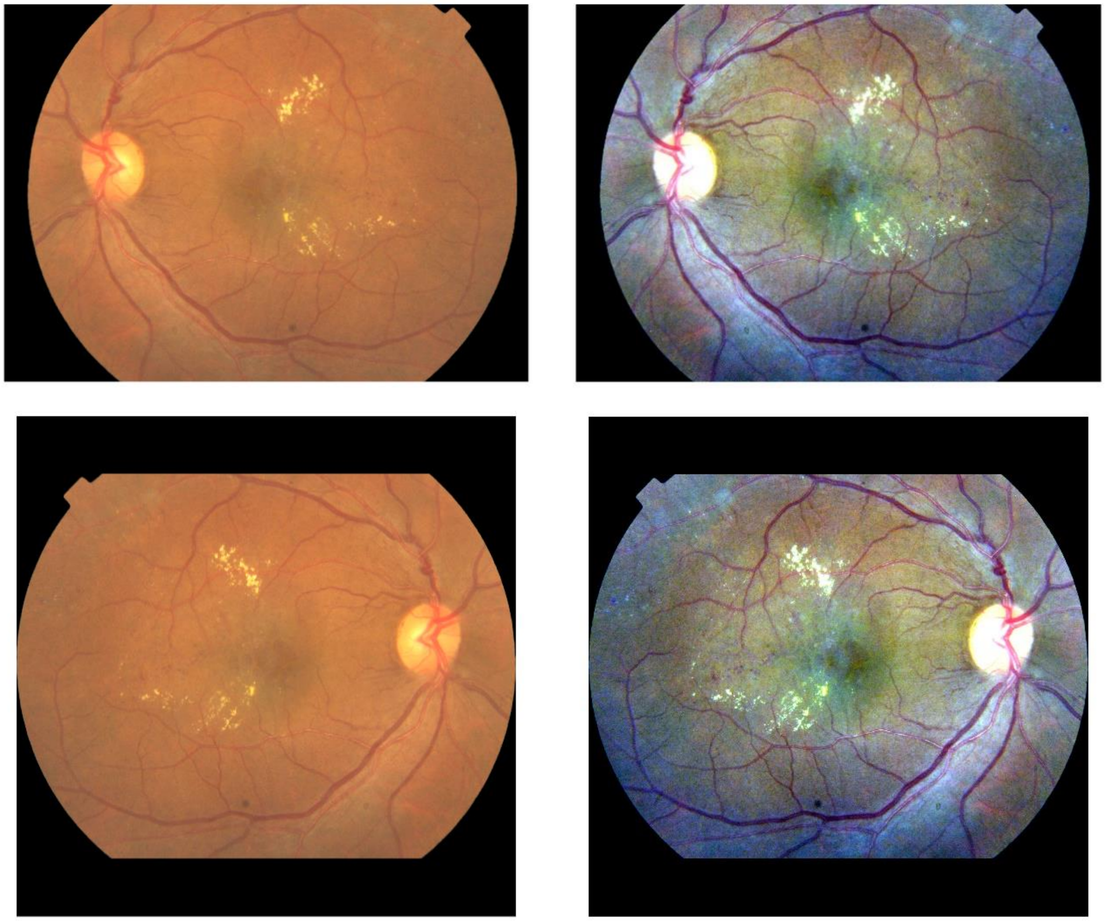

Figure 9: Preprocessing techniques before resizing: "Norm." (top left), "color norm." (top right), "norm. + pad + flip" (bottom left), and "color norm. + pad + flip" (bottom right)

+ pad + flip” preprocessing technique results in the highest average accuracy, albeit with a slightly lower average AUC. “Color norm. + pad + flip” is the preprocessing technique we used for all other experiments, described earlier in Sec. 2.1.

3.2 Resolution Study

Next, we present the results from the resolution study described in Sec. 2.3. Figure 11 shows the accuracy and AUC of the Inception-v3 network used for DR grading after changing the input size to several different scales.

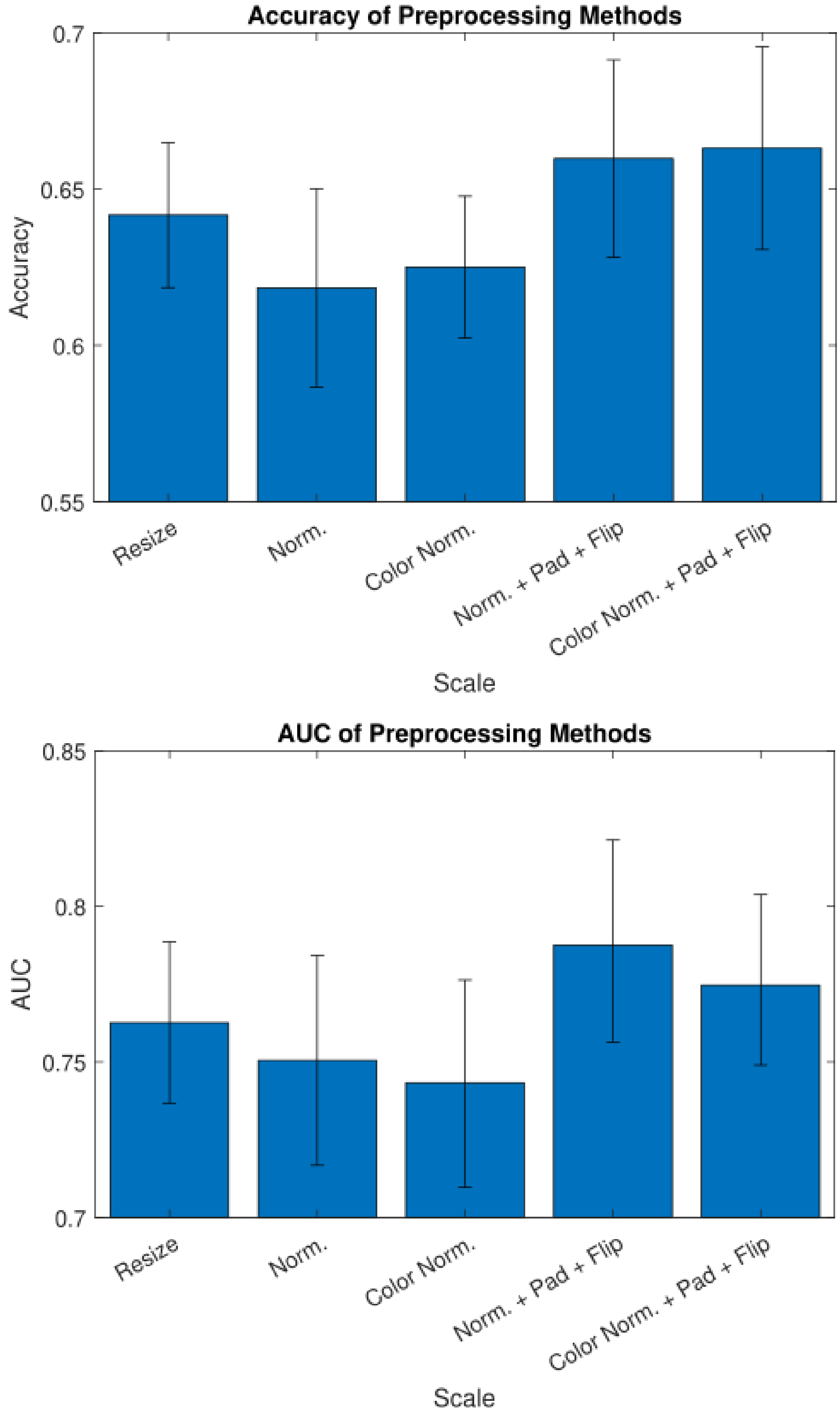

Figure 10: Accuracy and AUC of DR grading algorithm using different preprocessing methods

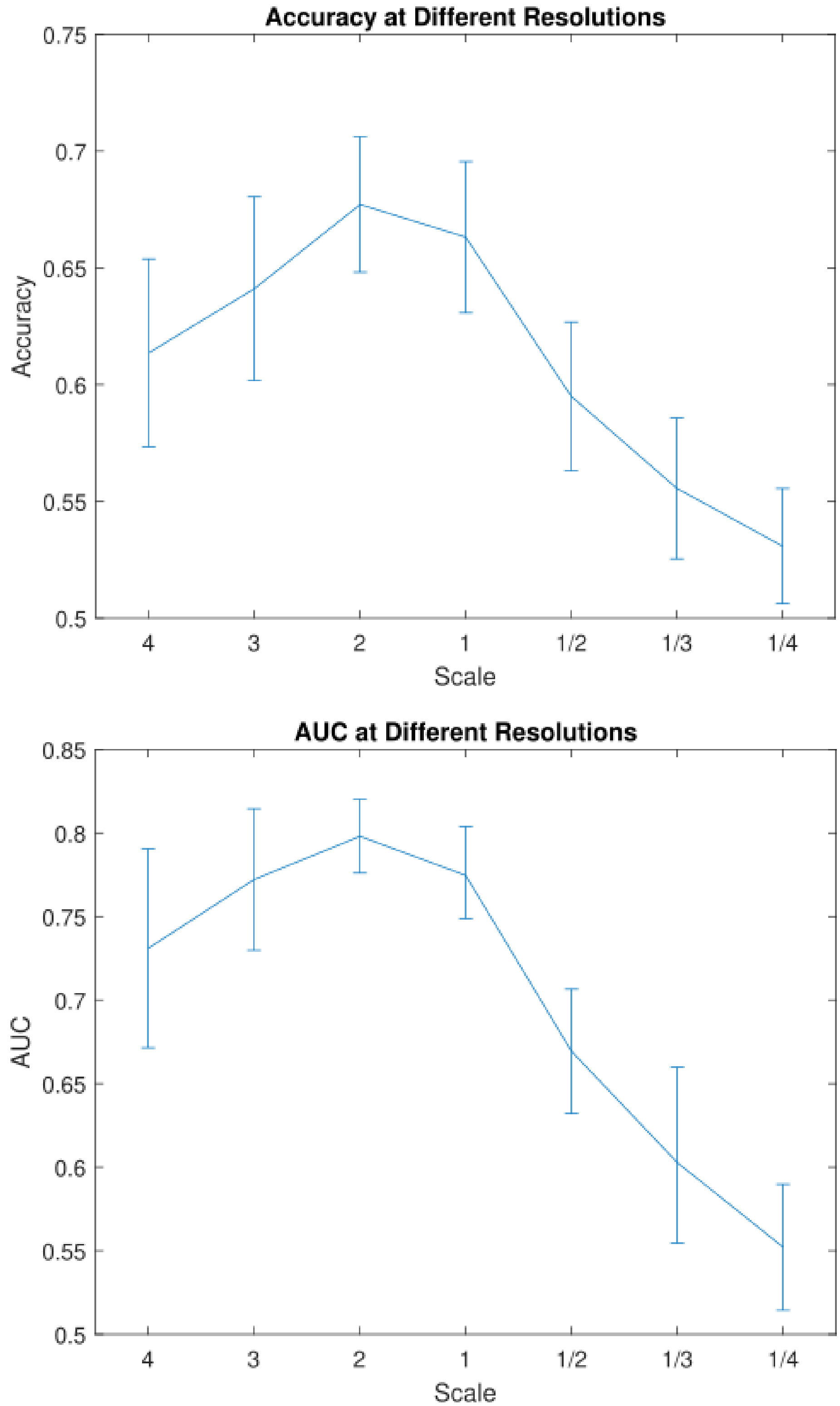

Figure 11: Accuracy and AUC of DR grading algorithm at different input resolutions

As we expect, the accuracy and AUC drop off at very low resolutions. However, we notice that the accuracy and AUC both peak at a scale higher than 1. That is, we achieve a higher accuracy and AUC by using a higher resolution than the default input size for Inception-v3, 299 by 299 pixels. This result indicates that we may want to take advantage of the higher resolution available from fundus photography. We also notice that the accuracy and AUC begin to drop off at very high scale factors. This decrease may be due to the fact that Inception-v3 cannot handle such large inputs with the current number of parameters.

3.3 ROI Input and Circle to Square Spatial Transformation

Next, we present the results when using the ROI as the input to the network as well as the results of the circle to square spatial transformation. Figure 12 shows the accuracy and AUC of the different image transformation methods for DR detection, while Figure 13 shows the results for DR grading. The image transformations do not achieve a higher accuracy or AUC than the unmodified images for DR grading. However, the circle to square transformations provides a slight boost in accuracy and AUC for DR detection.

The ROI input achieves a much lower accuracy and AUC than the other methods for both DR grading and detection. This decrease in accuracy and AUC might be due to a poor selection of the ROI. Additionally, the ROI may be too focused on one area of the image. Instead of inputting the ROI directly into the network, perhaps we could try inputting the ROI as a channel alongside the entire image. So, the network would have two channels: one containing a low-resolution fundus image and one containing a high-resolution ROI. Alternatively, we could concatenate the low-resolution fundus image and ROI to create a single input for the network.

The circle to square transformations performed slightly worse than the unmodified images for DR grading, though not as poorly as the ROI input. It might be that the spatial relationships in a fundus image are very important for the grading of DR. The transformation may have resulted in too much corruption of the spatial relationships present in the image. On the other hand, the circle to square transformations provided a slight increase in performance for DR detection. Cropping the rows that contain only black pixels along the y-axis results in an increase in the average accuracy and AUC compared to the transformation without the cropping for both DR grading and detection. This result seems to indicate that the DR detection and grading algorithms may perform better when we remove as much of the black background area as possible. Although all circle to square transformations will result in some level of deformation, we could also try other transformations than the elliptical grid mapping.

4. Conclusion

In this thesis, we presented and discussed several methods for increasing the performance of DR detection grading algorithms. First, we showed that preprocessing techniques like padding and color normalization can help standardize highly varied fundus image datasets. We also showed that DR grading algorithms perform better using higher resolutions than the default input size of many common deep learning algorithms used for transfer learning. Inspired by this result, we focused on ways to increase the resolution content of images for DR detection and grading. We tried using a smaller ROI selected from a high-resolution image. Additionally, we attempted to use a mathematical transformation to convert the circular fundus images to square images.

While none of our transformations outperformed the unmodified preprocessed images for DR grading, the circle to square transformation provided a slight increase in performance for DR detection. The transformations presented here provide a useful basis for future research. Future studies could focus on more robust methods for selecting the ROI, as well as inputting the high-resolution ROI alongside the entire low-resolution image. There are also many other circle to square image transformations that may result in less spatial deformation. We also are interested in studying nonlinear downsampling methods that sample the areas relevant for diagnosis using a higher sampling frequency. Overall, this thesis characterizes the importance of the relationship between deep learning-based DR detection and grading algorithm performance and the resolution of fundus images.

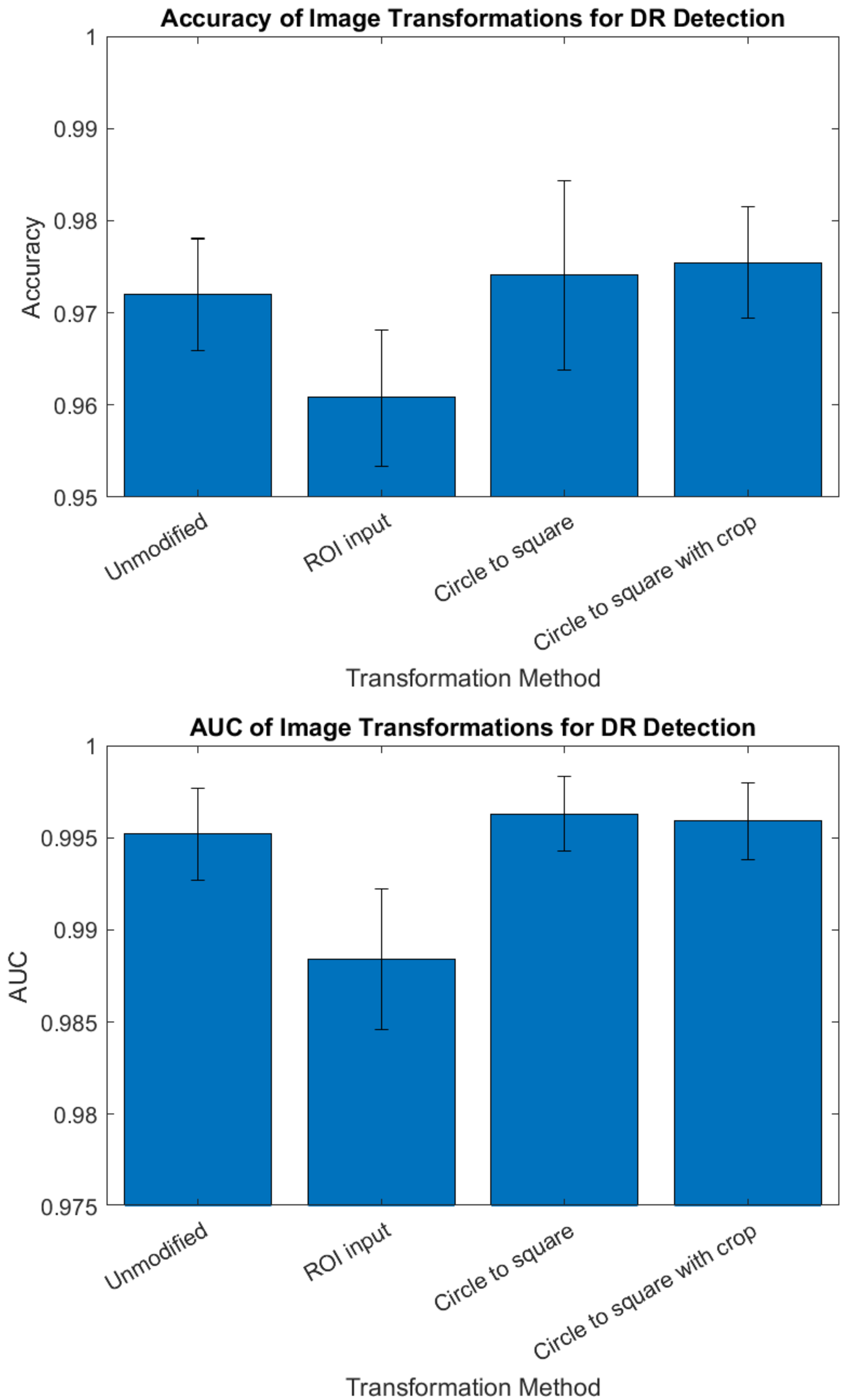

Figure 12: Accuracy and AUC of DR detection algorithms using different image transformation methods

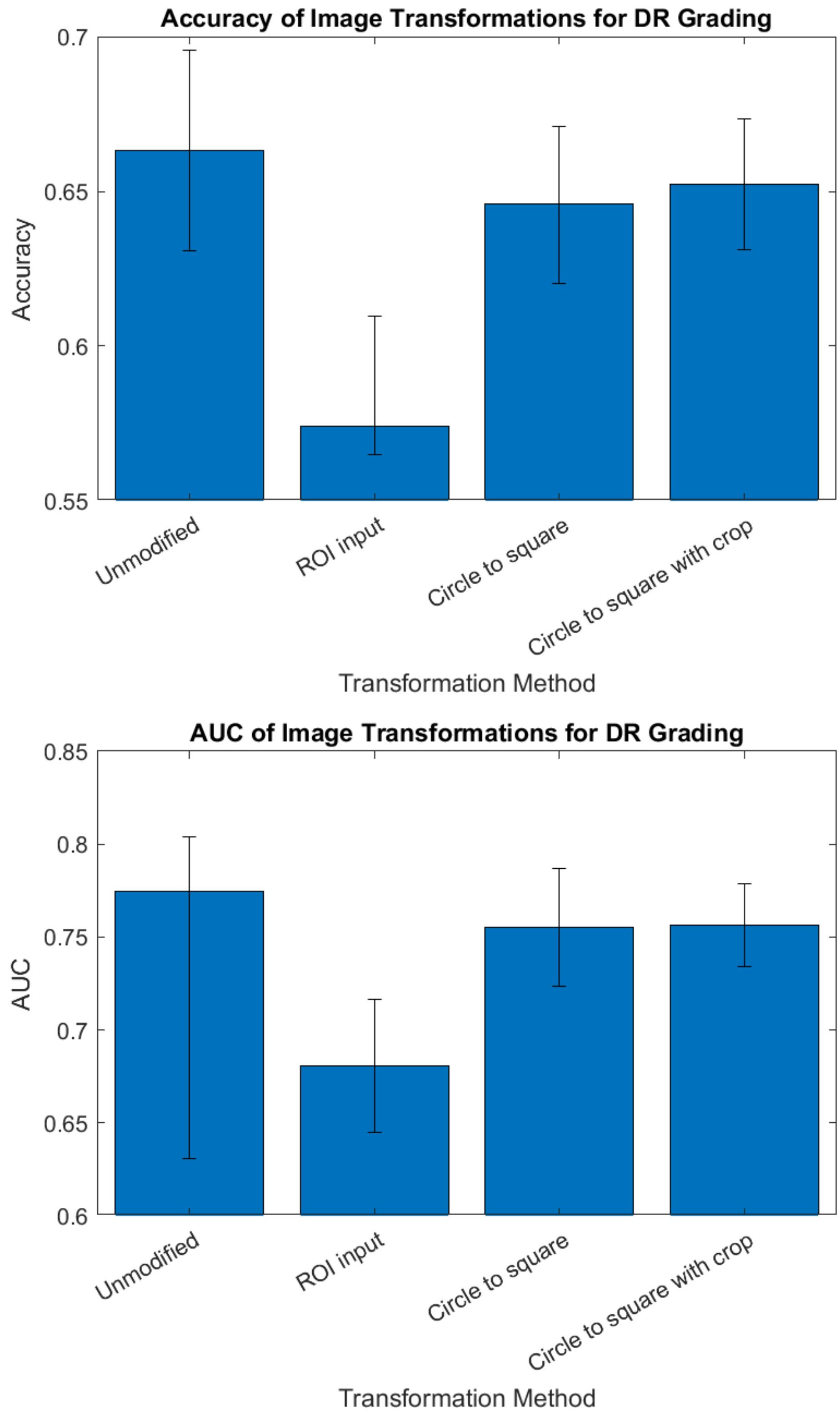

Figure 13: Accuracy and AUC of DR grading algorithms using different image transformation methods

References

[1] K.-Y. Lin, W.-H. Hsih, Y.-B. Lin, C.-Y. Wen and T.-J. Chang, "Update in the epidemiology, risk factors, screening, and treatment of diabetic retinopathy," *Journal of Diabetes Investigation,* vol. 12, no. 8, pp. 1322-1325, 2021.

[2] A. Jones, T. B. Vijayan, S. Subbarayan and S. John, "A systematic review on screening of diabetic retinopathy and maculopathy using artificial intelligence," in *Advances in Signal Processing and Communication Engineering: Select Proceedings of ICASPACE 2021*, Hyderabad.

[3] S. Masood, T. Luthra, H. Sundriyal and M. Ahmed, "Identification of diabetic retinopathy in eye images using transfer learning," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017.

[4] *APTOS 2019 Blindness Detection Dataset,* Kaggle, 2019.

[5] *Diabetic Retinopathy Detection,* Kaggle, 2015.

[6] F. Li, Z. Liu, H. Chen, M. Jiang, X. Zhang and Z. Wu, "Automatic Detection of Diabetic Retinopathy in Retinal Fundus Photographs Based on Deep Learning Algorithm," *Translational Vision and Science,* vol. 8, no. 6, 2019.

[7] C. Bhardwaj, S. Jain and M. Sood, "Transfer learning based robust automatic detection system for diabetic retinopathy grading," *Neural Computing and Applications,* vol. 33, pp. 13999-14019, 2021.

[8] B. N. Narayanan, R. C. Hardie, M. S. D. Silva and N. K. Kueterman, "Hybrid machine learning architecture for automated detection and grading of retinal images for diabetic retinopathy," *Journal of Medical Imaging,* vol. 7, p. 1 – 18, 2020.

[9] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma and W. Qian, "An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019.

[10] M. K. Jabbar, J. Yan, H. Xu, Z. U. Rehman and A. Jabbar, "Transfer Learning-Based Model for Diabetic Retinopathy Diagnosis Using Retinal Images," *Brain Sciences*, vol. 12, no. 5, 2022.

[11] R. Haripriya, B. Gopalakrishnan and P. Purusothaman, "Retinopathy Detection on Diabetic Patients through Transfer Learning based Inceptionv3 Algorithm," in *International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022)*, 2022.

[12] F. N. M. Noor, W. H. M. Isa, I. M. Khairuddin, M. A. Mohd, J. A. M. J. Razman, A. F. A. Nasir, R. M. Musa and A. P. P. A. Majeed, "The Diagnosis of Diabetic Retinopathy: A Transfer Learning with Support Vector Machine Approach," in *International Conference on Innovative Technology, Engineering and Science (iCITES 2020)*, 2021.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[14] A.-O. Asia, C.-Z. Zhu, S. A. Althubiti, D. Al-Alimi, Y.-L. Xiao, P.-B. Ouyang and M. A. A. Al-Qaness, "Detection of diabetic retinopathy in retinal fundus images using CNN classification models," *Electronics*, vol. 11, no. 17, 2022.

[15] C. Fong, "Analytical Methods for Squaring the Disc," in *International Congress of Mathematicians*, Seoul, 2014.