4-2019

A deep learning approach to detect diabetic retinopathy in fundus images.

Winston R. Furtado
University of Louisville

A DEEP LEARNING APPROACH TO DETECT DIABETIC RETINOPATHY IN FUNDUS IMAGES

By

Winston Furtado
B.Sc. in Bioengineering, University of Louisville, 2018

A Thesis
Submitted to the Faculty of the
J.B. Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree

Master of Engineering in Bioengineering

Department of Bioengineering
University of Louisville
Louisville, Kentucky

April 2019

A DEEP LEARNING APPROACH TO DETECT DIABETIC RETINOPATHY IN FUNDUS IMAGES

Submitted by: _____
Winston Furtado

A Thesis Approved on

_____
(Date)

by the Following Reading and Examination Committee:

_____
Ayman El-Baz, Ph.D. Thesis Director

_____
Adel Elmaghraby, Ph.D.

_____
Guruprasad A. Giridharan, Ph.D.

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr. El-Baz for his support over the past year. His mentorship has had a positive impact on me as a student and as a researcher; it has truly been rewarding. To the rest of my thesis committee, Dr. Elmaghraby and Dr. Giridharan, thank you for being a part of my thesis committee and helping me to achieve a special milestone in my college career. Also, I would like to thank Dr. Ahmed Shalaby for his guidance and assisting me whenever I needed help. Finally, I would like to thank my family for helping me stay motivated and determined.

ABSTRACT

A DEEP LEARNING APPROACH TO DETECT DIABETIC RETINOPATHY IN FUNDUS IMAGES

Winston Furtado

April 2019

Background: Diabetic retinopathy is a disease caused due by complications of diabetes mellitus which can lead to blindness. About 33% of the US population with diabetes also show symptoms for diabetes retinopathy. If not treated, diabetic retinopathy worsens over time by progressing through two main pathological stages of non-proliferative and proliferative and four clinical stages. While the diagnostic accuracy of detecting diabetic retinopathy through machine learning have shown to be successful for OCT images, the accuracy of ultra-widefield fundus images have yet to be fully reported. This paper describes a method to non-invasively detect and diagnose diabetic retinopathy from ultra-widefield fundus images.

Methods: A total of 62 graded-images were obtained from the Cleveland Clinic. A deep learning algorithm was developed to identify and extract features from the images. The algorithm was then simulated to classify the test images into one of three clinical classes. Data was collected on the accuracy and probability of the diagnosis/classification.

Results: The classification algorithm had an average accuracy that ranged from 92% to 97% for the training images and 50% for the test images. Confusion matrices were created to obtain statistical measures of performance such as sensitivity, false negative rate, precision, and the false discovery rate. The sensitivity decreased from 70% to 50% as the image size increased. The precision also decreased from 65% to 50% as the image size increased. Validation methods such as image normalization and transfer learning showed no improvement in classification accuracy.

Conclusion: This study demonstrates the potential for applying deep learning algorithms to classify ultra-widefield images. This study also demonstrates the need for doctors to further examine the diagnosis to account for false positives and/or misdiagnosis. Additionally, limitations and their impact on the simulation of the deep learning algorithm were explored.

Keywords: diabetic retinopathy; deep learning; CNN; image classification; ultra-wide field imaging

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Diabetic retinopathy (DR) is the leading cause of blindness among adults in the United States. It is the result of diabetes mellitus which itself is a global epidemic leading to a range of complications [1]. The main risk factors associated with diabetic retinopathy are duration of the diabetes, high blood sugar levels, and arterial hypertension [2][3][4]. At the onset, it affects the blood vessels in the retina, a photosensitive tissue found at the back of the eye. Overtime, under high elevation of glucose, the blood vessels can cause capillary endothelial damage which results in the cells along the capillary to become "leaky". The high glucose levels combined with endothelial damage lead to capillary occlusion which, in turn, leads to reduced blood flow [5]. If not treated, diabetic retinopathy progresses through four clinical stages (Figure 1) of mild non-proliferative, moderate non-proliferative, severe non-proliferative, which is sometimes subclassified into severe non-proliferative and very severe non-proliferative, and proliferative which is sometimes subclassified into non-high risk proliferative, and high risk proliferative, and advanced proliferative. In each stage there are problems that can occur (shown in Figure 2): such as microaneurysms (mild, moderate, severe, and proliferative stages), hemorrhages (moderate, severe, and proliferative stages), exudates (moderate, severe and proliferative stages), cotton-wool spots (severe, pre-proliferative, and proliferative stages), and intraretinal microvascular anomalies (proliferative stage) [5]. Microaneurysms are one of the earliest signs of diabetic retinopathy; classified as mild non-proliferative. They appear as red dots in the retina due to an expansion of the capillaries. In small numbers, microaneurysms do not affect the eyesight of a patient. Regardless, the number of

microaneurysms can be used to estimate the regression or progression of diabetic retinopathy. Hemorrhages occur due to the rupture of small blood vessels in the retina, which results in red lesions. Hemorrhages occur deep in the retina where their shape is determined by their location or position. Even though hemorrhages are a cause of concern, they do not functionally disrupt the eye. Like microaneurysms, the number of hemorrhages can be used to estimate the progression or regression of diabetic retinopathy. Exudates define the yellow-white lesions that occur due to the leakage of plasma from capillaries near the macular region. Exudates away from the macular region have been observed but have not been found to cause serious problems. The first sighting of exudates means that diabetic retinopathy has progressed from the mild to the moderate non-proliferative stage. Cotton-wool spots are caused by failure of capillary circulation, which leads to the swelling of nerve fibers. They appear as pale or fuzzy areas on the retina. Solely, the cotton-wool spots do not cause visual problems, but their presence is graded as pre-proliferative or severe depending on how widespread they are. Intraretinal microvascular anomalies (IRMAs) or microvascular abnormalities represent the first sign of new blood vessels appearing. The cause of IRMAs is due to the dilation of capillaries and the shunts between the arteries and veins. Additionally, the retinal veins appear tortuous and irregular. Any patient with IRMAs is classified as proliferative, in which case they must seek medical help immediately.

It is estimated that about 30% to 50% of the diabetic population has retinopathy. In longitudinal studies, about 30% of the affected individuals showed symptoms of diabetic retinopathy after 10 to 20 years of progression with diabetes

mellitus. Of the 30% affected, 10% progressed to proliferative retinopathy, which puts patients at risk of blindness. In 90% of the cases examined, non-proliferative retinopathy is responsible for a serious loss of visual capacity. Non-proliferative retinopathy is also observed to occur more often in cases where diabetes onset started in adulthood, whereas proliferative retinopathy is present “generally”, though not always in juvenile forms of the disease [5]. Another study, the classical Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR), reported that about 75% of people diagnosed with diabetes mellitus showed signs and symptoms of diabetic retinopathy after 10 years. For those who developed early symptoms, about 66% of the people exhibited the severe stage of diabetic retinopathy. About 20% of those affected had the disease progress to proliferative diabetic retinopathy or diabetic macular odema, a subtype of diabetic retinopathy that involves the macula [2][3][6][7].

Before 1961, to detect diabetic retinopathy, various diagnostic techniques were utilized. A visual acuity test was used to detect how well a person could see at various distances. A dilated eye exam was done to closely examine the retina and optic nerve for signs of damage. Additionally, a tonometry was done to detect abnormal pressure (<12 mmHg or >22 mmHg) inside the eye [8]. With the advance of diagnostic medicine came the arrival of new imaging technologies. One such technology was fluorescein angiography. Fluorescein angiography involved using a fluorescent dye to mark the eye after which pictures were taken to identify vascular leakage and the presence of ischemia [9] [10]. Although traditionally used as the primary method in diagnosing diabetic retinopathy, fluorescein angiography was phased out over time in

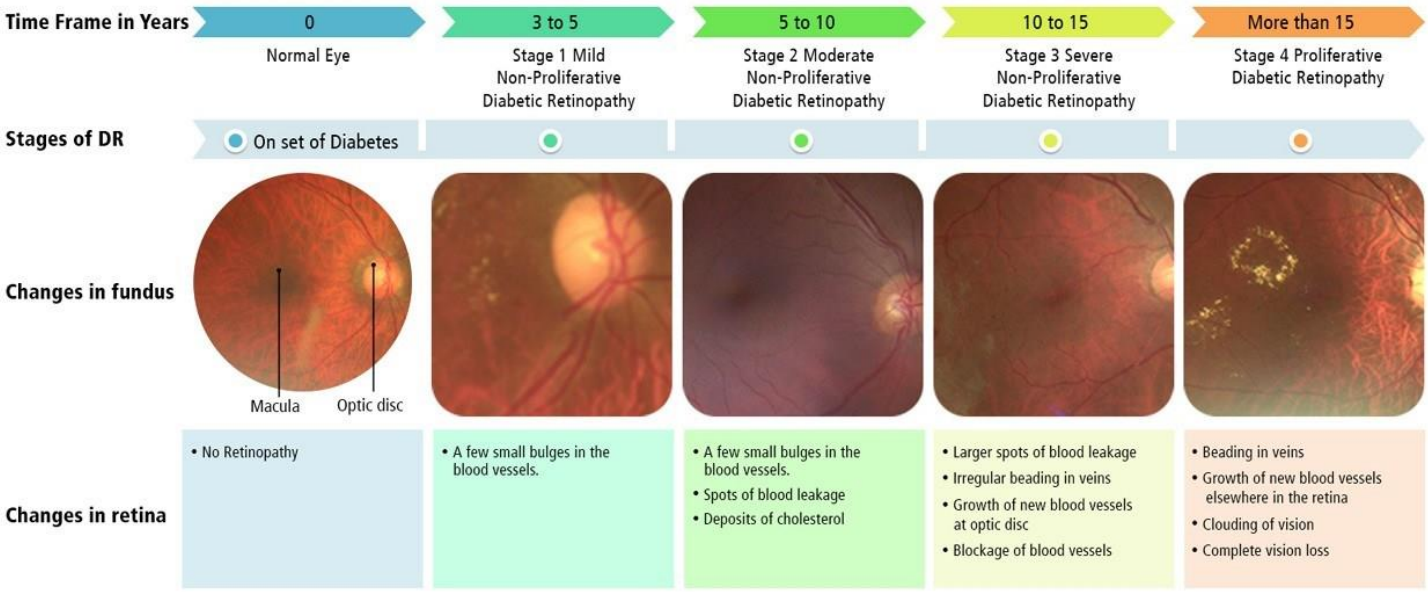


Figure 1 - Stages of diabetic retinopathy [31].

preference of optical coherence tomography (OCT). Optical coherence tomography was considered a major technical innovation that did not require the need for pupil dilation and could be adopted with telemedicine to offer fast, accurate, systematic solutions in both developed and developing countries. [11] [12] [13]. As a non-invasive diagnostic technology, it creates high-resolution images of the retina to record 3D structural changes (e.g. odema) which otherwise cannot be deduced through a clinical examination of retinal pictures. At present, new fundus photography techniques are being tested and developed (e.g. wide-field fluorescein angiography) with some already introduced to the market, including the ultra-widefield (UWF) retinal imaging device, Optos 200 Tx, sold by Optos plc. Compared to current imaging techniques, ultra-widefield devices can image up to 200 degrees, thus providing more information to help clinicians accurately diagnose the condition. Additionally, these devices have a shorter image processing time and allow for image

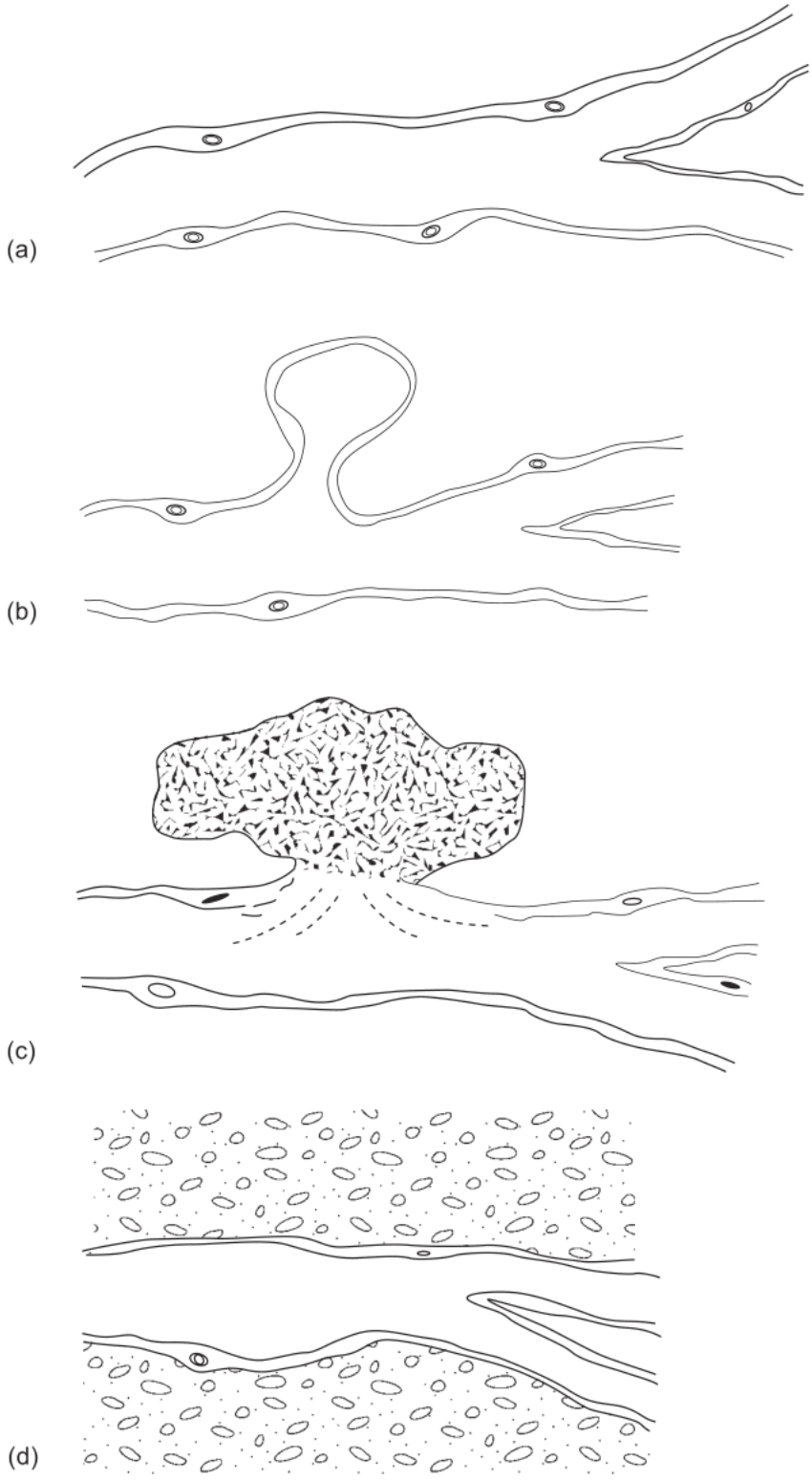


Figure 2 – Representations of various types of damage to capillaries in DR. (a) normal capillary. (b) microaneurysm. (c) hemorrhage. (d) exudate [5].

duplication and manipulation with ease [32].

With a recent rise in machine learning, applications in medical diagnostics have been explored due to their high classification accuracy. Deep learning, a subset of machine learning, uses multiple layers for feature extraction from an image. The input into each layer uses the output from the previous layer thus allowing every layer in the algorithm to independently learn the features. Deep learning models are typically trained using large datasets with an appropriate neural network to learn the features in an image without prior knowledge. The use of deep learning for medical images has been studied by LeCun et al. [14], Liu et al. [15], and Litjens et al. [16]. While most of the research on optical coherence tomography images has been explicitly based on segmenting features such as blood vessels and identifying the stage of diabetic retinopathy through a classifier, new research in the ophthalmologic field has been done by Pratt et al. [17], Sangeethaa et al. [18], and Chandrakumar [19] to detect diabetic retinopathy using the convoluted neural network (CNN). This network consists of three layers: input layer, hidden layer, and output layer. The hidden layer encompasses convolution and pooling layers to reduce computational time. As a part of deep learning, CNNs eliminate the need for manual feature extraction and can be retrained for new classification tasks. While the diagnostic accuracy of detecting diabetic retinopathy from OCT images has proven to be successful through CNN, the accuracy of ultra-widefield fundus images have yet to be fully investigated, especially involving various classes/stages. Therefore, the goals of this study are to develop and utilize CNN to help classify ultra-widefield fundus images into one of the 3 major classes of diabetic retinopathy; evaluate and compare

the accuracy of the classification between different areas and sizes of fundus images; and explore the limitations of using CNN for ultra-widefield images.

II. METHODS

2.1 Image Dataset

The images for this research were received from the Cleveland Clinic. There was a total of 62 images sent, 26 TIF format and 36 JPEG format, each of which represented 62 patients. The images were taken by a technician on an ultra-wide field ophthalmoscope with the resulting dimensions of most images being 3900 × 3071 (JPEG) and the remaining images being 3900 × 3072 (TIF). The horizontal and vertical resolution was 96 dpi (dots per inch). These images were then graded by an ophthalmologist and categorized into 1 of 7 classes based on the Early Treatment Diabetic Retinopathy Study (ETDRS) scale: mild (11 images), moderate (11 images), severe (15 images), very severe (14 images), non-high risk (4 images), high risk (3 images), and advanced (4 images).

2.2 Preparation and Pre-processing of Images

Due to the small number of images and uneven classes, the fundus images were organized into three main classes: mild (mild + moderate images), severe (severe + very severe images), and high risk (non-high risk + high risk + advanced images). This was done to balance the classes and provide enough training data for the machine learning algorithm. It is important to note that if there were “normal” images or images that showed no signs of diabetic retinopathy, then four classes would’ve been created to differentiate between normal, mild, severe, and high-risk diabetic retinopathy. Once the images were organized, using MATLAB, five masks at various diameters following a concentric pattern from the center of the fundus image were created. The masks were created to help isolate the region of interest (ROI) in the

fundus images. After the creation of the mask, desired images with the ROI were cut out and converted to TIF format to prepare for training through the CNN. To begin pre-processing, the desired image dimension (e.g. 32 × 32) was chosen for the input layer. Once the images were inputted, they were shuffled to make MATLAB re-train with new images for each new training session. Then, image augmentation was utilized to randomly rotate and reflect the images to prevent overfitting and memorization of image features due to the small size of the image dataset.

2.3 Neural Network Design

A CNN was developed to classify fundus images into 3 classes. This network consists of a 2D convolution layer, batch normalization layer, rectified linear unit (ReLU) layer, max pooling layer, fully connected layer, softmax layer, and classification layer (Figure 3).

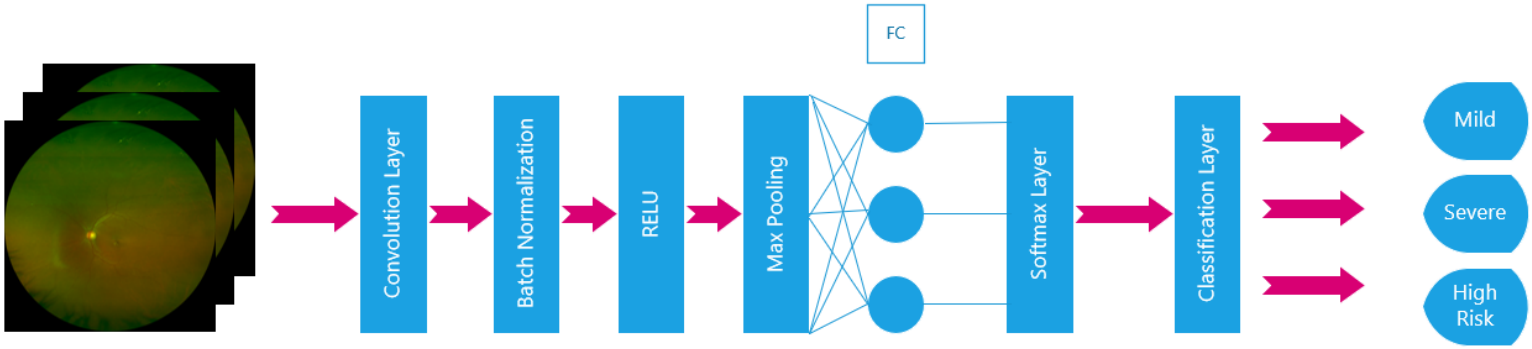


Figure 3 – Neural network schematic.

First, a 2D convolution layer was deployed, consisting of 5 filters with a stride (step size in which the filter moves) of 1, which convolved the input layer by applying sliding filters horizontally and vertically. Given that the input pixel value of each fundus image examined is *x(a,b,c)*, the output of the convolutional layer is *y(w,y,z)* -

which represents the feature maps extracted by convolving *x(a,b,c)* with *f(w,a,b,c)* - is specified as follows:

$$y(w,y,z) = \sum_{a=1}^{5} \sum_{b=1}^{5} \sum_{c=1}^{3} f(w,a,b,c) \cdot x(a+y-1, b+z-1, c) + b(w) \quad (1)$$

w represents the filter index where *w = 1,2,3,4, 5*. Similarly, *m = 1,2,3,4…q* and *n = 1,2,3,4…q* where *q* represents the input dimension of the image. The terms *b(w)* and *f(w,a,b,c)* represent the biases and weights of the filters, respectively [20]. Due to the small dataset and to avoid underfitting, the bias values and weights were kept at default values. The weights were initialized with the Glorot initializer, which is optimal for neural networks with a single hidden layer, with a small Gaussian value with zero mean and variance = 2/(fan-in + fan-out) where fan-in = *filterSize1* × *filterSize2* × *Number of Input Channels* and fan-out = *filterSize1* × *filterSize2* × *Number of Hidden Channels*. The bias values were initialized with zeros since symmetry-breaking is provided through the weights. The number of neurons in a convolution layer is equivalent to the *Map Size* × *Number of Filters*. For example, if the input image is 32-by-32-by-3 the map size would be 15-by-15 since (32-5+(1×1))/2 +1 ≈ 15. The *Map Size* is then multiplied by the *Number of Filters* resulting in the total number of neurons being 15 × 15 × 5 = 1125.

After the output was obtained, a batch normalization layer was included after each convolution layer, normalizing the input and speeding up training while reducing the sensitivity to network initialization. As shown in equation (2), the layer functions by calculating the normalized activations by using the mean μ_B and σ^2_B over the input channel and the mini batch. If the variance is very small, the ε property comes into play by improving the numerical stability [21].

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{2}$$

The batch normalization layer was then followed by the rectified linear unit (ReLU) layer. This layer sets any value less than zero to zero; a type of threshold operation equivalent to equation (3) [22].

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

A max pooling layer was implemented to perform down sampling by separating the input into "rectangular pooling regions and computing the maximum of each region"; specifically, a pooling layer with 2 filters and a stride of 2 [23]. Next, the fully connected layer was deployed to acquire sub-feature maps to help classify them into three classes. The *outputSize* was set to 3 which represented the three classes: mild, severe, and high risk. The fully connected layer functioned by multiplying the input by a weight matrix and then adding a bias to that matrix as shown in equation (4) [24] [25] [26]:

$$f(t) = \sum_{k=1}^{4} \sum_{j=1}^{32} \sum_{i=1}^{32} \bar{w}(l,m,n) \cdot g(l,m,n) + \bar{b}(t) \tag{4}$$

The terms $\bar{w}(l,m,n)$ and $\bar{b}(t)$ represent the weights and bias; *t* is the index where *t* = 1, 2, and 3. As mentioned before, no bias values or weights were specified to avoid underfitting. Following the fully connected layer was the softmax function, which calculates the probability distribution. The softmax function is also known as the normalized exponential function and is defined as such for multi-class classification problems:

$$P(d_r|y,\theta) = \frac{P(y,\theta \mid d_r)P(d_r)}{\sum_{i=1}^{m} P(y,\theta \mid d_j)P(d_j)} = \frac{e^{(b_r(y,\theta))}}{\sum_{i=1}^{m} e^{(b_j(y,\theta))}} \quad (5)$$

The limits for the probability, $P(d_r|y,\theta)$, are from 0 to 1 and the term $\sum_{i=1}^{m} P(d_j \mid y,\theta)$ is equal to 1. Additionally, $b_r = ln(P(y,\theta \mid d_r)P(d_r))$ where $P(y,\theta \mid d_r)$ is the conditional probability of the sample class r and the prior probability $P(d_r)$ [27] . To successfully classify each fundus image, the cross-entropy loss was calculated in the final layer called the classification layer. The cross-entropy loss is defined as such:

$$\text{Loss} = -\sum_{j=1}^{M}\sum_{i=1}^{H} s_{ji} \ln u_{ji} \quad (6)$$

H is the number of classes, *M* is the number of samples, s_{ji} is the indicator that correlates the *j*th sample to the *i*th class, and u_{ji} is the output of the softmax function, associating the *j*th input with the *i*th class [27].

2.4 Simulation Options

Once the neural network was designed, the simulation options were specified. The learn rate of 0.01 was specified along with the maximum number of epochs of 100 and a mini-batch size of 48. For network training, 52 randomly selected images were used. The remaining images (10) were used for classification. The command *shuffle every-epoch* was used to shuffle the training data before each epoch and also to take advantage of the batch normalization effect, thus making the training of the

neural network an effortless optimization problem. The algorithm used for training is the stochastic gradient descent with momentum (SGDM), which is defined as

$$\theta_{n+1} = \theta n - \beta \nabla F(\theta_n) + \eta(\theta_n - \theta_{n-1}) \quad (7)$$

where θ is the parameter vector n is the iteration number, β is the learn rate (β>0), $\eta(\theta_n - \theta_{n-1})$ is the momentum term/parameter, and $F(\theta)$ is the loss function. This SGDM algorithm minimizes the cross-entropy loss across the mini-batch by proceeding incrementally at each iteration in the direction of the gradient loss. Adding the momentum term, value of 0.9, helps reduce oscillations along the path of steep descent toward the optimum value [28].

III. DISCUSSION

3.1 Results

The neural network was simulated for a total of 25 times with 5 trials for each image size: 32×32, 64×64, 128×128, 256×256, and 512×512. The average simulation/processing time ranged from 25 seconds for 32×32 images, to 22 minutes for 512×512 images. A confusion matrix was created for each of the image sizes to obtain statistical measures of performance, such as sensitivity, false negative rate, precision, and the false discovery rate. The sensitivity was 70% for 32×32 images, 65% for 64×64 images, 60% for 128×128 images, 50% for 256×256 images and 50% for 512×512 images. The sensitivity decreased as the size of the images increased; the opposite is true for the false negative rate. Similarly, the precision was 65% for 32×32 images, 65% for 64×64 images, 60% for 128×128 images, 50% for 256×256 images, and 50% for 512×512 images. The precision decreased as the size of the images increased, while the false discovery rate increased as the size of the images increased (Appendix II). These statistical measures suggest that a combination of a small image dataset and uneven classes may contribute to an increase in type I and type II errors thus leading to irrelevant results. The classification accuracy of the training and the test sets was determined by adding the number of correctly classified images divided by the total number of classified images; 52 images total for training and 10 images total for testing. As shown in Table 1, the training accuracy for the 32×32, 64×64, and 128×128 images ranged from 92% to 97%. For both the 256×256 and 512×512 images, the training accuracy was 46%. The training accuracy represents the accuracy of the neural network model on the image sets on which it was constructed. While within 100 epochs, most of the images up to size of

128×128 approached >90% training accuracy, the images of sizes 256×256 and 512×512 had their respective training accuracy plateau at 46% with the 45th iteration (or epoch).

Table 1 – Average classification accuracy of training and test images for 5 trials

Image Sizes	Avg. Training Accuracy (%)	Avg. Test Accuracy (%)
32×32	92.692	50.000
64×64	97.307	48.000
128×128	96.923	52.000
256×256	46.153	50.000
512×512	46.538	50.000

The plateauing of the accuracy is significant because it means that the algorithm has hit convergence, where further training is likely to degrade the test accuracy and the training accuracy and result in overfitting. The test accuracy is the accuracy of the images the model has yet to see. Although the test accuracy of all the images was around 50%, an important observation to note is that the test accuracy did not improve as the image size increased, as one would intuitively believe since larger images have more pixel data. This shows, as is commonly known in the ophthalmologic field, that the majority (~90%) of cases, the information for proper diagnosis is around the optic nerve and macula region of the eye. Moreover, this suggests that different ROIs do not impact the diagnosis of DR.

3.2 Validation

While the accuracy results were not ideal, they provide some insight into why they are low. Noticing the discrepancy between training accuracy and test accuracy, one would think that overfitting might be the reason. So, to test this hypothesis and to verify

the results, two methods were used: image normalization and transfer learning.

Image normalization is the process of changing the histogram values of an image to bring it to a range of intensity values that is “normal” (normal distribution). This avoids the influence of very low and very high noise in the image. In machine learning, image normalization is sometimes used to make learning more stable and to make the neural network generalize favorably to unseen data by reducing the variability between the training set and the testing set [18]. For this simulation, all images were normalized by taking each individual color channel and dividing them by the magnitude of all three (RGB) color channels (Figure 4).

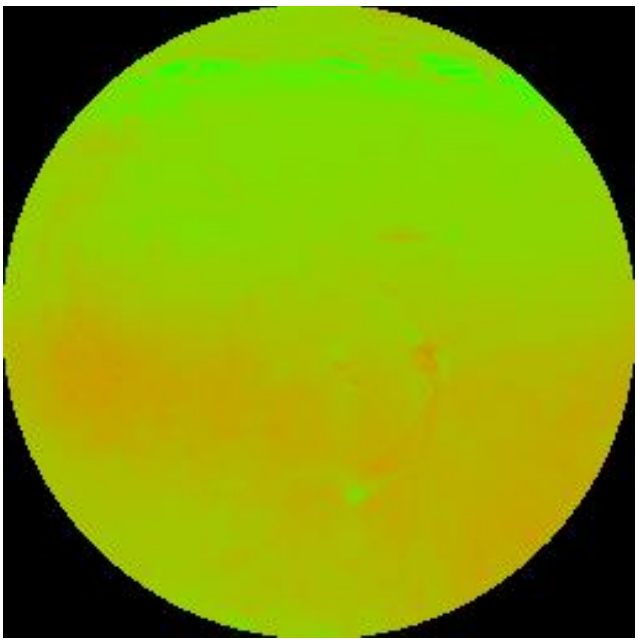


Figure 4 – Normalized fundus image.

After normalization, the images were trained and tested using the CNN algorithm and produced the following results (Table 2):

Table 2 – Average classification accuracy of normalized images for 5 trials

Image Sizes	Avg. Training Accuracy (%)	Avg. Test Accuracy (%)
32×32	88.461	50.000
64×64	96.153	50.000
128×128	94.423	50.000
256×256	46.153	50.000
512×512	46.153	50.000

The results show a similar pattern to the original results where the training accuracy is relatively high for image sizes of 32×32, 64×64, and 128×128 followed by a steep decrease to 46% for image sizes of 256×256 and 512×512. The test accuracy did not show a significant difference from the original results, staying around 50%. Two things that can be concluded from this method is that overfitting is not a cause for the discrepancy in the accuracy values and that the varied levels of lighting in the original images, which could possibly affect pixel intensity values and create unnecessary variation, did not impact the test accuracy.

Transfer learning utilizes a model that has been trained on a larger dataset which can be fine-tuned for a new task or new dataset; especially small datasets (<1000 images). Compared to the performance of a model that is trained from scratch, it achieves higher network performance in a short amount of time. The pre-trained model used for transfer learning for this study is called AlexNet. This model was developed by Alex Krizhevsky utilizing the CNN algorithm. It has been trained on 1.2 million images from the ImageNet Dataset (**http://image-net.org/index**) with image dimensions of 227×227. The structure of the model consists of 5 convolution layers which together with the

remaining layers add to a total of 23 [29]. Once the model is selected, the last classification layer is replaced for a new classification task. Since an image size of 227×227 is required as the input, fundus images with dimensions of 256×256 and 512×512 were used by utilizing the function *augmentedImageDatastore* to automatically resize the images with no loss of data. Then, the model was simulated, and the training accuracy and test accuracy were obtained. Similar to the results of the normalized images and the original results, after 5 trials the average test accuracy was found to be 50%. On the other hand, the training accuracy was 100%. The high training accuracy is most likely due to overfitting since the transfer learning model did not include an image augmentation step to account for overfitting. While the use of transfer learning to improve classification of DR has been done using AlexNet [30], in this case transfer learning showed no significant improvement compared to the original method that was done from scratch. This strongly suggests that any additional convolution layers are unnecessary since the algorithm has already extracted all the features with one convolution layer for successful classification.

3.3 Limitations

The reason that neither the original method nor the validation methods were able to achieve greater than 50% accuracy can be attributed to one of many reasons which are also the limitations of this study. One limitation, perhaps the most impactful limitation, was the small amount of data available. While researchers have been able to use a small dataset to detect DR from 140 OCT images with a classification accuracy of 93%, no current literature exists demonstrating this with ultra-widefield fundus images. One could argue that with a more complex algorithm, i.e. with additional convolution layers, one

could achieve better performance [18]. However, as stated previously that even with additional layers, the algorithm showed no improvement in performance.

Another limitation was the unbalanced image classes caused by a lack of diseased images. Initially, the images were rearranged into four classes where the classification accuracy (test accuracy) was found to be 35%. With the images rearranged into 2 classes, the classification accuracy was 60% (see Appendix III for probability values). Although decreasing the number of classes does increase accuracy, it is not beneficial for doctors since they would need to further examine the fundus images to determine the best treatment plan. Other ways to tackle the unbalanced image classes were to balance the classes by removing images from larger classes or by adjusting weights and bias components in the algorithm. By having an equal number of images in each of the three classes, the training accuracy decreased to around 80% for most of the image sizes and the test accuracy decreased for three of the image sizes ranging from 28% to 39% (Appendix I). There were additional trials done by adjusting the weights to “zeros” and changing the bias to “ones”. The average classification accuracy for all images was about 23% thus severely underfitting and resulting in high training loss; neither able to model the training images nor able to generalize the test images.

A third limitation in this study was the presence of unnecessary features such as eyelashes (Figure 5) and camera artifacts (Figure 6) in some of the images. The presence of eyelashes caused the CNN algorithm to learn the unnecessary features and thus deteriorate the classification accuracy, especially for larger image sizes (256×256 and 512×512).

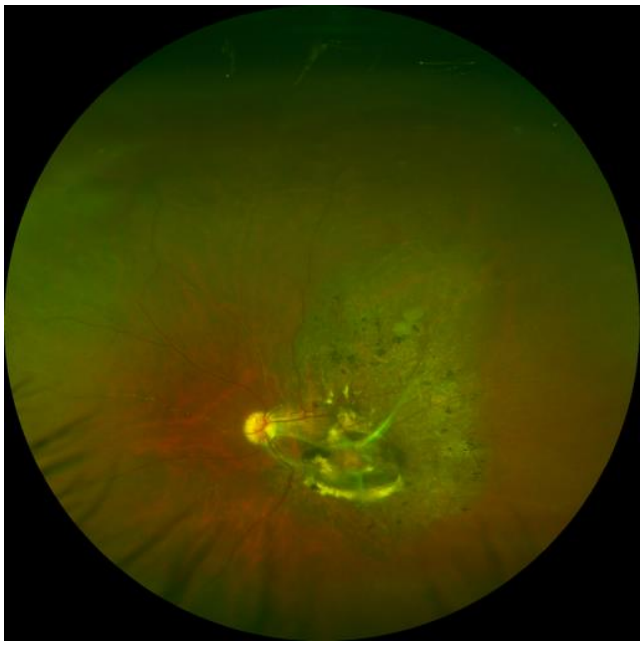


Figure 5 – Presence of eyelashes towards the bottom of the fundus image.

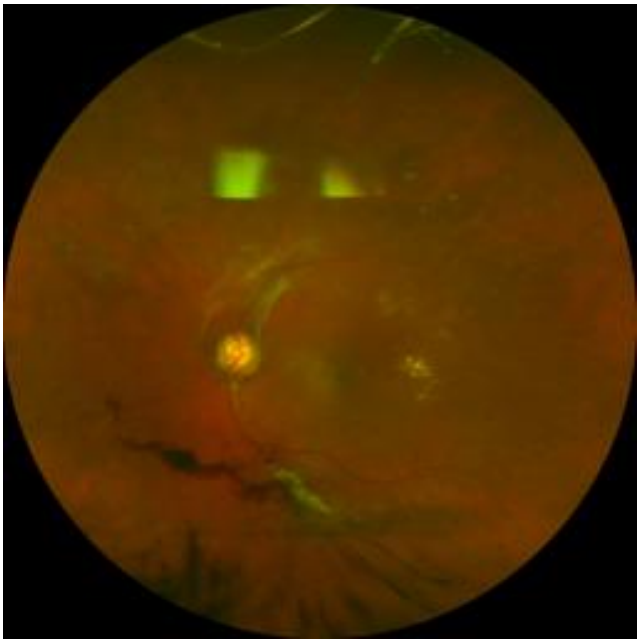


Figure 6 – Presence of camera artifacts (i.e. camera lens reflection) towards the top of the fundus image

IV. CONCLUSION

The goals of this study were to develop an algorithm to classify ultra-widefield fundus images into one of the 3 major classes of diabetic retinopathy, evaluate and compare the accuracy of the classification between different areas and sizes of fundus images, and explore the limitations of using CNN for ultra-widefield images. Validation methods such as image normalization and transfer learning were also done to calculate the classification accuracy and justify the algorithm that was developed from scratch. While the results are not ideal, they demonstrate the potential for applying deep learning algorithms to ultra-widefield images and the need for doctors to further examine the diagnosis to account for false positives and/or misdiagnosis.

Future work in this subject would include improvements to the CNN algorithm as more images are obtained in order to improve the classification accuracy. Additional research would be done to determine if the file type of an image (e.g. JPEG or TIF) impacts the classification accuracy. Also, instead of concentric circles which mainly focus on the optic nerve, different sections of an image may be isolated/cut out where the accuracy of different regions of the eye may be compared to see if the classification accuracy solely relies on the optic nerve. Lastly, some pre-processing would be done to remove artifacts such as eyelashes from the images to reduce the number of features to be learned by the CNN algorithm; only if the flexibility is available to do so i.e. large amount of data available where removal of some images will not drastically impact the classification accuracy.

APPENDIX I – Classification Data

A. Balanced classes

Trials	Image size	Training Accuracy	Test Accuracy
1	32*32	0.5439	0.4000
2		0.6667	0.6000
3		0.5995	0.6000
4		0.5789	0.6000
5		0.6140	0.4000
	Avg. (%)	60.06	52.000
1	64*64	0.8421	0.4000
2		0.6140	0.4000
3		0.7719	0.2000
4		0.6140	0.4000
5		0.8421	0.4000
	Avg. (%)	73.682	36.000
1	128*128	1	0.2759
2		0.7018	0.2000
3		0.7719	0.6000
4		0.7576	0.4000
5		0.9394	0.4828
	Avg. (%)	83.414	39.174
1	256*256	0.8485	0.3103
2		0.6061	0.3793
3		0.9394	0.1724
4		0.6970	0.2069
5		0.9394	0.3448
	Avg. (%)	80.608	28.274
1	512*512	0.8788	0.2414
2		0.9091	0.3103
3		0.8596	0.6000
4		0.6667	0.8000
5		0.8741	0.6333
	Avg. (%)	83.766	51.7

B. Unbalanced classes

Trials	Image size	Training Accuracy	Test Accuracy
1	32*32	0.9038	0.6000
2		0.9615	0.7000
3		0.9038	0.4000
4		0.9423	0.3000
5		0.9230	0.5000
	Avg. (%)	92.692	50.000
1	64*64	1.0000	0.4000
2		0.9807	0.5000
3		0.9807	0.6000
4		0.9615	0.5000
5		0.9423	0.4000
	Avg. (%)	97.308	48.000
1	128*128	0.9038	0.6000
2		0.9615	0.5000
3		1.0000	0.4000
4		0.9807	0.5000
5		1.0000	0.6000
	Avg. (%)	96.923	52.000
1	256*256	0.4615	0.5000
2		0.4615	0.5000
3		0.4615	0.5000
4		0.4615	0.5000
5		0.4615	0.5000
	Avg. (%)	46.153	50.000
1	512*512	0.4807	0.5000
2		0.4615	0.5000
3		0.4615	0.5000
4		0.4615	0.5000
5		0.4615	0.5000
	Avg. (%)	46.538	50.000

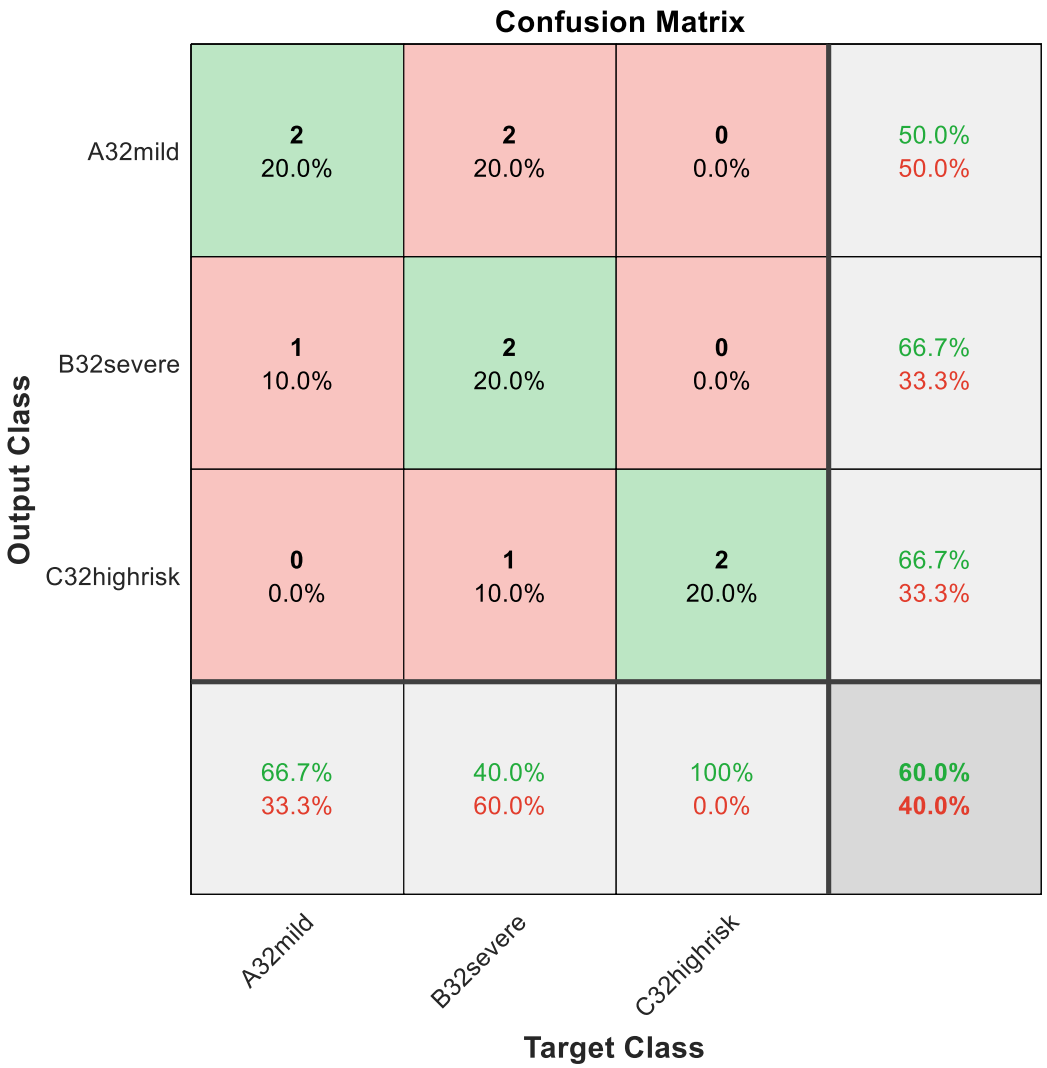
APPENDIX II – Confusion Matrices

Far right column = precision, false discovery rate

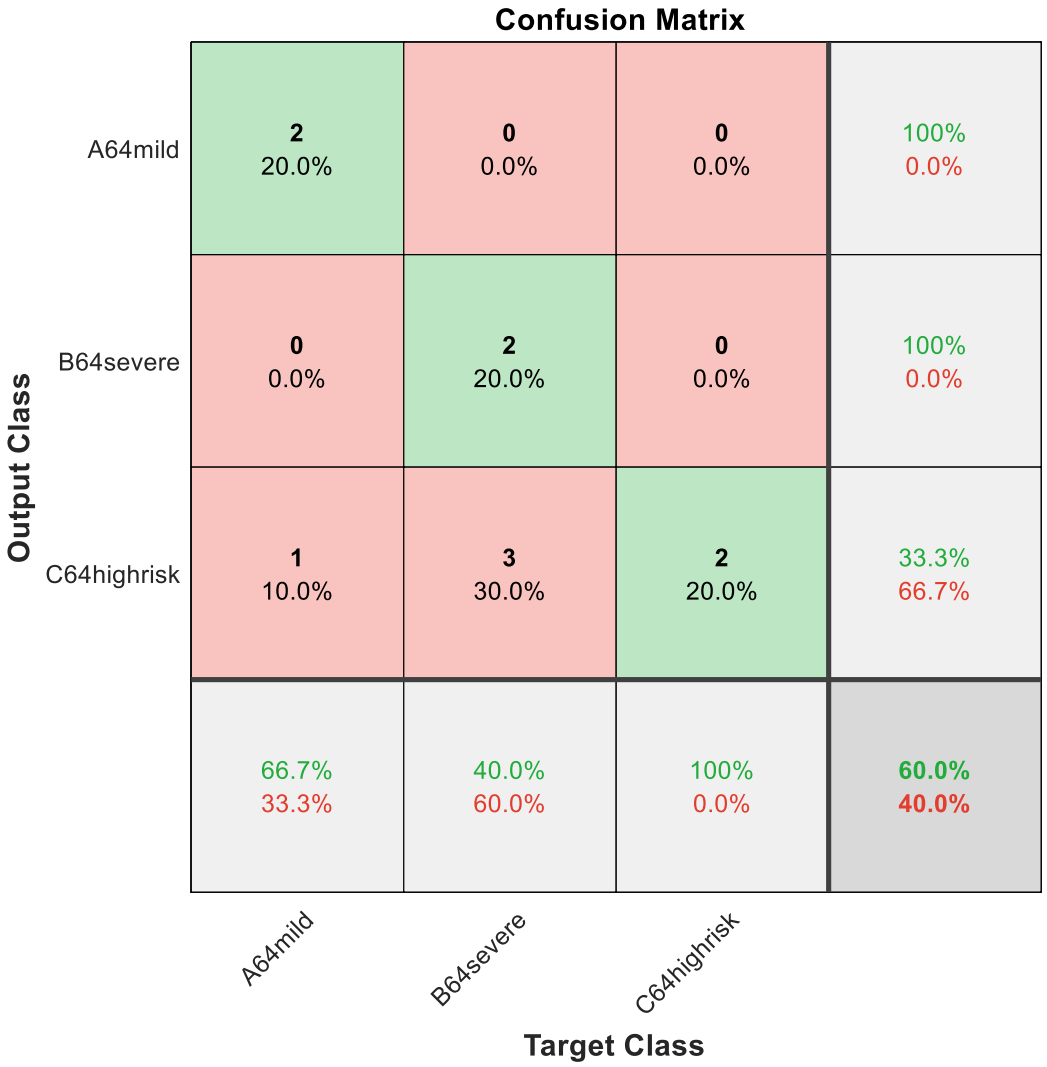
Bottom row = sensitivity, miss rate

Bottom right corner = overall accuracy of current trial, overall inaccuracy of current trial

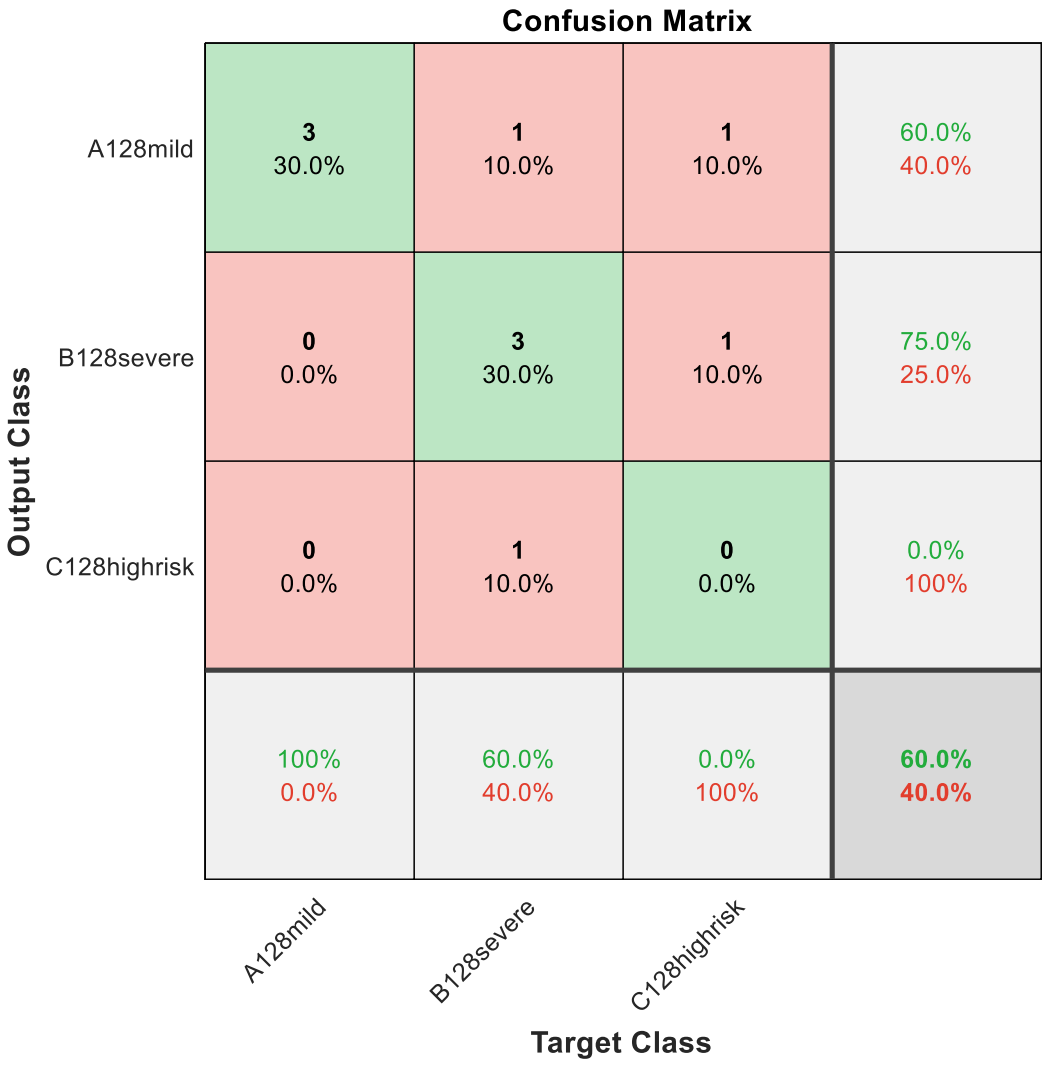
32×32



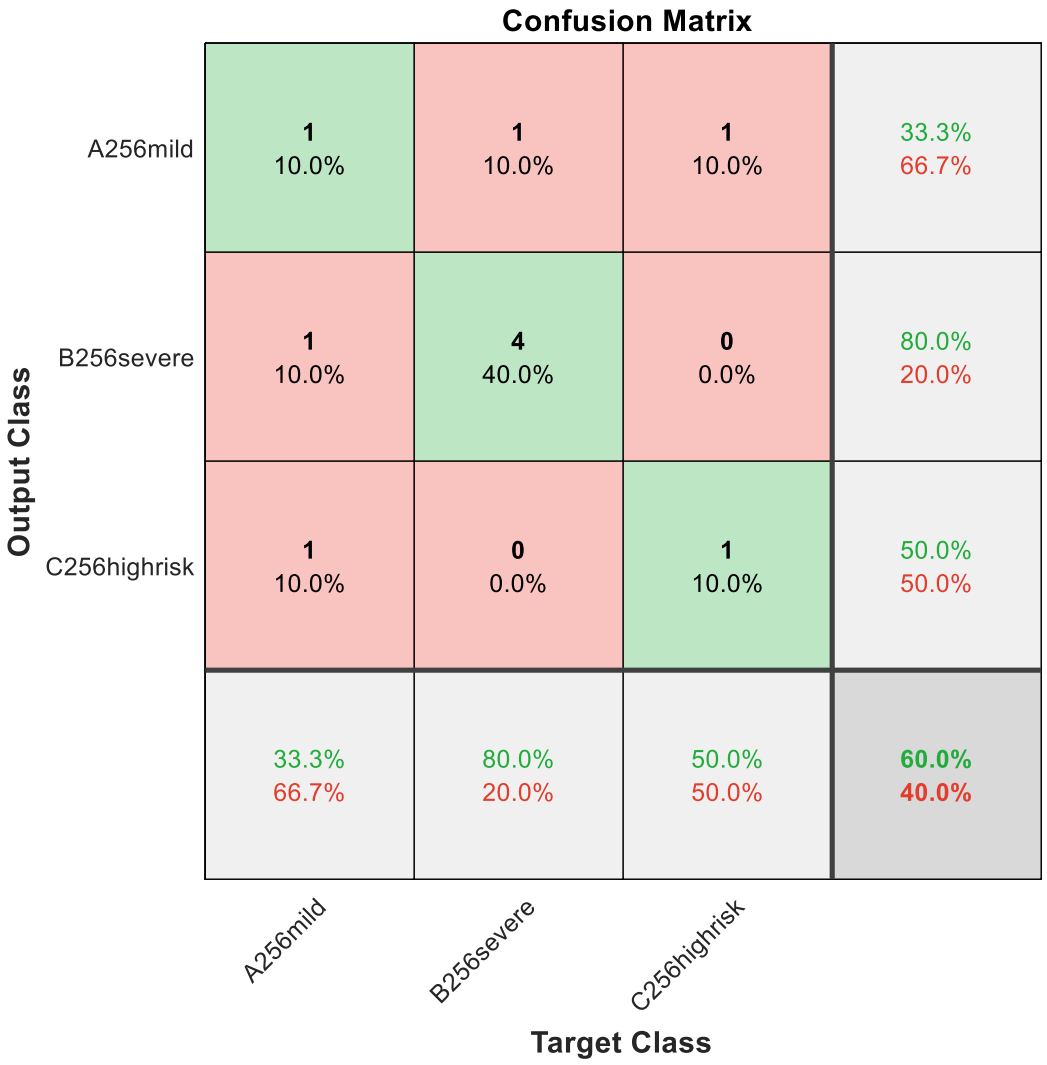
64×64


Confusion Matrix
Output Class
A64mild
B64severe
C64highrisk
2
20.0%
0
0.0%
0
0.0%
100%
0.0%
0
0.0%
2
20.0%
0
0.0%
100%
0.0%
1
10.0%
3
30.0%
2
20.0%
33.3%
66.7%
66.7%
33.3%
40.0%
60.0%
100%
0.0%
60.0%
40.0%
A64mild
B64severe
C64highrisk
Target Class

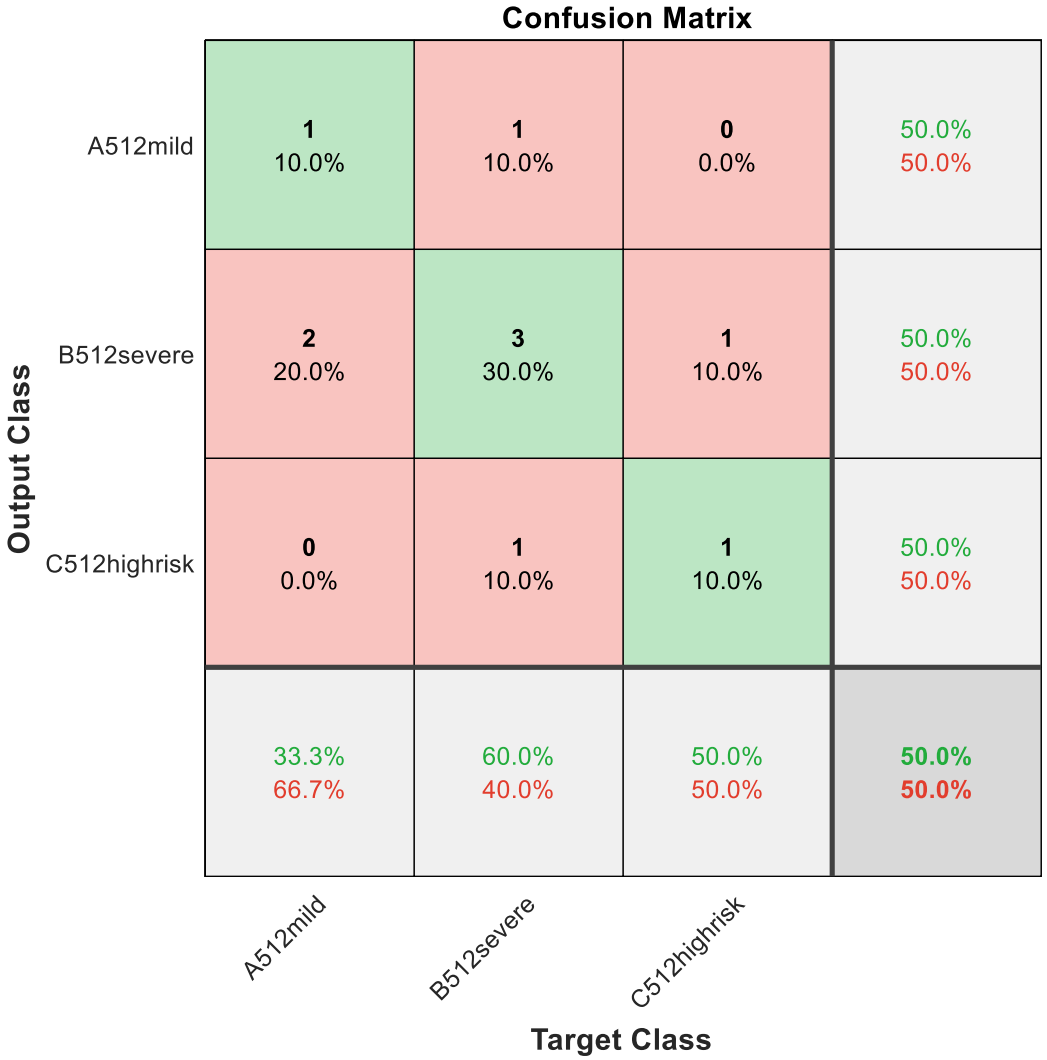
128×128


Confusion Matrix
A128mild
B128severe
C128highrisk
Output Class
3
30.0%
1
10.0%
1
10.0%
60.0%
40.0%
0
0.0%
3
30.0%
1
10.0%
75.0%
25.0%
0
0.0%
1
10.0%
0
0.0%
0.0%
100%
100%
0.0%
60.0%
40.0%
0.0%
100%
60.0%
40.0%
A128mild
B128severe
C128highrisk
Target Class

256×256


Confusion Matrix
A256mild
B256severe
C256highrisk
Output Class
1
10.0%
1
10.0%
1
10.0%
33.3%
66.7%
1
10.0%
4
40.0%
0
0.0%
80.0%
20.0%
1
10.0%
0
0.0%
1
10.0%
50.0%
50.0%
33.3%
66.7%
80.0%
20.0%
50.0%
50.0%
60.0%
40.0%
A256mild
B256severe
C256highrisk
Target Class

512×512


Confusion Matrix
A512mild
B512severe
C512highrisk
Output Class
1
10.0%
1
10.0%
0
0.0%
50.0%
50.0%
2
20.0%
3
30.0%
1
10.0%
50.0%
50.0%
0
0.0%
1
10.0%
1
10.0%
50.0%
50.0%
33.3%
66.7%
60.0%
40.0%
50.0%
50.0%
50.0%
50.0%
A512mild
B512severe
C512highrisk
Target Class

APPENDIX III – Probability Data

Red = incorrectly predicted

32×32

Patient	Type	P (Abnormal) [32*32]	P(Normal) [32*32]
10001418 OS	abnormal	0.99998534	1.47E-05
10022953 OS	abnormal	2.43E-07	0.99999976
10029705-20180416@112408-L4	abnormal	0.99998879	1.12E-05
10029705-20180416@112408-R3	abnormal	0.13589142	0.86410856
10041101 OD	abnormal	0.005335684	0.99466431
10041101 OS	abnormal	3.72E-08	1
10046318 OS	abnormal	0.14558524	0.85441476
10055620 OD	abnormal	0.99221849	0.007781468
10055620 OS	abnormal	0.00250658	0.99749339
10065949 OS	abnormal	0.0087861	0.99121398
10068165 OD	abnormal	0.040095784	0.95990425
10075046 OD	abnormal	0.80176336	0.19823663
10075046 OS	abnormal	0.80258894	0.19741113
10077406 OD	abnormal	0.66958404	0.33041599
10077406 OS	abnormal	0.99926466	0.000735337
10084968 OS	abnormal	0.16454743	0.83545256
10086708 OS	abnormal	6.59E-10	1
10117924-20180220@102014-L4	abnormal	0.79679787	0.20320213
10145754-20180415@105849-L2	abnormal	0.99968874	0.000311262
10145754-20180415@105849-R1	abnormal	0.99999964	3.16E-07
10148705-20180425@100700-R3	abnormal	0.9999944	5.65E-06
10149811-20180225@090046-R1	abnormal	0.99999905	9.38E-07
10158969-20180506@090506-R1	abnormal	0.99999976	2.90E-07
10001091 OD	abnormal	0.38385868	0.61614132
10001091 OS	abnormal	0.23049685	0.76950318
10000464 OS	mild/moderate	2.74E-09	1
10000484 OD	mild/moderate	0.9986406	0.001359399
10001418 OD	mild/moderate	0.99864513	0.001354832
10004898 OS	mild/moderate	4.69E-15	1
10008285 OS	mild/moderate	0.73723733	0.26276267
10013404 OD	mild/moderate	0.96721429	0.032785717
10018169 OD	mild/moderate	0.96985573	0.03014431
10022953 OD	mild/moderate	0.026244124	0.9737559
10022979-20180213@155058-R1	mild/moderate	0.99999571	4.26E-06
10025329 OD	mild/moderate	0.99999142	8.55E-06

10028832 OD	mild/moderate	0.99961448	0.000385448
10028832 OS	mild/moderate	3.78E-06	0.99999619
10037058 OD	mild/moderate	0.44739011	0.55260986
10037058 OS	mild/moderate	0.1597831	0.84021693
10041971 OD	mild/moderate	0.048188951	0.95181108
10041971 OS	mild/moderate	5.42E-10	1
10047851	mild/moderate	0.5488714	0.45112863
10047851 OD	mild/moderate	0.99706465	0.002935403
10051334 OD	mild/moderate	8.39E-06	0.99999166
10051334 OS	mild/moderate	0.018040443	0.98195964
10059609-20180325@094033-L3	mild/moderate	0.090828925	0.9091711
10065949 OD	mild/moderate	0.99982738	0.000172549
10068165 OS	mild/moderate	1.01E-05	0.99998987
10072783-20180121@092503-L2	mild/moderate	0.97144598	0.028554037
10075925 OD	mild/moderate	0.99999118	8.78E-06
10081192 OS	mild/moderate	8.48E-06	0.99999154
10084968 OD	mild/moderate	0.01174185	0.98825818
10086708 OD	mild/moderate	0.000257731	0.99974221
10088791 OD	mild/moderate	1	2.72E-23
10088791 OS	mild/moderate	1.82E-05	0.99998188
10093928 OS	mild/moderate	1.98E-07	0.99999976
10117924-20180220@102014-R2	mild/moderate	0.99999177	8.24E-06
10131726-20180319@085722-L3	mild/moderate	0.007010281	0.99298966
10131726-20180319@085722-R2	mild/moderate	0.99999964	3.62E-07
10158969-20180506@090506-L2	mild/moderate	0.99999547	4.52E-06
10166525-20180402@083736-L4	mild/moderate	0.8872788	0.11272123
10167812-20180424@134719-L2	mild/moderate	0.05154828	0.94845176
# of Incorrect Classification		16	12

64×64

Patient	Type	P (Abnormal) [64*64]	P(Normal) [64*64]
10001418 OS	abnormal	1	1.64E-09
10022953 OS	abnormal	1	8.43E-10
10029705-20180416@112408-L4	abnormal	1	8.60E-09
10029705-20180416@112408-R3	abnormal	0.99998713	1.29E-05
10041101 OD	abnormal	1	8.74E-11
10041101 OS	abnormal	1	1.68E-32
10046318 OS	abnormal	1	1.55E-21
10055620 OD	abnormal	1	1.93E-20
10055620 OS	abnormal	1	2.76E-22
10065949 OS	abnormal	0.99999964	3.26E-07

10068165 OD	abnormal	0.99998391	1.62E-05
10075046 OD	abnormal	0.99999857	1.49E-06
10075046 OS	abnormal	1	4.30E-18
10077406 OD	abnormal	1	1.64E-12
10077406 OS	abnormal	0.99997497	2.51E-05
10084968 OS	abnormal	1	2.28E-14
10086708 OS	abnormal	1	2.85E-15
10117924-20180220@102014-L4	abnormal	0.99986422	1.36E-04
10145754-20180415@105849-L2	abnormal	0.99998403	1.60E-05
10145754-20180415@105849-R1	abnormal	1	8.69E-10
10148705-20180425@100700-R3	abnormal	1.85839E-15	1.00E+00
10149811-20180225@090046-R1	abnormal	1	6.47E-14
10158969-20180506@090506-R1	abnormal	0.99999356	6.40E-06
10001091 OD	abnormal	1	8.70E-14
10001091 OS	abnormal	1	4.15E-21
10000464 OS	mild/moderate	1.58362E-24	1.00E+00
10000484 OD	mild/moderate	1.956E-08	1.00E+00
10001418 OD	mild/moderate	5.52701E-09	1.00E+00
10004898 OS	mild/moderate	1.45156E-40	1.00E+00
10008285 OS	mild/moderate	0.06106123	0.9389388
10013404 OD	mild/moderate	1.74044E-10	1
10018169 OD	mild/moderate	2.02048E-09	1
10022953 OD	mild/moderate	1.57295E-14	1
10022979-20180213@155058-R1	mild/moderate	1.11E-06	0.99999893
10025329 OD	mild/moderate	4.45E-11	1
10028832 OD	mild/moderate	1.11E-06	0.99999893
10028832 OS	mild/moderate	2.92E-06	0.99999714
10037058 OD	mild/moderate	1.01E-12	1
10037058 OS	mild/moderate	1.00E+00	1.16596E-11
10041971 OD	mild/moderate	9.82E-01	0.017991316
10041971 OS	mild/moderate	1.79E-08	1
10047851	mild/moderate	2.03E-10	1
10047851 OD	mild/moderate	4.06E-09	1
10051334 OD	mild/moderate	1.91E-23	1
10051334 OS	mild/moderate	3.85E-06	0.99999619
10059609-20180325@094033-L3	mild/moderate	1.00E+00	3.93148E-06
10065949 OD	mild/moderate	8.95E-06	0.99999106
10068165 OS	mild/moderate	1.65E-05	0.99998355
10072783-20180121@092503-L2	mild/moderate	1.07E-05	0.99998927
10075925 OD	mild/moderate	1.13E-08	1
10081192 OS	mild/moderate	2.30E-11	1
10084968 OD	mild/moderate	4.49E-13	1
10086708 OD	mild/moderate	1.62E-11	1

10088791 OD	mild/moderate	2.49E-05	0.99997509
10088791 OS	mild/moderate	0.00E+00	1
10093928 OS	mild/moderate	4.34E-07	0.99999952
10117924-20180220@102014-R2	mild/moderate	5.35E-11	1
10131726-20180319@085722-L3	mild/moderate	4.32E-07	0.99999952
10131726-20180319@085722-R2	mild/moderate	2.71E-11	1
10158969-20180506@090506-L2	mild/moderate	2.51E-05	0.99997485
10166525-20180402@083736-L4	mild/moderate	1.07E-05	0.99998927
10167812-20180424@134719-L2	mild/moderate	1.01E-07	0.99999988
# of Incorrect Classification		2.00E+00	0

128×128

Patient	Type	P (Abnormal) [128*128]	P(Normal) [128*128]
10001418 OS	abnormal	0.99999952	4.87E-07
10022953 OS	abnormal	0.99995208	4.79E-05
10029705-20180416@112408-L4	abnormal	0.99989486	1.05E-04
10029705-20180416@112408-R3	abnormal	0.99989188	0.000108118
10041101 OD	abnormal	0.99984419	0.000155748
10041101 OS	abnormal	1	6.96896E-11
10046318 OS	abnormal	0.99984348	1.56E-04
10055620 OD	abnormal	1	8.19795E-10
10055620 OS	abnormal	0.9999969	3.10E-06
10065949 OS	abnormal	0.99955744	0.000442587
10068165 OD	abnormal	0.99990392	9.60951E-05
10075046 OD	abnormal	0.99993837	6.16E-05
10075046 OS	abnormal	0.99998069	1.92567E-05
10077406 OD	abnormal	0.9999969	3.14E-06
10077406 OS	abnormal	0.99981743	0.000182557
10084968 OS	abnormal	0.99996579	3.43E-05
10086708 OS	abnormal	0.99994266	5.73716E-05
10117924-20180220@102014-L4	abnormal	0.99999535	4.61176E-06
10145754-20180415@105849-L2	abnormal	0.9999938	6.17E-06
10145754-20180415@105849-R1	abnormal	0.9999516	4.84E-05
10148705-20180425@100700-R3	abnormal	0.9998908	0.000109201
10149811-20180225@090046-R1	abnormal	0.99998224	1.77666E-05
10158969-20180506@090506-R1	abnormal	0.99998629	1.37E-05
10001091 OD	abnormal	0.99988723	1.13E-04
10001091 OS	abnormal	0.99999547	4.55215E-06
10000464 OS	mild/moderate	4.67553E-06	0.99999535
10000484 OD	mild/moderate	0.000164147	1.00E+00

10001418 OD	mild/moderate	1.82634E-05	1.00E+00
10004898 OS	mild/moderate	3.41719E-12	1.00E+00
10008285 OS	mild/moderate	1.72229E-06	0.99999833
10013404 OD	mild/moderate	5.08921E-05	0.9999491
10018169 OD	mild/moderate	0.000141378	0.99985862
10022953 OD	mild/moderate	6.75234E-05	0.99993253
10022979-20180213@155058-R1	mild/moderate	1.48E-04	0.99985194
10025329 OD	mild/moderate	1.44E-01	0.8564226
10028832 OD	mild/moderate	8.03106E-06	0.99999201
10028832 OS	mild/moderate	7.74E-02	0.92256796
10037058 OD	mild/moderate	2.03E-08	1
10037058 OS	mild/moderate	0.000249152	0.99975079
10041971 OD	mild/moderate	4.99E-05	0.99995017
10041971 OS	mild/moderate	8.07E-05	0.9999193
10047851	mild/moderate	0.99851722	0.001482814
10047851 OD	mild/moderate	1.63E-04	0.99983692
10051334 OD	mild/moderate	4.39E-05	0.99995601
10051334 OS	mild/moderate	0.000212336	0.99978763
10059609-20180325@094033-L3	mild/moderate	3.41E-05	0.99996591
10065949 OD	mild/moderate	1.00E+00	5.6119E-05
10068165 OS	mild/moderate	0.88677871	0.11322127
10072783-20180121@092503-L2	mild/moderate	0.000187775	0.9998123
10075925 OD	mild/moderate	0.000161167	0.99983883
10081192 OS	mild/moderate	1.24959E-06	0.99999881
10084968 OD	mild/moderate	1.11E-05	0.99998891
10086708 OD	mild/moderate	6.08E-07	0.9999994
10088791 OD	mild/moderate	1.00E+00	5.53622E-09
10088791 OS	mild/moderate	2.88794E-05	0.99997115
10093928 OS	mild/moderate	1.25E-04	0.99987519
10117924-20180220@102014-R2	mild/moderate	0.22493188	0.7750681
10131726-20180319@085722-L3	mild/moderate	1.88E-04	0.99981159
10131726-20180319@085722-R2	mild/moderate	2.67E-10	1
10158969-20180506@090506-L2	mild/moderate	0.9999944	5.60941E-06
10166525-20180402@083736-L4	mild/moderate	0.99999702	3.02197E-06
10167812-20180424@134719-L2	mild/moderate	9.99E-01	0.000615534
# of Incorrect Classification		7	0

256×256

Patient	Type	P (Abnormal) [256*256]	P(Normal) [256*256]

10001418 OS	abnormal	1	9.18E-11
10022953 OS	abnormal	0.99997783	2.22E-05
10029705-20180416@112408-L4	abnormal	0.99998868	1.14E-05
10029705-20180416@112408-R3	abnormal	0.99999261	7.3702E-06
10041101 OD	abnormal	0.99998462	1.53752E-05
10041101 OS	abnormal	1	1.09164E-08
10046318 OS	abnormal	0.99999988	9.78637E-08
10055620 OD	abnormal	1	2.52E-11
10055620 OS	abnormal	0.99998379	1.63E-05
10065949 OS	abnormal	0.99996638	3.36E-05
10068165 OD	abnormal	0.99999881	1.25081E-06
10075046 OD	abnormal	0.99996507	3.49861E-05
10075046 OS	abnormal	0.99999797	2.01E-06
10077406 OD	abnormal	0.99999964	3.7115E-07
10077406 OS	abnormal	0.99999583	4.12848E-06
10084968 OS	abnormal	0.99999988	1.14369E-07
10086708 OS	abnormal	1	3.93208E-10
10117924-20180220@102014-L4	abnormal	1	3.17E-13
10145754-20180415@105849-L2	abnormal	1	2.96E-08
10145754-20180415@105849-R1	abnormal	1	4.2894E-15
10148705-20180425@100700-R3	abnormal	0.99994671	5.32E-05
10149811-20180225@090046-R1	abnormal	0.99999869	1.30E-06
10158969-20180506@090506-R1	abnormal	0.9999795	2.05E-05
10001091 OD	abnormal	0.99997497	2.51E-05
10001091 OS	abnormal	1	9.43E-15
10000464 OS	mild/moderate	1.80009E-07	1.00E+00
10000484 OD	mild/moderate	0.9988721	0.001127839
10001418 OD	mild/moderate	9.05139E-09	1.00E+00
10004898 OS	mild/moderate	5.21877E-11	1.00E+00
10008285 OS	mild/moderate	0.58756328	0.41243672
10013404 OD	mild/moderate	9.22898E-05	0.99990773
10018169 OD	mild/moderate	0.99998701	1.29961E-05
10022953 OD	mild/moderate	1.32348E-05	0.99998677
10022979-20180213@155058-R1	mild/moderate	2.89E-05	0.99997103
10025329 OD	mild/moderate	5.43E-09	1
10028832 OD	mild/moderate	0.028293183	0.97170681
10028832 OS	mild/moderate	9.00074E-09	1
10037058 OD	mild/moderate	3.18E-14	1
10037058 OS	mild/moderate	6.41E-15	1
10041971 OD	mild/moderate	7.89E-06	0.99999213
10041971 OS	mild/moderate	2.33375E-07	0.99999976
10047851	mild/moderate	7.12E-03	0.99288416
10047851 OD	mild/moderate	3.11E-05	0.99996889

10051334 OD	mild/moderate	8.69E-06	0.9999913
10051334 OS	mild/moderate	2.20107E-09	1
10059609-20180325@094033-L3	mild/moderate	1.88E-10	1
10065949 OD	mild/moderate	4.30446E-05	0.99995697
10068165 OS	mild/moderate	0.99974483	0.000255199
10072783-20180121@092503-L2	mild/moderate	4.13E-05	0.99995875
10075925 OD	mild/moderate	9.99E-01	0.000701777
10081192 OS	mild/moderate	3.14926E-05	0.99996853
10084968 OD	mild/moderate	3.49794E-13	1.00E+00
10086708 OD	mild/moderate	1.06E-06	0.99999893
10088791 OD	mild/moderate	1.19E-21	1
10088791 OS	mild/moderate	9.07903E-06	0.99999094
10093928 OS	mild/moderate	9.8197E-07	0.99999905
10117924-20180220@102014-R2	mild/moderate	0.99951839	0.000481647
10131726-20180319@085722-L3	mild/moderate	9.93E-10	1
10131726-20180319@085722-R2	mild/moderate	2.40E-18	1
10158969-20180506@090506-L2	mild/moderate	9.26337E-06	0.9999907
10166525-20180402@083736-L4	mild/moderate	3.21697E-05	0.99996781
10167812-20180424@134719-L2	mild/moderate	9.34E-06	0.9999907
# of Incorrect Classification		6	0

512×512

Patient	Type	P (Abnormal) [512*512]	P(Normal) [512*512]
10001418 OS	abnormal	0.99936146	0.000638496
10022953 OS	abnormal	0.9960413	0.003958674
10029705-20180416@112408-L4	abnormal	0.99952638	4.74E-04
10029705-20180416@112408-R3	abnormal	0.99940002	6.00E-04
10041101 OD	abnormal	0.97922426	0.020775726
10041101 OS	abnormal	0.98724443	0.01275562
10046318 OS	abnormal	0.92326951	0.07673049
10055620 OD	abnormal	0.9999845	1.55044E-05
10055620 OS	abnormal	0.9982987	1.70E-03
10065949 OS	abnormal	0.003065838	0.99693418
10068165 OD	abnormal	0.97903001	0.020969966
10075046 OD	abnormal	0.96919304	0.030806985
10075046 OS	abnormal	0.034994565	0.96500546
10077406 OD	abnormal	0.96866268	0.031337354
10077406 OS	abnormal	1.77919E-07	0.99999988
10084968 OS	abnormal	0.9262225	0.073777564

10086708 OS	abnormal	0.004060022	0.99593997
10117924-20180220@102014-L4	abnormal	0.999982	1.7998E-05
10145754-20180415@105849-L2	abnormal	0.99996924	3.07769E-05
10145754-20180415@105849-R1	abnormal	0.99999881	1.21834E-06
10148705-20180425@100700-R3	abnormal	0.99959844	0.00040161
10149811-20180225@090046-R1	abnormal	0.97366387	0.026336132
10158969-20180506@090506-R1	abnormal	0.99999619	3.78E-06
10001091 OD	abnormal	0.96762341	0.032376532
10001091 OS	abnormal	0.99683553	3.16E-03
10000464 OS	mild/moderate	6.08433E-06	0.99999392
10000484 OD	mild/moderate	0.01140972	0.9885903
10001418 OD	mild/moderate	0.084231965	0.91576803
10004898 OS	mild/moderate	0.000502195	0.99949777
10008285 OS	mild/moderate	0.005196332	0.99480361
10013404 OD	mild/moderate	0.15164208	0.84835792
10018169 OD	mild/moderate	0.000914671	0.99908531
10022953 OD	mild/moderate	0.046061426	0.9539386
10022979-20180213@155058-R1	mild/moderate	6.56536E-06	0.99999344
10025329 OD	mild/moderate	2.68299E-08	1
10028832 OD	mild/moderate	1.29917E-05	0.99998701
10028832 OS	mild/moderate	0.001329623	0.99867034
10037058 OD	mild/moderate	6.37248E-08	0.99999988
10037058 OS	mild/moderate	4.54236E-09	1
10041971 OD	mild/moderate	0.72162044	0.27837956
10041971 OS	mild/moderate	0.000409728	0.99959034
10047851	mild/moderate	0.027813982	0.97218597
10047851 OD	mild/moderate	0.062140908	0.93785918
10051334 OD	mild/moderate	0.16369528	0.83630472
10051334 OS	mild/moderate	0.001955796	0.99804413
10059609-20180325@094033-L3	mild/moderate	1.27724E-08	1
10065949 OD	mild/moderate	0.53620446	0.46379557
10068165 OS	mild/moderate	0.001630107	0.99836987
10072783-20180121@092503-L2	mild/moderate	0.004932033	0.99506795
10075925 OD	mild/moderate	0.10453955	0.89546043
10081192 OS	mild/moderate	0.010845487	0.98915452
10084968 OD	mild/moderate	0.33051535	0.66948467
10086708 OD	mild/moderate	0.000114546	0.99988544
10088791 OD	mild/moderate	0.68852872	0.31147122
10088791 OS	mild/moderate	7.09177E-07	0.99999928
10093928 OS	mild/moderate	0.001435881	0.99856418
10117924-20180220@102014-R2	mild/moderate	0.000862961	0.9991371
10131726-20180319@085722-L3	mild/moderate	0.004823742	0.9951762
10131726-20180319@085722-R2	mild/moderate	0.000321771	0.99967825

10158969-20180506@090506-L2	mild/moderate	0.091181904	0.90881807
10166525-20180402@083736-L4	mild/moderate	0.11425235	0.88574767
10167812-20180424@134719-L2	mild/moderate	0.027037384	0.97296256
# of Incorrect Classification		2	4

REFERENCES

1. Wong T. Y., Cheung C. M., Larsen M., Sharma S., Simó R. Diabetic retinopathy. Nature Reviews Disease Primers. 2016;**2**, p. 16012.
2. Klein, R., Klein, B. E., Moss, S. E., Davis, M. D. & DeMets, D. L. The Wisconsin epidemiologic study of diabetic retinopathy. II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. Arch. Ophthalmol. **102**, 520–526 (1984).
3. Klein, R., Klein, B. E., Moss, S. E., Davis, M. D. & DeMets, D. L. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch. Ophthalmol. **102**, 527–532 (1984).
4. Mohamed, Q., Gillies, M. C. & Wong, T. Y. Management of diabetic retinopathy: a systematic review. JAMA **298**, 902–916 (2007). **A systematic review of all major trials of treatment options for DR**.
5. Lumbroso, Bruno, and Marco Rispoli. Diabetic Retinopathy. 1st ed., Jaypee Brothers Medical Publisher, 2015.
6. Klein, R., Klein, B. E., Moss, S. E., Davis, M. D. & DeMets, D. L. The Wisconsin epidemiologic study of diabetic retinopathy. IV. Diabetic macular edema. Ophthalmology **91**, 1464–1474 (1984).
7. Klein, R., Klein, B. E., Moss, S. E. & Cruickshanks, K. J. The Wisconsin epidemiologic study of diabetic retinopathy: XVII. The 14-year incidence and progression of diabetic retinopathy and associated risk factors in type 1 diabetes. Ophthalmology **105**, 1801–1815 (1998).
8. McPherson, Richard, and Matthew Pincus. Henry's Clinical Diagnosis and Management by Laboratory Methods. 22nd ed., Elsevier, 2011.
9. Sim, D. A. et al. Patterns of peripheral retinal and central macula ischemia in diabetic retinopathy as evaluated by ultra-widefield fluorescein angiography. Am. J. Ophthalmol. **158**, 144–153 (2014).
10. Wessel, M. M. et al. Peripheral retinal ischaemia, as evaluated by ultra-widefield fluorescein angiography, is associated with diabetic macular oedema. Br. J. Ophthalmol. **96**, 694–698 (2012).
11. Javitt, J. C. & Aiello, L. P. Cost-effectiveness of detecting and treating diabetic retinopathy. Ann. Intern. Med. **124**, 164–169 (1996). **This paper demonstrates the cost-effectiveness of screening for DR**.
12. Singer, D. E., Nathan, D. M., Fogel, H. A. & Schachat, A. P. Screening for diabetic retinopathy. Ann. Intern. Med. **116**, 660–671 (1992).
13. Williams, G. A. et al. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. Ophthalmology **111**, 1055–1062 (2004).
14. LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature **521**, 436–444.
15. Liu S, Cai W, Che H, Pujol S, Kikinis R, Feng D, Fulham MJ. 2015. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Transactions on Biomedical Engineering **62**,1132–1140.
16. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I,

Hulsbergenvan de Kaa C, Bult P, Van Ginneken B, Van der Laak J. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific Reports **6**, 26286.
17. Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. 2016 Convolutional neural networks for diabetic retinopathy. Procedia Comp. Sci. **90**, 200–205.
18. S. N. Sangeethaa, P. U. Maheswari, "An intelligent model for blood vessel segmentation in diagnosing DR using CNN", J. Med. Syst., vol. 42, no. 10, pp. 175, Oct. 2018.
19. Chandrakumar T, R Kathirvel, 2016, Classifying Diabetic Retinopathy using Deep Learning Architecture, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 05, Issue 06 (June 2016).
20. "conv2." MathWorks, MathWorks, www.mathworks.com/help/matlab/ref/conv2.html.
21. Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift."
22. Nair, Vinod, and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." In Proceedings of the 27th international conference on machine learning (ICML-10), pp. 807-814. 2010.
23. Nagi, J., F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L. M. Gambardella. "Max-Pooling Convolutional Neural Networks for Vision-based Hand Gesture Recognition". IEEE International Conference on Signal and Image Processing Applications (ICSIPA2011), 2011.
24. Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." In Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 249-256. 2010.
25. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." In Proceedings of the IEEE international conference on computer vision, pp. 1026-1034. 2015.
26. Saxe, Andrew M., James L. McClelland, and Surya Ganguli. "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks."
27. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, New York, NY, 2006.
28. Murphy, K. P. Machine Learning: A Probabilistic Perspective. The MIT Press, Cambridge, Massachusetts, 2012.
29. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." Advances in neural information processing systems. 2012.
30. X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification", Image and Signal Processing BioMedical Engineering and Informatics (CISP-BMEI) 2017 10th International Congress on. IEEE, pp. 1-11, 2017.

31. “Diabetic Retinopathy Stages.” *Tipsense.me*, tipsense.me/editorz/.
32. “The Benefits of Optomap.” *Optos*, Optos Plc., www.optos.com/en/products/the-benefits-of-optomap/.