Proceedings — AMIA Joint Summits on Translational Science

AMIA Jt Summits Transl Sci Proc. 2018; 2018: 147–155. PMCID: PMC5961805
Published online 2018 May 18. PMID: 29888061

Automated Detection of Diabetic Retinopathy using Deep Learning

Carson Lam, MD,[1] Darvin Yi,[1] Margaret Guo,[2] and Tony Lindsey, PhD[1,3]

Abstract

Diabetic retinopathy is a leading cause of blindness among working-age adults. Early detection of this condition is critical for good prognosis. In this paper, we demonstrate the use of convolutional neural networks (CNNs) on color fundus images for the recognition task of diabetic retinopathy staging. Our network models achieved test metric performance comparable to baseline literature results, with validation sensitivity of 95%. We additionally explored multinomial classification models, and demonstrate that errors primarily occur in the misclassification of mild disease as normal due to the CNNs inability to detect subtle disease features. We discovered that preprocessing with contrast limited adaptive histogram equalization and ensuring dataset fidelity by expert verification of class labels improves recognition of subtle features. Transfer learning on pretrained GoogLeNet and AlexNet models from ImageNet improved peak test set accuracies to 74.5%, 68.8%, and 57.2% on 2-ary, 3-ary, and 4-ary classification models, respectively.

1. Introduction

Approximately four hundred and twenty million people worldwide have been diagnosed with diabetes mellitus. The prevalence of this disease has doubled in the past 30 years[24] and is only expected to increase, particularly in Asia[7]. Of those with diabetes, approximately one-third are expected to be diagnosed with diabetic retinopathy (DR), a chronic eye disease that can progress to irreversible vision loss [8]. Early detection, which is critical for good prognosis, relies on skilled readers and is both labor and time-intensive. This poses a challenge in areas that traditionally lack access to skilled clinical facilities. Moreover, the manual nature of DR screening methods promotes widespread inconsistency among readers. Finally, given an increase in prevalence of both diabetes and associated retinal complications throughout the world, manual methods of diagnosis may be unable to keep apace with demand for screening services[12].

Automated techniques for diabetic retinopathy diagnoses are essential to solving these problems. While deep learning for binary classification in general has achieved high validation accuracies, multi-stage classification results are less impressive, particularly for early-stage disease.

In this paper we introduce an automatic DR grading system capable of classifying images based on disease pathologies from four severity levels. A convolutional neural network (CNN) convolves an input image with a defined weight matrix to extract specific image features without losing spatial arrangement information. We initially evaluate different architectures to determine the best performing CNN for the binary classification task and aim to achieve literature reported performance levels. We then seek to train multi-class models that enhance sensitivities for the mild or early stage classes,

including various methods of data preprocessing and data augmentation to both improve test accuracy as well as increase our effective dataset sample size. We address concerns of data fidelity and quality by collating a set of ophthalmologist verified images. Finally, we address the issue of insufficient sample size using a deep layered CNN with transfer learning on discriminant color space for the recognition task. We then trained and tested two CNN architectures, AlexNet and GoogLeNet, as 2-ary, 3-ary and 4-ary classification models. They are tuned to perform optimally on a training dataset using several techniques including batch normalization, L2 regularization, dropout, learning rate policies and gradient descent update rules[3]. Experimental studies were conducted using two primary data sources, the publicly available Kaggle dataset of 35,000 retinal images with 5-class labels (normal, mild, moderate, severe, end stage) and a physician-verified Messidor-1 dataset of 1,200 color fundus images with 4-class labels. Throughout this study we aim to elucidate a more effective means of classifying early stage diabetic retinopathy for potential clinical benefits.

2. Background and Related Work

Diagnosis of pathological findings in fundoscopy, a medical technique to visualize the retina, depends on a complex range of features and localizations within the image. The diagnosis is particularly difficult for patients with early stage diabetic retinopathy as this relies on discerning the presence of microaneursyms, small saccular outpouching of capillaries, retinal hemorrhages, ruptured blood vessels —among other features—on the fundoscopic images. Prototypical retinal disease stages are shown in Fig. 1.

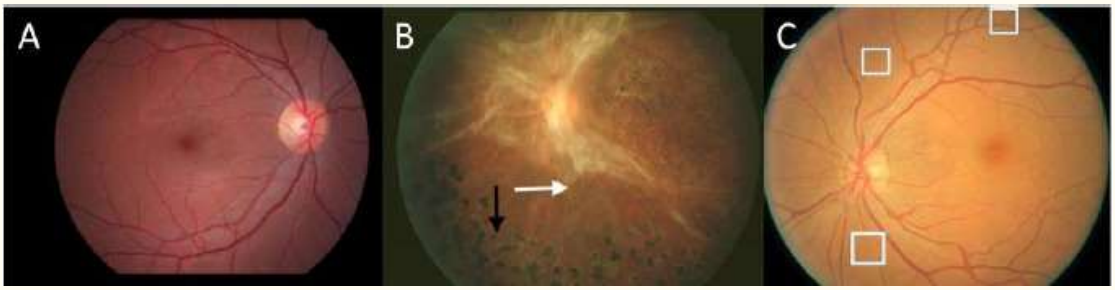

Figure 1.

Representative retinal images of DR at various stages of the disease, as labeled: A- normal, B- end stage, C- early stage. Arrows in B point to pathological indications. White boxes in C enclose very small lesions that the CNNs have difficulty discerning.

Computer-aided diagnosis of diabetic retinopathy has been explored in the past to reduce the burden on opthamologists and mitigate diagnostic inconsistencies between manual readers[16]. Automated methods to detect microaneursyms and reliably grade fundoscopic images of diabetic retinopathy patients have been active areas of research in computer vision[19]. The first artificial neural networks explored the ability to classify patches of normal retina without blood vessels, normal retinas with blood vessels, pathologic retinas with exudates, and pathologic retinas with microaneurysms. The accuracy of being able to detect microaneursyms compared to normal patches of retina was reported at 74% [10].

Past studies using various high bias, low variance digital image processing techniques have performed well at identifying one specific feature used in the detection of subtle disease such as the use of top-hat algorithm for microaneurysm detection [17,23,16]. However, a variety of other features besides microaneurysms are efficacious for disease detection.

Additional methods of detecting microaneurysms and grading DR involving k-NN[5,20], support vector machines [22], and ensemble-based methods [6] have yielded sensitivities and specificities within the 90% range using various feature extraction techniques and preprocessing algorithms.

Previous CNN studies[14,11] for DR fundus images achieved sensitivities and specificities in the range of 90% for binary classification categories of normal or mild vs moderate or severe on much larger private datasets of 80,000 to 120,000 images. However, accuracy measures for the detection of four classes of DR, that is: no DR (R0), mild (R1), moderate (R2), and severe (R3) depend nontrivially on disease graded class collection ratios. While R0 and R3 stages are capable of achieving high sensitivity, the R1 and R2 computed recall rates are often low. Experiments from publicly available datasets suggest this is primarily attributable to the relative difficulty of detecting early stage DR. Furthermore, current accuracies for R1 and R2 stages are reported at 0% and 41%, respectively. We will determine the sensitivity and specificity of our 4-ary classification model and evaluate performance by comparing results to currently published research data.

3. Dataset

We used two fundoscope image datasets to train an automated classifier for this study. Rapid prototyping was facilitated by pretrained models obtained from the ImageNet visual object recognition database. Diabetic retinopathy images were acquired from a Kaggle dataset of 35,000 images with 5-class labels (normal, mild, moderate, severe, end stage) and Messidor-1 dataset of 1,200 color fundus images with 4-class labels (normal, mild, moderate, severe)[9,13]. Both datasets consist of color photographs that vary in height and width between the low hundreds to low thousands. Compared to Messidor-1, the Kaggle dataset consists of a larger proportion of uninterpretable images due to artifact preponderance, faulty labeling and poor quality. After training on the larger Kaggle datasets and identifying limitations of the conventional approach to retinal image classification, we performed experiments on higher fidelity datasets with improved image quality and reliable labeling.

In the interest of efficient model building, we progressed to a smaller but more ideal dataset for learning difficult features. The Messidor dataset was supplemented with a Kaggle partition (MildDR) consisting of 550 images that was verified for its efficacy by direct physician interpretation. The dataset contains images from a disparate patient population with extremely varied levels of fundus photography lighting and is labeled in a consistent manner. The lighting affects pixel intensity values within the images and creates variation unrelated to classification pathology. Our study only uses the retinopathy grade as a reference, a description of which is provided in Table 1 together with the number of images for each category.

Table 1.

Retinopathy grades in Messidor dataset

Grade	Description	Nb Images
R0	($N_{MA} = 0$) AND ($N_{HE} = 0$)	546
R1	($0 < N_{MA} \leq 5$) AND ($N_{HE} = 0$)	153
R2	($5 < N_{MA} < 15$) AND ($0 <$ NHE < 5) AND ($N_{NV} = 0$)	247
R3	($N_{MA} \geq 15$) OR ($N_{HE} \geq 5$) OR ($N_{NV} > 0$)	254

Open in a separate window

N_{MA}, N_{HE}, N_{NV}: number of MAs, HEs and neovessels (NV), respectively

To assess the suitability of transfer learning, we considered several frameworks to design, train and deploy a modified version of GoogLeNet as a baseline 2-ary, 3-ary and 4-ary classification model. Our final model was implemented in Tensorflow and was influenced by results from a deep learning GPU interactive training system (DIGITS) that enabled rapid neural network prototype training, performance monitoring and real-time visualizations.

4. Methods

4.1. CNN Architectures

In order to assess the strengths and limitations of CNNs, several architectures were trained and tested with particular focus on a 22 layers deep model called GoogLeNet. This very efficient network achieves state-of-the-art accuracy using a mixture of low-dimensional embeddings and heterogeneous sized spatial filters[21]. Increased convolution layers and improved utilization of internal network computing resources allow the network to learn deeper features. For example, the first layer might learn edges while the deepest layer learns to interpret hard exudate, a DR classification feature. The network contains convolution blocks with activation on the top layer that defines complex functional mappings between inputs and response variable, followed by batch normalization after each convolution layer. As the number of feature maps increase, one batch normalization per block is introduced in succession.

The max pooling sample-based discretization process was performed with kernel size 3x3 and stride 2[1]. The network was then flattened to one dimension after the final convolutional block. Dropout of network layers was performed until reaching the dense five node output layer, which uses a softmax activation function to compute the probability of classification labels. Leaky rectified linear unit activation was also applied with gradient value 0.01 to mitigate dead neuron bottlenecks during back-propagation[2]. The network uses convolutional layer L2 regularization to reduce model overfitting, cross-entropy computed error loss, and the Xavier method of initializing weights so that neuron activation functions start out in unsaturated regions.

4.2. Preprocessing

All images were converted to a hierarchical data format for preprocessing, data augmentation, and training. Preprocessing involved several steps: images were cropped using Otsu's method to isolate the circular colored image of the retina. Images were normalized by subtracting the minimum pixel

intensity from each channel and dividing by the mean pixel intensity to represent pixels in the range 0 to 1. Contrast adjustment was performed using the contrast limited adaptive histogram equalization (CLAHE) filtering algorithm.

4.3. Data Augmentation

We augmented the number of images in real-time to improve network localization capability and reduce overfitting. During each epoch, a random augmentation of images that preserve collinearity and distance ratios was performed. We implemented random padding with zeros, zoom, rolling and rotation. These affine transformations are particularly effective when applied to disease class R1 which are the most difficult to grade and fewest in number.

4.4. Training and Testing Models

A Deep Learning GPU Training System (DIGITS) with prebuilt convolutional neural networks for image classification facilitated data management, model prototyping and real-time performance monitoring. DIGITS is an interactive system and was first used to build a classification dataset by splitting the Messidor and MildDR fundus folder into training and validation subsets of 1077 and 269 images respectively. The images were cropped to area size 256x256 and used as input data by Imagenet models previously trained for generic classification tasks. The test subset folder contained 400 images from the Lariboisiere hospital Messidor partition and was disjoint from training data. This training system, which offered extensive hyperparameter selections, was then used to build model prototypes over 100 epochs requiring approximately 20 minutes each to complete. A Tesla K80 GPU hardware device powered the training and a form of early stopping determined the optimal test set model epoch. Advanced visualization monitoring and confusion matrix statistical analysis provided important insights. The knowledge gained guided construction of more complex model architectures finely tuned for improved interpretation of our datasets see Appendix.

The more refined models were developed in Tensorflow using modified versions of the open source GitHub package (https://github.com/yidarvin/FirstAid) and evaluated with additional digital image preprocessing techniques.

4.5. Transfer Learning

Transfer learning based approaches were executed with pretrained AlexNet and GoogLeNet architectures from ImageNet. The last fully connected layer was removed, then a transfer learning scenario[18] was followed by treating the remaining network components as a fixed feature extractor for the new dataset. The transfer learning retains initial pretrained model weights and extracts image features via a final network layer.

5. Experiments

5.1. Digital image processing improves sensitivity for mild class detection

The dataset contained images from a disparate patient population with extremely varied levels of lighting in the fundus photography. The lighting affects pixel intensity values within the images and creates unnecessary variation unrelated to classification levels. A contrast limited adaptive histogram equalization filtering algorithm, using the OpenCV (http://opencv.org/) package was applied to address this artifact. Results from this preprocessing step are visually depicted in Fig. 2. We discovered that 3-ary classifier sensitivity for the mild case increased from 0 to 29.4%, while this measure was approximately the same for the remaining two classes Fig. 3. Our digital image preprocessing technique enabled improved detection of pinpoint subtle features and microaneurysms via convolutional filters, which were previously imperceptible by the CNN. We hypothesize this change is attributable to the channel wise contrast enhancing effect of histogram equalization.

Figure 2.

Contrast Limited Adaptive Histogram Equalization enhances contrast and the detection of subtle features. Shown are fundoscopic illustrations before and after CLAHE application.

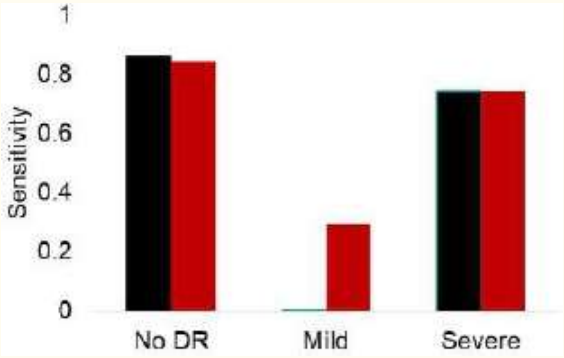

Figure 3.

Sensitivity of a 3-ary (no DR, mild, and severe classes) GoogLeNet classifier before (black) and after (red) CLAHE application on the Messidor dataset.

5.2. Binary model classification attains benchmark performance from literature

AlexNet, VGG16 and GoogLeNet models were trained on the binary-labeled (normal or mild vs moderate to end stage) Kaggle dataset to explore strengths and weaknesses of CNNs. This implementation was performed in Tensorflow, and all model weights were allowed to be updated. The GoogLeNet model achieved the highest sensitivity of 95% and specificity of 96% using our real time data augmentation and preprocessing techniques see Fig. 4. Thus, we successfully demonstrate state-of-the-art accuracy levels that have previously been published in this field.

Figure 4.

Training Curve for model on the binary classified Kaggle data set of DR fundoscope images. Sensitivity of 95% and specificity of 96% was achieved.

5.3. Multi-class training sensitivities is highly dependent on dataset fidelity

However, when we trained 3-ary and 4-ary classifiers with a GoogLeNet model on the Kaggle dataset, we were unable to achieve significant sensitivity levels for the mild class. As shown in the confusion matrix (Table 2), the sensitivity of the no DR and severe DR classes were 98% and 93% respectively; however the sensitivity for the mild class was only 7%. Thus, we find that our performance is limited by the inability of CNNs to detect very subtle features. We hypothesize that this can be explained by the considerable noise and camera artifacts as well as poor fidelity labeling (misleading or incorrect) of the Kaggle images.

Table 2.

Confusion matrix on test set of Kaggle dataset

-	Pred R0	Pred R1	Pred R2 or R3
True R0	149	1	2
True R1	21	2	7
True R2 or R3	1	15	202

Open in a separate window

We gained insight into ways which might improve our results by producing visualizations that reveal at the image level our CNNs usage of high and low level features to improve probability of disease detection. Visualization of pathological areas shown in a heat map generated by sliding window patch-wise occlusion, Fig. 5 demonstrate a large number of false positive and false negative zones, indicating that, while large disease features and those that have a different color from the background are easily interpreted by the CNN, small and subtle features remain undetected.

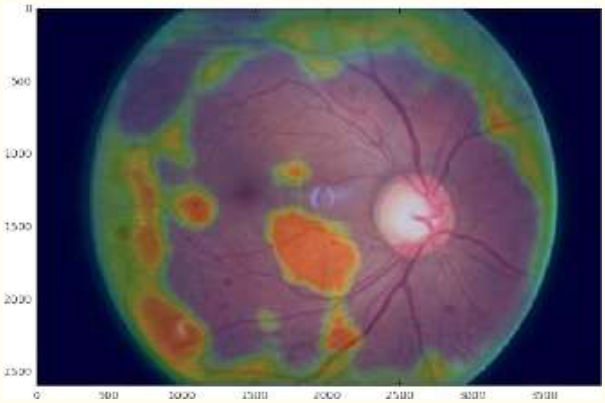

Figure 5.

Heat map on a representative DR image. Green: Regions that do not change the probability of an abnormal binary classification (neutral areas or unfamiliar areas); Orange: Regions that increase the probability of an abnormal binary classification (suspicious areas); Clear or light blue: Regions that decrease the probability of abnormal binary classification (normal areas).

Thus, we consider our Messidor collection which addresses fidelity concerns of the Kaggle dataset (see **Dataset** section for more details on data acquisition procedures).

After following our preprocessing steps, the 3-ary classifier achieves sensitivities for no DR and severe DR of 85% and 75% respectively as well as 29% mild class sensitivity. Given the dramatic measurement increase for the mild class, even with 5% the amount of data, it is evident that data fidelity has a strong impact on multi-class training model performance.

However, the 4-ary classifier encounters a problem of simply not having enough images to effectively train a deep CNN such as GoogLeNet. The multi-class model is unable to distinguish between different classes and behaves as a majority classifier, attempting to classify all images into a single class. In this instance overfitting is not occurring since the crossentropy loss is nondecreasing. Rather the problem is more likely one of underfitting, where the model parameters have not been exposed to enough data requisite for discriminating between the different classes.

5.4. Transfer learning as a parallel means of exploring optimal CNN models

Our previous models relied on a Tensorflow[4] implementation without fixed weights. The practicality of transfer learning was investigated by using a baseline prototype consisting of pretrained GoogLeNet model obtained from the ImageNet visual object recognition database. The prototype was trained on the Messidor dataset for 30 epochs using stochastic gradient descent optimization with step decay learning rate initialized at 0.002. The classification model validation achieved 66.03% as the best accuracy.

Introducing loss (and accuracy layers) to intermediate representations of the deep network allowed for faster propagation during training and avoided the vanishing gradients issue. We found that training converged much faster using the preprocessed images—in a matter of 25 epochs for the transfer learning scenario rather than 90 epochs when training on the raw data.

The accuracy measures indicate how well these intermediate inception layer representations have converged to produce the best classifier. We observed spiking behavior in the training loss curves which suggested a need for weight decay hyperparameter tuning. Additionally, tuning hyper-parameters such as L2 regularization, dropout and batch normalization produced a greater degree of accuracy layer convergence. Full training and test results can be found in the **Appendix** Tables 3 and 4.

Table 3.

Hyperparameter optimization of the Messidor dataset trained using transfer learning on a pretrained GoogLeNet model from ImageNet. 2-ary dataset classes were group C0:R0, R1 and C1:R2, R3. 3-ary dataset classes were C0: R0, C1:R1, and C2:R2, R3. 4-ary dataset classes were C0:R0, C1:R1, C2:R2, C3:R3. C represents the label within the CNN architecture, and *R* represents the label from the dataset.

GoogLeNet Rapid Prototyping Results-Raw Images					
Model	**Solver**	**Learning Rate**	**Policy**	**Validation Accuracy%**	**Test Set Accuracy%**
2-ary	SGD	1e-3	Step Down	83.82	72.75
2-ary	NAG	1e-3	Step Down	82.36	72.75
2-ary	Adam	1e-4	Step Down	86.40	71.75
2-ary	AdaGrad	1e-3	Exponential Decay	84.55	64.25
2-ary	RMSProp	1e-4	Sigmoid Decay	79.04	64.25
3-ary	RMSProp	1e-4	Exponential Decay	63.97	66.25
3-ary	SGD	1e-3	Step Down	71.69	64.25
3-ary	Adam	1e-4	Step Down	72.40	61.50
3-ary	NAG	1e-3	Step Down	69.85	58.75
3-ary	AdaGrad	1e-3	Exponential Decay	72.43	58.25
4-ary	Adam	1e-4	Step Down	67.65	57.25
4-ary	SGD	1e-3	Step Down	65.07	55.25
4-ary	AdaGrad	1e-3	Exponential Decay	66.54	53.25
4-ary	NAG	1e-3	Step Down	66.18	52.75
4-ary	RMSProp	1e-4	Step Down	62.50	49.75

Open in a separate window

Table 4.

Hyperparameter optimization of preprocessed Messidor dataset trained using transfer learning on a pretrained GoogLeNet model from ImageNet. 2-ary dataset classes were group C0:R0, R1 and C1:R2, R3. 3-ary dataset classes were C0: R0, C1:R1, and C2:R2, R3. 4-ary dataset classes were C0:R0, C1:R1, C2:R2, C3:R3. C represents the label within the CNN architecture, and *R* represents the label from the dataset.

GoogLeNet Rapid Prototyping Results-Data Augmentation, Contrast Filtering & Regularization						
Model	**Solver**	**Learning Rate**	**Policy**	**Drop Out**[1]	**Validation Accuracy%**	**Test Set Accuracy%**
2-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	88.35	74.50
2-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	88.07	71.75
2-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	87.50	70.00
2-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	85.80	69.50
2-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	87.22	66.25
3-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	63.28	68.75
3-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	65.63	67.00
3-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	67.71	64.75
3-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	60.68	64.00
3-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	64.32	63.50
4-ary	SGD	1e-3	Step Down	(0.6,0.6,0.4)	60.00	51.25
4-ary	Adam	1e-4	Step Down	(0.8,0.7,0.4)	53.75	49.50
4-ary	NAG	1e-3	Step Down	(0.7,0.8,0.4)	55.00	47.75
4-ary	AdaGrad	1e-3	Exponential Decay	(0.7,0.7,0.5)	57.50	47.00
4-ary	RMSProp	1e-4	Exponential Decay	(0.7,0.7,0.4)	54.75	44.25

Open in a separate window

[1]counting only bypass procedures with two or more *Procedure site – direct* attributes

However, the final test time 3-ary accuracy was similar for both, 67.2% training on the raw data and 71.25% for the transfer learning task. Grouping mild and abnormal together the accuracy was 71.5% and 74.5% respectively. The best test set accuracies can be found in Fig. 6. It is interesting to note that the accuracy of the 4-ary classifier decreases after preprocessing. We hypothesize this is due to the loss of important image features during downsampling. However, we note that there was a gain in the overall detection of subtle features, microaneurysms, of 17% with transfer learning. This method should be further explored to determine if we can enhance sensitivity of these models to mild class DR.

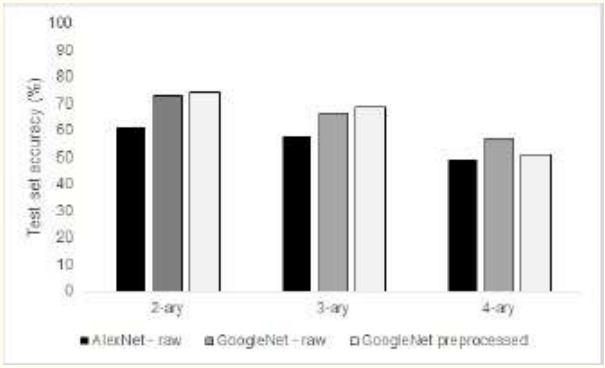

Figure 6.

Test set accuracies for 2-ary, 3-ary, and 4-ary classifiers for transfer learning models based on AlexNet and GoogLeNet. Preprocessed images indicates the presence of real-time data augmentation and histogram equalization.

6. Conclusion

Automated detection and screening offers a unique opportunity to prevent a significant proportion of vision loss in our population. In recent years, researchers have added CNNs into the set of algorithms used to screen for diabetic disease. CNNs promise to leverage the large amounts of images that have been amassed for physician interpreted screening and learn from raw pixels. The high variance and low bias of these models could allow CNNs to diagnose a wider range of nondiabetic diseases as well.

However, while we achieve state-of-the-art performance with CNNs using binary classifiers, the model performance degrades with increasing number of classes. Though it is tempting to surmise that more data may be better, previous work in the field has corroborated that CNN ability to tolerate scale variations is restricted and others have suggested that in the case of retinal images, more data cannot supplement for this inherent limitatio[25,14]. Gulshan et al. reported a 93-96% recall for binary classification of disease but reports that recall is not improved when training with 60,000 samples vs 120,000 samples of a private dataset.

Visualizations of the features learned by CNNs reveal that the signals used for classification reside in a portion of the image clearly visible by the observer[26]. Moderate and severe diabetic retinal images contain macroscopic features at a scale that current CNN architectures, such as those available from the ImageNet visual database, are optimized to classify. Conversely, the features that distinguish mild vs normal disease reside in less than 1% of the total pixel volume, a level of subtleness that is often difficult for human interpreters to detect.

Medical images are fraught with subtle features that can be crucial for diagnosis. Fortunately, the most often deployed architectures have been optimized to recognize macroscopic features such as those present in the ImageNet dataset. We may therefore require a new paradigm for diagnosing diseases via CNN models. This could be a two stage lesion detection pipeline that involves feature localization followed by classification and further preprocessing steps to segment out pathologies difficult to discern by manual inspection, and finally rebalancing network weights to account for class imbalances seen in medical datasets. Overall, our future goals involve improving detection of mild disease and transitioning to more challenging and beneficial multi-grade disease detection.

Acknowledgements

We thank Professor Daniel Rubin MD and the National Library of Medicine for providing grant funds supporting Carson Lam and Darvin Yi, PhD Candidate in Rubin Lab, Department of Biomedical Informatics at Stanford University School of Medicine. Also acknowledged are medical students and physicians Laura C. Huang, Caroline Yu, Emma Zhao, Robert Kleinman, Ryan Smith and Ryan Shields for assistance with curation of the dataset. A related paper was presented at the ARVO annual meeting[15].

A. Appendix

A.1. Messidor and MildDR Raw Images

A.2. Messidor and MildDR Data with Data Augmented and CLAHE Images

References

[1] http://cs231n.github.io/convolutional-networks/

[2] http://cs231n.github.io/neural-networks-1/

[3] http://cs231n.github.io/neural-networks-2/

[4] https://github.com/yidarvin/firstaid.

[5] Abràmoff M. D., Reinhardt J. M., Russell S. R., Folk J. C., Mahajan V. B., Niemeijer M., Quellec G. Automated early detection of diabetic retinopathy. *Ophthalmology,* 2010;117(6):1147–1154. [PMC free article] [PubMed] [Google Scholar]

[6] Antal B., Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE transactions on biomedical engineering,* 2012;59(6):1720–1726. [PubMed] [Google Scholar]

[7] Chan J. C., Malik V., Jia W., Kadowaki T., Yajnik C. S., Yoon K.-H., Hu F. B. Diabetes in asia: epidemiology, risk factors, and pathophysiology. *JAMA,* 2009;301(20):2129–2140. [PubMed] [Google Scholar]

[8] Congdon N. G., Friedman D. S., Lietman T. Important causes of visual impairment in the world today. *Jama,* 2003;290(15):2057–2060. [PubMed] [Google Scholar]

[9] Decenciere E., Zhang X., Cazuguel G., Lay B., Cochener B., Trone C., Gain P., Ordonez R., Massin P., Erginay A., et al. Feedback on a publicly distributed image database: the messidor database. 2014;33:231–234. [Google Scholar]

[10] Gardner G., Keating D., Williamson T., Elliott A. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British journal of Ophthalmology.* 1996;80(11):940–944. [PMC free article] [PubMed] [Google Scholar]

[11] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Elsevier.* 2017 [PubMed] [Google Scholar]

[12] Goh J. K. H, Cheung C. Y., Sim S. S., Tan P. C., Tan G. S. W, Wong T. Y. Retinal imaging techniques for diabetic retinopathy screening. *Journal of diabetes science and technology,* 2016;10(2):282–294. [PMC free article] [PubMed] [Google Scholar]

[13] Graham B. Kaggle diabetic retinopathy detection competition report. 2015 [Google Scholar]

[14] Gulshan V., Peng L., Coram M., Stumpe M. C., Wu D., Narayanaswamy A., Venugopalan S., Widner K., Madams T., Cuadros J., et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA,* 2016;316(22):2402–2410. [PubMed] [Google Scholar]

[15] Huang L. C., Yu C., Kleinman R., Smith R., Shields R., Yi D., Lam C., Rubin D. Opening the black box: Visualization of deep neural network for detection of disease in retinal fundus photographs. *The Association for Research in Vision and Ophthalmology.* 2017 [Google Scholar]

[16] Mookiah M., Acharya U., Chua C., Lim C., Ng E., Laude A. Computer-aided diagnosis of diabetic retinopathy: A review. *In Computers in Biology and Medicine.* 2013:2136–2155. [PubMed] [Google Scholar]

[17] Niemeijer M., Van Ginneken B., Cree M. J., Mizutani A., Quellec G., Sanchez C. I., Zhang B., Hornero R., Lamard M., Muramatsu C., et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE transactions on medical imaging,* 2010;29(1):185–195. [PubMed] [Google Scholar]

[18] Oquab M., Bottou L., Laptev I., Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* 2014:1717–1724. [Google Scholar]

[19] Philip S., Fleming A., Goatman K., Mcnamee P., Scotland G. The efficacy of automated disease/no disease grading for diabetic retinopathy in a systematic screening programme. *In British Journal of Ophthalmology,* 2007:1512–1517. [PMC free article] [PubMed] [Google Scholar]

[20] Quellec G., Lamard M., Josselin P. M., Cazuguel G., Cochener B., Roux C. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transactions on Medical Imaging,* 2008;27(9):1230–1241. [PMC free article] [PubMed] [Google Scholar]

[21] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2016:2818–2826. [Google Scholar]

[22] UR A. Decision support system for diabetic retinopathy using discrete wavelet transform. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine,* 2013;227(3):251–261. [PubMed] [Google Scholar]

[23] Wang S., Tang H. L., Hu Y., Sanei S., Saleh G. M., Peto T., et al. Localizing microaneurysms in fundus images through singular spectrum analysis. *IEEE Transactions on Biomedical Engineering,* 2017;64(5):990–1002. [PubMed] [Google Scholar]

[24] WHO. *Global report on diabetes.* 2016.

[25] Xu Y., Xiao T., Zhang J., Yang K., Zhang Z. Scale-invariant convolutional neural networks. *arXiv preprint arXiv:1411.6369,* 2014 [Google Scholar]

[26] Yosinski J., Clune J., Nguyen A., Fuchs T., Lipson H. Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579,* 2015 [Google Scholar]