Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **4**, July **2018**

A Novel Diabetic Retinopathy Classification Scheme based on Compact Bilinear Pooling CNN and Gradient Boosted Decision Tree

Chun-Yan Lian, Yi-Xiong Liang*, Rui Kang, Yuan Mao and Yao Xiang

School of Information Science and Engineering
Central South University, Changsha, China
yxliang@csu.edu.cn

Ling Gao

Ophthalmology Department of the Second Xiangya Hospital
Central South University, Changsha, China

Received August, 2017; revised February, 2018

Abstract. *Diabetic retinopathy (DR) is one of the leading causes of blindness, but the classification of DR requires experienced ophthalmologist to distinguish the presence of various small features, which is time-consuming and difficult. Therefore, automated DR classification is essential for medical treatment. In this paper, a novel scheme is proposed for automated DR classification, in which a compact bilinear pooling Convolutional Neural Network (CNN) is applied to extract DR features and a Gradient Boosted Decision Decision Tree classifier is trained based on these extracted features to classify DR. Our results on the EyePACS dataset demonstrate the proposed scheme which combines deep learning and tree based approaches achieves a superior performance for automated DR classification with a Kappa score of 0.73, a average F1-score of 0.79 and a micro-average AUC of 0.95.*

Keywords: Diabetic Retinopathy, Convolutional Neural Networks, Compact Bilinear Pooling, Gradient Boosted Decision Tree, Deep Learning

1. **Introduction.** Diabetic Retinopathy (DR), a main complication caused by diabetes, is one of the most common and severe eye diseases causing vision impairment and blindness among adults for age 20-64 [21]. DR progresses through five stages: normal, mild non-proliferative DR (NPDR), moderate NPDR, severe NPDR and proliferative DR (PDR) as shown in Fig.1. Early diagnosis and treatment are crucial to slow down the DR progression and prevent vision loss [4, 19]. Normally, diagnosing DR patients is performed manually by ophthalmologists, which is time-consuming and prone to errors. Therefore, automated DR classification is sorely needed.

A number of automated DR classification schemes have been proposed in the past decade, which can be mainly divided into conventional image-analysis based schemes and deep learning based schemes.

The conventional image-analysis schemes are based on the experiences of ophthalmologists and manually designed features. Sinthanayothin et al [7] proposed a binary DR classification algorithm based on Neural Networks using morphological features. Their method demonstrates a sensitivity of 80.21% and a specificity of 70.66% on a small dataset of 767 images. Nayak et al [12] used Neural Networks based on texture and morphological

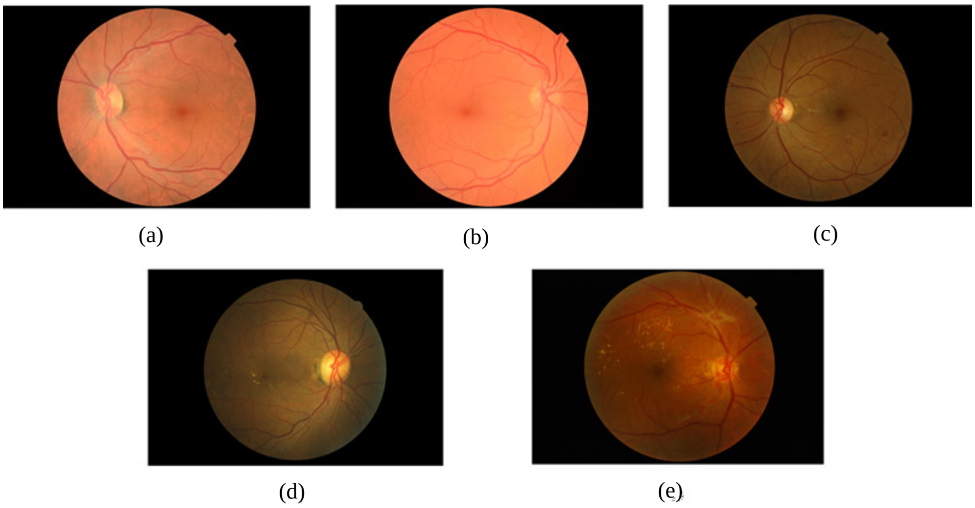

Figure 1. fundus images with different DR stage. (a): normal; (b): mild NPDR; (c): moderate NPDR; (d): severe NPDR; (e): PDR;

features to classify DR into normal, NPDR, PDR with an accuracy of 93%, a sensitivity of 90%, and a specificity of 100% on a 140 images dataset. Acharya et al. [22] used support vector machine (SVM) with Higher Order Spectra (HOS) features to classify image into five class with an accuracy of 82%, a sensitivity of 82% and a specificity of 88% on a dataset of 300 images. These hand-craft features are low-level and conducted on small datasets and thus the results of these schemes tend to overfit. Therefore, the generalization abilities of these models are poor in actual scenes.

Deep learning based schemes have achieved remarkable success and outperform image-analysis based schemes in common computer vision tasks [1, 15, 18, 13]. Although common images and medical images differ significantly, deep learning based schemes still can obtain higher level features with stronger representation abilities and thus achieve superior performances in terms of generalization for medical applications than image-analysis based schemes [23, 8, 14]. Actually, there are already some attempts utilizing CNNs for DR classification. Marco Alban [16] classified image to five classes using convnets, which were trained with off-the-shelf CNN features. It achieves an micro-average the area under curve (AUC) of 0.79. Harry Pratt [9] proposed a CNN structure with ten convolutional layers and three fully-connected layers to extract features and classify images. It achieves an accuracy of 75% and a average F1-score of 0.68. These deep learning based schemes [16, 9] show better generalization abilities than image-analysis based schemes. However, all of them use fully connected layers for pooling whereas ignore the local pairwise feature interactions, which needs further improvements. Furthermore, they use the softmax classifier, which can not represent complex features well and will decrease the performance of DR classification.

To address the above two problems, a scheme which combines compact bilinear pooling CNNs and tree based approaches is proposed in this paper to provide an effective solution for automatic DR classification. First, a compact bilinear pooling CNN is trained to extract DR features. The compact bilinear pooling [25] can gather second-order statistics of local features to improve the performance of automatic DR classification because it consists of two feature extractors, of which the outputs are multiplied using approximate outer product at each location of the image and pooled to obtain an image descriptor. Second, the Gradient Boosted Decision Tree (GBDT) classifier is trained on these features to classify DR. The GBDT combines the output of many weak classifiers into a powerful

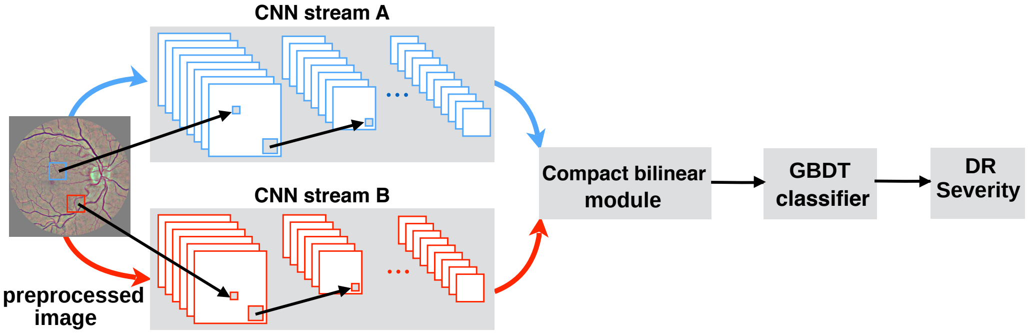

FIGURE 2. Schema of the proposed pipeline

ensemble classifiers, which enhances the performance for automatic DR classification. Our experiment results indicate that the proposed scheme achieves a superior performance in terms of the Kappa score, F1-score and AUC.

The remaining of this paper is organized as follows. Section 2 presents the detailed description of the proposed scheme. The dataset and data preparation are described in section 3. Section 4 evaluates and discusses the results of the experiments. Section 5 concludes the paper.

2. **Proposed Scheme.** In this section,the details of our proposed scheme for DR classification are presented. The scheme consists of two main components, which are a compact bilinear pooling CNN for feature extraction and a GBDT for classification. The schema of the proposed scheme is shown in Fig.2.

2.1. **Compact Bilinear Pooling CNN.** In this paper, we train a CNN model via compact bilinear pooling and then extract the feature from L2 normalization layer. Instead of using fully connected layers for pooling and encoding as conventional CNN models, which loses local features, compact bilinear pooling is used to gather second-order statistics of local features from the whole image for more discriminative representation. This architecture can train end-to-end and learn highly discriminative feature with low dimension.

2.1.1. *Compact Bilinear Pooling.* The bilinear pooling calculates the outer product of the vector $x \in \mathbb{R}^c$ as $w = x \otimes x$, where x represents the output from the CNN stream in our paper, $\otimes$ denotes the outer product (xx^T) and w is a $c \times c$ matrix. However, the representation is very high-dimensional, which makes it is impractical. As suggested by [25], a compact bilinear pooling scheme with Count Sketch [17] projection function is used to reduce the feature dimensionality with little-to-no loss. The detailed procedures of compact bilinear pooling with Count Sketch are as follows. First, Count Sketch is used to project the vector $x \in \mathbb{R}^c$ to $y \in \mathbb{R}^d$ through $\Psi(x, h, s)$, where $s \in \{-1, 1\}^n$ and $h \in \{1, ..., d\}^n$. This step allows us to compute outer product in a lower dimensional space and reduces the number of parameters. Then, the outer product of these two Count Sketch vectors is expressed as the convolution to reduce computation cost.

$$\Psi(x \otimes x, h, s) = \Psi(x, h, s) * \Psi(x, h, s) \tag{1}$$

where $*$ is the convolution operator. Finally, the convolution in the time domain is expressed by the multiplication in the frequency domain as the output of compact bilinear pooling.

$$\Psi(x, h, s) * \Psi(x, h, s) = \mathrm{FFT}^{-1}(\mathrm{FFT}(\Psi(x, h, s)) \circ \mathrm{FFT}(\Psi(x, h, s))) \tag{2}$$

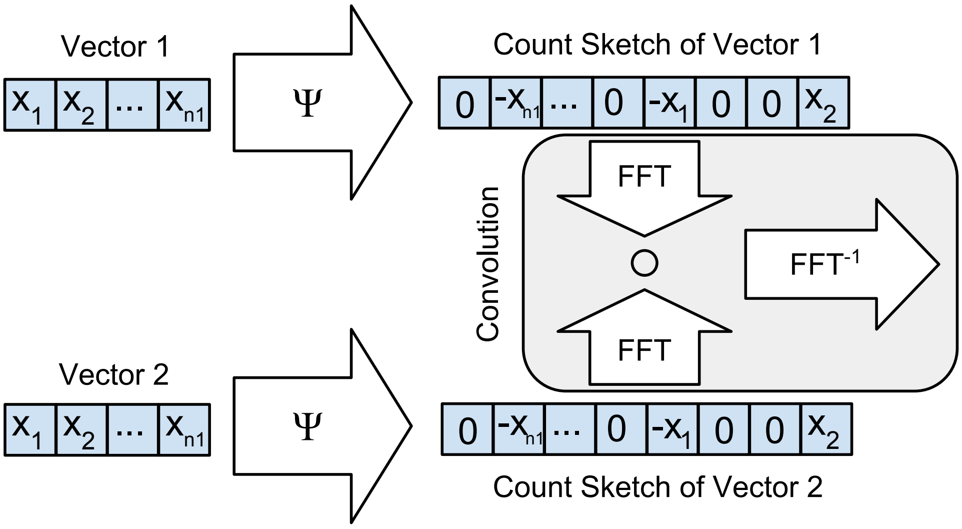

FIGURE 3. Compact bilinear pooling

where $\circ$ refers to element-wise multiplication and FFT refers to the Fast Fourier Transform Algorithm. These ideas are summarized in Fig.3.

2.1.2. *Network Architecture.* We adopt the VGG-16 [15] architecture as the basis of our compact bilinear pooling CNN. To construct the compact bilinear pooling CNN, the layers of FC6 and FC7 in VGG-16 are discarded. And a compact bilinear layer, a signed square root layer and a L2 normalization layer are appended. The network architecture is shown in Table 1.

2.2. **Gradient Boosted Decision Tree.** The conventional CNNs use softmax as the classifier, but the softmax is a linear classifier, which can hardly represent complex features well. In this paper, we use the Gradient Boosted Decision Tree [11] to train a classifier based on the features extracted from L2 normalization layer. The GBDT combines many weak classifiers into a powerful ensemble classifier in an iterative fashion. In each iteration, a weak classifier (decision tree) is added to minimize the loss of ensemble models using the gradient descent. We implement GBDT with XGBoost [20], which uses a better regularized model formalization for over-fitting controlling to ensures better performance.

The objective function of XGBoost is a sum of a specific loss function evaluated over all predictions and a sum of regularization term for all predictors (K trees).

$$obj = \sum_{i}^{n} l(y_i - \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

where obj denotes the objective function, l presents the loss between the prediction $\hat{y}_i$ and target y_i, n is the number of samples, Ω denotes the regularization term, which is used to avoid over fitting. And f_k presents the prediction coming from the k-th tree. The loss function depends on the task (classification, regression, etc.) and the regularization term is described by the equation below:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{4}$$

where T denotes the number of leaves in a tree, w_j^2 is the square of the weight in j-th leaf. γ and λ are hyperparameters which control the degree of regularization. The first part of

Table 1. Network structure

ID	Layer type	Activation maps	Window size	Stride size	Pading size
1	Input	$\emptyset$			
2	Conv	64	3 x 3	1	1
3	Conv	64	3 x 3	1	1
4	MaxPool	$\emptyset$	3 x 3	2	0
5	Conv	128	3 x 3	1	1
6	Conv	128	3 x 3	1	1
7	MaxPool	$\emptyset$	3 x 3	2	0
8	Conv	256	3 x 3	1	1
9	Conv	256	3 x 3	1	1
10	Conv	256	3 x 3	1	1
11	MaxPool	$\emptyset$	3 x 3	2	0
12	Conv	512	3 x 3	1	1
13	Conv	512	3 x 3	1	1
14	Conv	512	3 x 3	1	1
15	MaxPool	$\emptyset$	3 x 3	2	0
16	Conv	512	3 x 3	1	1
17	Conv	512	3 x 3	1	1
18	Conv	512	3 x 3	1	1
19	MaxPool	$\emptyset$	3 x 3	2	0
20	Compact Bilinear Pooling	8192	$\emptyset$		
21	Signed Square Root	8192			
22	L2 normalization	8192			
23	Inner Product	5			

this equation (γT) is responsible for controlling the overall number of created leaves, and the second part ($\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$) watches over their scores.

3. Dataset and Data Preparation.

3.1. Dataset.

The dataset used in this paper is provided by EyePACS [10] and contains 35126 color fundus images. In this dataset, each image is graded by a human reader according to the presence of DR: 0 (normal), 1 (mild NPDR), 2 (moderate NPDR), 3 (severe NPDR) and 4 (PDR). The images are highly heterogeneous because they are captured with various type of digital fundus cameras from different fields of views. Some images in the dataset are in poor quality which are out of focus, underexposed or overexposed or contain artifacts. In addition, both the images and labels involve noises. Furthermore, the distribution of the class in the dataset is extremely imbalanced as shown in Table 2. In our paper, the dataset is split into three part: 80% for training, 10% for validation and 10% for testing.

3.2. Data Preparation.

3.2.1. *Image Preprocessing.* The resolution of these fundus images ranges from 2592×1944 to 4752×3168. To reduce the computational complexity, these images are resized into 448 × 448. And then, color enhancement is performed on fundus image to make the details of the images more clearly inspired by [2].

$$I_c(x,y) = \alpha I(x,y) + \beta G(x,y,\rho) * I(x,y) + \gamma \tag{5}$$

Table 2. The proportion of classes

Label	Class	Number	Percentage
0	Normal	20653	73.46%
1	Mild NPDR	1657	6.69%
2	Moderate NPDR	4234	15.06%
3	Severe NPDR	701	2.50%
4	PDR	569	2.02%

Table 3. Comparison of the Kappa scores for DR classification

Scheme	Harry's [9]	Fully connect CNN	Compact bilinear CNN	Proposed scheme
Kappa score	0.44	0.55	0.70	0.73

where $*$ denotes the convolution operator, $G(x, y, \rho)$ represents the Gaussian filter with a standard deviation of ρ, $I(x, y)$ represents the pixel of the raw image and $I_c(x, y)$ denotes the pixel of the image after preprocessing. The values of $\alpha, \beta, \rho, \gamma$ are designed empirically as 4, -4, 10, 128 respectively.

3.2.2. *Data Augmentation.* As shown in Table 2, the training dataset exhibits imbalance in the class distribution. However, the CNNs schemes tend to be biased on majority class and thus get poor performance on the minority class [5]. To balance the samples across different classes, re-sampling is performed in the training set. For the overrepresented classes, random sub-sampling is applied. For underrepresented classes, spatial translation rotation and crop are employed to increase their numbers artificially. After re-sampling, the number of samples for each class is 7000 in the training set.

4. Experiments and Results.

4.1. **Experiment Settings.** First, an compact bilinear CNN model is trained with Caffe framework [26] on a NVidia GeForce GTX TITAN X GPU. Suggested by [25], Stochastic Gradient Descent is used as optimization method with a learning rate of 0.001. The learning rate decreases by a factor of 2 every 45 epochs with a mini-batch of 64. The weight decay of 0.0005 is added to penalize large weight parameters during back-propagation of the gradient optimization routine. The momentum is fixed as 0.9. Then, we extract the features of L2 normalization layer from the trained compact bilinear CNN model and feed them into a GBDT classifier implemented with XGBoost. The hyperparameters of XGBoost are fine tuned using grid search and 5-fold cross validation.

4.2. **The Evaluation of DR Classification Performance.** To evaluate the DR classification performance of proposed scheme, a fully connect CNN and a compact bilinear pooling CNN are also trained using the same settings and compared with the proposed scheme. In addition, Harry's scheme [9] is also compared. All these schemes are evaluated at the test set and the proportion of each class in the test set is as same as the proportion in the dataset via EyePACS. In our paper, the Cohens Kappa score [3] and the F1-score are used to evaluate the performance. Besides, the receiver operating characteristics (ROC) curve and the area under curve (AUC) metric for each class are conducted to show the performance of the proposed scheme more concretely.

First, we calculate the Kappa scores of the Harry's scheme, the fully connect CNN, the compact bilinear CNN and the proposed scheme for DR classification. The evaluated results are listed in Table 3. As shown in Table 3, the Harry's scheme plays the worst

Table 4. Comparison of the F1-scores for DR classification

Scheme	Harry's [9]	Fully connect CNN	Compact bilinear CNN	Proposed scheme
Normal	0.86	0.75	0.91	0.91
Mild NPDR	0.00	0.18	0.29	0.27
Moderate NPDR	0.30	0.48	0.46	0.54
PDRSevere NPDR	0.14	0.28	0.41	0.41
PDR	0.37	0.53	0.51	0.59
Average	0.68	0.66	0.78	0.79

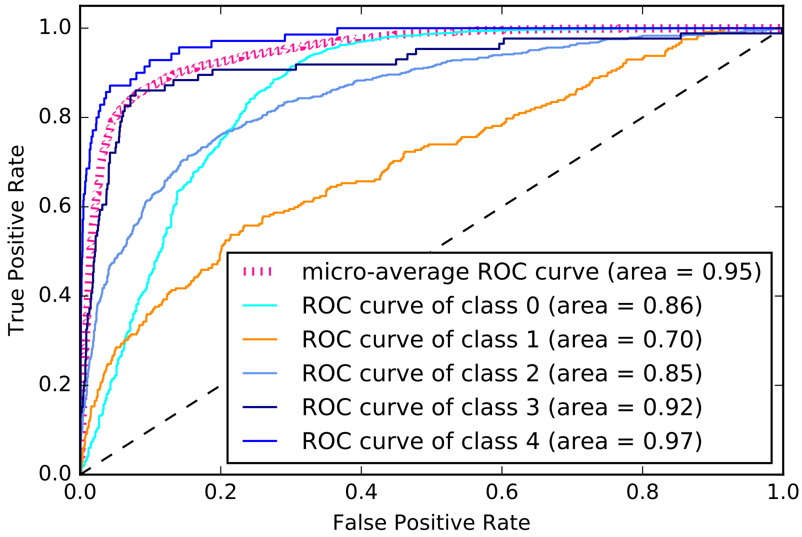

Figure 4. Multiclass ROC-Curves of the proposed scheme

performance with a Kappa score of 0.44, and the proposed scheme achieves the hightest Kappa score of 0.73. Compact bilinear pooling is consistently beneficial for automated DR classification, as evidenced by the comparison of the Kappa scores of the fully connect CNN and the compact bilinear CNN, which are 0.55 and 0.70, respectively. GBDT is also beneficial for automated DR classification, as evidenced by the comparison of the Kappa scores of the compact bilinear CNN and our proposed scheme, which are 0.70 and 0.73, respectively.

Second, we calculate the F1-scores of these four schemes to evaluate the performance for DR classification in another perspective. We classify DR into five classes: normal, mild NPDR, moderate NPDR, severe NPDR and PDR. The F1-scores are reported in each class and average, as illustrated in Table 4. The reported average is a prevalence-weighted average across classes. The proposed scheme yields notably higher F1-scores than other three schemes in all classes except for mild NPDR. The average F1-score of our proposed scheme is 0.79 and higher than those of the other three schemes, which are 0.68, 0.66 and 0.78, respectively. The results demonstrate that our proposed scheme achieves remarkable performance for automated DR classification. However, all the F1-scores of mild DR are especially poor because the representation of mild DR in fundus image is microaneurysm, which appears as a small round dot and its diameter usually ranges from 10 μm to 100 μm[6]. When resizing the image from high resolution such as 2592×1944 into 448×448, almost all microaneurysms disappear and the images are similar to the normal images. The model classifies most mild DR into normal with the indiscriminative inputs.

Finally, we draw the ROC curves and calculate AUC metric for each class to show the performance of proposed scheme more concretely. The ROC-Curves of the proposed scheme with five classes are shown in Fig.4. Class 0-4 corresponds to normal, mild NPDR, moderate NPDR, severe NPDR and PDR. The AUC is also used as a performance metric. The test dataset exhibits imbalance in the class distribution, hence micro-average AUC is used to measure the performance. The details of micro-average can be found in [24]. As shown in Fig.4, the proposed scheme yields the least competitive AUC of 0.70 for class 1, the best AUC of 0.97 for class 4 and a micro-average AUC of 0.95. These results indicate our proposed scheme performs well for DR classification.

5. **Conclusions.** In this paper, a novel scheme which combines compact bilinear pooling CNNs and tree based approaches is proposed to classify DR automatically based on the fundus images. The compact bilinear pooling gather second-order statistics of local features to generate more highly discriminative representation. The tree based approach can fit complex features well and build a powerful classifier. The experiment results on the EyePACS dataset demonstrate that the proposed scheme achieves superior performance for automated DR classification.

6. **Acknowledgement.** This research is partially supported by the Natural Science Foundation of Hunan Province, China (No. 14JJ2008), the Fundamental Research Funds for the Central Universities of Central South University(No. 2017zzts513) and the National Natural Science Foundation of China under Grant No. 61602522, No. 61573380, No. 61672542.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[2] B. Graham, Kaggle diabetic retinopathy detection competition report, Technical report, 2015.

[3] B. Ted, B. Janet and C. John B, Bias, prevalence and kappa, *Journal of clinical epidemiology* vol. 46, no. 5, pp. 423–429, 1993.

[4] C. A. Corr and D. M. Corr, *Death & dying, life & living*, Nelson Education, Ontario, Canada, 2012.

[5] C. Huang, Y. Li, C. Change Loy, and X. Tang, Learning deep representation for imbalanced classification, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5375–5384, 2016.

[6] C. Pereira, D. Veiga, J. Mahdjoub, Z. Guessoum, L. Gonçalves, M. Ferreira, and J. Monteiro, Using a multi-agent system approach for microaneurysm detection in fundus images, *Artificial intelligence in medicine*, vol. 60, no. 3, pp. 179–188, 2014.

[7] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, Automated screening system for diabetic retinopathy, in *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, vol. 2, pp. 915–920, 2003.

[8] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, Mitosis detection in breast cancer histology images with deep neural networks, in *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 411–418, 2013.

[9] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, Convolutional neural networks for diabetic retinopathy, *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[10] J. Cuadros and G. Bresnick, Eyepacs: an adaptable telemedicine system for diabetic retinopathy screening, *Journal of diabetes science and technology*, vol. 3, no. 3, pp. 509–516, 2009.

[11] J. H. Friedman, Greedy function approximation: a gradient boosting machine, *Annals of statistics*, pp. 1189–1232, 2001.

[12] J. Nayak, P. S. Bhat, R. Acharya, C. Lim, and M. Kagathi, Automated identification of diabetic retinopathy stages using digital fundus images, *Journal of medical systems*, vol. 32, no. 2, pp. 107–115, 2008.

[13] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, *arXiv preprint arXiv:1512.03385*, 2015.

[14] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation, *arXiv preprint arXiv:1603.05959*, 2016.

[15] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[16] M. Alban and T. Gilligan, Automated detection of diabetic retinopathy using fluorescein angiography photographs, 2016.

[17] M. Charikar, K. Chen, and M. Farach-Colton, Finding frequent items in data streams, *Theoretical Computer Science*, vol. 312, no. 1, pp. 3–15, 2004.

[18] M. Lin, Q. Chen, and S. Yan, Network in network, *arXiv preprint arXiv:1312.4400*, 2013.

[19] R. Hazin, M. Colyer, F. Lum, and M. K. Barazi, Revisiting diabetes 2000: challenges in establishing nationwide diabetic retinopathy prevention programs, *American journal of ophthalmology*, vol. 152, no. 5, pp. 723–729, 2011.

[20] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* pp. 785–794, 2016.

[21] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, Automatic detection of microaneurysms in color fundus images, *Medical image analysis*, vol. 11, no. 6, pp. 555–566, 2007.

[22] U. R. Acharya, E. Ng, J.-H. Tan, S. V. Sree, and K.-H. Ng, An integrated index for the identification of diabetic retinopathy stages using texture parameters, *Journal of medical systems*, vol. 36, no. 3, pp. 2011–2020, 2012.

[23] V. Murthy, L. Hou, D. Samaras, T. M. Kurc, and J. H. Saltz, Center-focusing multi-task cnn with injected features for classification of glioma nuclear images, *arXiv preprint arXiv:1612.06825*, 2016.

[24] V. Vincent, Macro-and micro-averaged evaluation measures, 2013.

[25] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, Compact bilinear pooling, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 317–326, 2016.

[26] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, 2014.