Combining Fine- and Coarse-Grained Classifiers for Diabetic Retinopathy Detection *

Muhammad Naseer Bajwa[0000−0002−4821−1056]1,2, Yoshinobu Taniguchi[0000−0002−0625−9123]3, Muhammad Imran Malik[4], Wolfgang Neumeier[5], Andreas Dengel[1,2], and Sheraz Ahmed[0000−−0002−4239−6520]2

[1] Technische Universität Kaiserslautern, Kaiserslautern 67663, Germany
[2] German Research Center for Artificial Intelligence GmbH (DFKI), Kaiserslautern 67663, Germany
bajwa@dfki.uni-kl.de {andreas.dengel,sheraz.ahmed}@dfki.de
[3] Osaka Prefecture University, Naka, Sakai, Osaka 599-8531, Japan
taniguchi@m.cs.osakafu-u.ac.jp
[4] School of Electrical Engineering and Computer Science, National University of Science and Technology (NUST), Islamabad 46000, Pakistan
malik.imran@seecs.edu.pk
[5] Opthalmology Clinic, Rittersberg 9, 67657 Kaiserslautern, Germany
dr.neumeier-kl@web.de

Abstract. Visual artefacts of early diabetic retinopathy in retinal fundus images are usually small in size, inconspicuous, and scattered all over retina. Detecting diabetic retinopathy requires physicians to look at the whole image and fixate on some specific regions to locate potential biomarkers of the disease. Therefore, getting inspiration from ophthalmologist, we propose to combine coarse-grained classifiers that detect discriminating features from the whole images, with a recent breed of fine-grained classifiers that discover and pay particular attention to pathologically significant regions. To evaluate the performance of this proposed ensemble, we used publicly available EyePACS and Messidor datasets. Extensive experimentation for binary, ternary and quaternary classification shows that this ensemble largely outperforms individual image classifiers as well as most of the published works in most training setups for diabetic retinopathy detection. Furthermore, the performance of find-grained classifiers is found notably superior than coarse-grained image classifiers encouraging the development of task-oriented find-grained classifiers modeled after specialist ophthalmologists.

Keywords: Computer-Aided Diagnosis, Medical Image Analysis, Automated Diabetic Retinopathy Detection, Convolutional Neural Network, Deep Learning in Ophthalmology

1 Introduction

Diabetic patients are at constant risk of developing Diabetic Retinopathy (DR) that may eventually lead to permanent vision loss if left unnoticed or untreated.

* Partially funded by Higher Education Commission (HEC) of Pakistan through Faculty Development Program of National University of Science and Technology (NUST), Pakistan

In such patients, increased blood sugar, blood pressure, and cholesterol can cause small blood vessels in retina to protrude and, in due course, haemorrhage blood into retinal layers and/or vitreous humour [4]. In severe conditions, scar tissues and newly proliferated fragile blood vessels blanket the retina and obstruct incoming light from falling on it. As a result, retina is unable to translate light into neural signals which results in blindness. Diabetic retinopathy advances slowly and gradually and may take years to reach proliferative stage, however, almost every diabetic patient is potentially susceptible to this complication.

Timely diagnosis is the key to appropriate prognosis. Ophthalmologists usually detect DR by examining retinal fundus and looking for any signs of microaneurysms (bulging of blood vessels), blood leakage, and/or neovascularization [2]. While the indications of advanced stages of DR are rather prominent, these symptoms remain largely discrete in early stages. Figure 1 shows progress of DR from healthy to proliferative stage in Retinal Fundus Images (RFIs) taken from EyePACS dataset[6]. It can be observed from the figure that the difference between healthy and early stages of DR are very subtle and not readily discernible. Manual analysis of these images requires highly qualified and specialized ophthalmologists who may not be easily accessible in developing countries or remote areas of developed countries. Even when medical experts are available, large scale analysis of RFIs is highly time-consuming, labour-intensive and prone to human error and bias. Furthermore, manual diagnosis by clinicians is largely subjective and rarely reproducible and, therefore, inter-expert agreement for a certain diagnosis is generally very poor.

Computer-Aided Diagnosis (CAD) based on deep learning can provide easily accessible, efficient and economical solution for large-scale initial screening of many diseases including diabetic retinopathy. CAD can perform objective analysis of the given image and predict standardized and reproducible diagnosis, which is free from any bias or tiredness. It can not only help physicians by reducing their workload but can also outreach to underprivileged population and afford them the opportunity of swift and cost-effective initial screening, which may effectively prevent advancement of disease into severer stage. Convolutional

(a) Healthy (b) Mild (c) Moderate (d) Severe (e) Proliferative

Fig. 1: Progression of diabetic retinopathy from healthy to proliferative stage is subtle and gradual. Images are taken from EyePACS train set.

[6] https://www.kaggle.com/c/diabetic-retinopathy-detection/data

Neural Networks (CNNs) are computer algorithms inspired by biological visual cortex. They work especially well in visual recognition tasks. CNNs have been used to perform at par with or even outperform humans in various challenging image recognition problems [14, 13]. Today automated image recognition can be divided into coarse-grained classification and fine-grained classification. In former case, images are classified into high-level categories like humans, animals, vehicles and other objects in a natural scene, for example. In later case, classification is focused on low-level categories like species of dogs or models of cars etc. Fine-grained classification is particularly challenging owning to high intra-class variations and low inter-class variations. Although DR is also a fine-grained classification task, it has normally been addressed using simple coarse-grained classification algorithms.

In this work we used a combination of general and fine-grained deep CNNs to analyze RFIs and predict automated diagnosis for DR. We used two of the most popular conventional image classification architectures i.e Residual Networks [11] and Densely Connected Networks [12], a network search framework called NASNet [25] and two recently proposed methods for fine-grained classification namely NTS-Net [24] and SBS Layer [18]. We tried to harvest the combined potential of these two approaches by training them separately and taking their ensemble during inference. We used EyePACS and Messidor datasets for evaluation. Since previous researches have used vastly disparate experimental setups, we cannot directly compare our results with most of them. However, we performed a broad range of experiments, following the most common problem settings in the literature like normal vs abnormal, referable vs non-referable, ternary and quaternary classification in order to define benchmarks which will afford future works with an opportunity of fair comparison.

1.1 Related Work

Over the past decade, machine learning and deep learning have been used to detect various pathologies, segment vessels and classify DR grades using RFIs. Welikala et al. [23] detected proliferative DR by identifying neovascularization. They used an ensemble of two networks trained separately on 100 different patches for each network. The patches are taken from a selected set of 60 images collected from Messidor [8] and a private dataset. Since the dataset had only 60 images they performed leave-one-out cross validation and achieved 0.9505 Area Under the Curve (AUC) and sensitivity of 1 with specificity of 0.95 at the optimal operating point. Wang et al. [22] identified suspicious regions in RFIs and classified DR into normal (nDR) vs abnormal (aDR) and referable (rDR) vs non-referable (nrDR). They developed a CNN based model called Zoom-in-Network to identify important regions. To classify an image the network uses the overview of the whole images and pays particular attention to important regions. They took 182 images from EyePACS dataset and let a trained ophthalmologist draw bounding boxes around 306 lesions. On Messidor dataset they achieved 0.921 AUC, 0.905 accuracy and 0.960 sensitivity at 0.50 specificity for nDR vs aDR.

Gulshan et al. [10] conducted a comprehensive study to distinguish rDR from nrDR grades. They trained a deep CNN on 128175 fundus images from a private dataset and tested on 9963 images from EyePACS-I and 1748 images of Messidor-2. They achieved AUC of 0.991 on EyePACS-I and 0.990 on Messidor-2. Guan et al. [9] proposed that modeling each classifier after individual human grader instead of training a single classifier using average grading of all human experts improves classification performance. They trained 31 classifiers using a dataset of total 126522 images collected from EyePACS and three other clinics. The method is tested on 3547 images from EyePACS-I and Messidor-2, and achieved 0.9728 AUC, 0.9025 accuracy, and 0.8181 specificity at 0.97 sensitivity. However, it would have been more interesting if they had provided comparison of their suggested approach with ensemble of 31 networks modeled after average grading. Costa et al. [7] used adversarial learning to synthesize color retinal images. However, the performance of their classifier trained on synthetic images was less than the classifier trained on real images. Aujih et al. [5] found that blood vessels play important role in disease classification and fundus images without blood vessels resulted in poor performance by the classifier.

The role of multiple filter sizes in learning fine-grained features was studied by Vo et al. [21]. To this end they used VGG network with extra kernels and combined kernels with multiple loss networks. They achieved 0.891 AUC for rDR vs nrDR and 0.870 AUC for normal vs abnormal on Messidor dataset using 10-fold cross validation. Somkuwar et al. [20] performed classification of hard exudates by exploiting intensity features using 90 images from Messidor dataset and achieved 100% accuracy on normal and 90% accuracy on abnormal images. Seoud et al. [19] focused on red lesions in RFIs, like haemorrhages and microaneurysms, and detected these biomarkers using dynamic shape features in order to classify DR. They achieved 0.899 AUC and 0.916 AUC for nDR vs aDR and rDR vs nrDR, respectively on Messidor. Rakhlin et al. [16] used around 82000 images taken from EyePACS for training and around 7000 EyePACS images and 1748 images from Messidor-2 for testing their deep learning based classifier. They achieved 0.967 AUC on Messidor and 0.923 AUC on EyePACS for binary classification. Ramachandran et al. [17] used 485 private images and 1200 Messidor images to test a third party deep learning based classification platform, which was trained on more than 100000 images. Their validation gave them 0.980 AUC on Messidor dataset for rDR vs nrDR classification. Quellec et al. [15] capitalized a huge private dataset of around 110000 images and around 89000 EyePACS images to train and test a classifier for rDR vs nrDR grades and achieved 0.995 AUC on EyePACS.

2 Materials and Methods

This section provides details on the datasets used in this work and the ensemble methodology employed to perform classification.

2.1 Datasets

We used EyePACS dataset published publicly by Kaggle for a competition on Diabetic Retinopathy Detection. Table 1 gives overview of EyePACS dataset.

Although this dataset is very large in size, only about 75% of its images are of reasonable quality that they can be graded by human experts [16]. EyePACS is graded on a scale of 0 to 4 in accordance with International Clinical Diabetic Retinopathy (ICDR) guidelines [3]. However, low gradability of this dataset raises suspicions on the fidelity of labels provided with each image. We pruned the train set to get rid of 657 completely uninterpretable images. For testing on EyePACS we used 33423 images randomly taken from test set.

Table 1: Overview of EyePACS Dataset. IRMA stands for IntraRetinal Microvascular Abnormalities

Severity Grade	Criterion	Train Set		Test Set	
		Images	Percentage	Images	Percentage
0	No Abnormalities	25810	73.48	39533	73.79
1	Microaneurysms Only	2443	6.95	3762	7.02
2	More than just microaneurysms but less than Grade 3	5292	15.07	7861	14.67
3	More than 20 intraretinal hemorrhages in each of 4 quadrants OR Definite venous beading in 2+ quadrants OR Prominent IRMA in 1+ quadrant AND no signs of proliferative retinopathy	873	2.48	1214	2.27
4	Neovascularization OR Vitreous/preretinal hemorrhage	708	2.02	1206	2.25
	Total	35126	100	53576	100

Messidor dataset consists of 1200 images collected at three different clinics in France. Each clinic contributed 400 images. This dataset is graded for DR on a scale of 0 to 3 following the criteria given in Table 2. Messidor dataset is validated by experts and is, therefore, of higher quality than EyePACS in terms of both image quality and labels.

2.2 Methodology

Figure 2 illustrates complete pipeline of the system combining coarse-grained and find-grained classifiers. Before feeding an image to the network, we first applied Otsu Thresholding to extract and crop retinal rim from RFI and get rid of irrelevant black background. Since the images in both datasets are taken with different cameras and under different clinical settings, they suffer from large brightness and color variations. We used adaptive histogram equalization to normalize brightness and enhance the contrast of visual artefacts which are critical for DR detection. Since the images are in RGB colour space, we first translate them into YCbCr colour space to distribute all luminosity information in Y channel and colour information in Cb and Cr channels. Adaptive histogram equalization is then applied on Y channel only and the resultant image is converted back

Table 2: Overview of Messidor Dataset

Severity Grade	Criterion	Images	Percentage
0	No microaneurysms AND No haemorrhages	546	45.50
1	Microaneurysms <= 5 AND No haemorrhages	153	12.75
2	5 <Microaneurysms <15 AND 0 <Haemorrhages <5 AND No Neovascularization	247	20.58
3	Microaneurysms >= 15 OR Haemorrhages >= 5 OR Neovascularization	254	21.17
	Total	1200	100

to RGB colour space. We further normalized the images by subtracting local average colour from each pixel to highlight the foreground and help our network detect small features. Figure 3 shows the effects of preprocessing steps on RFIs. These pre-processed images are then used to train all five networks individually. During inference, each network gives diagnosis which are ensembled to calculate the final prediction.

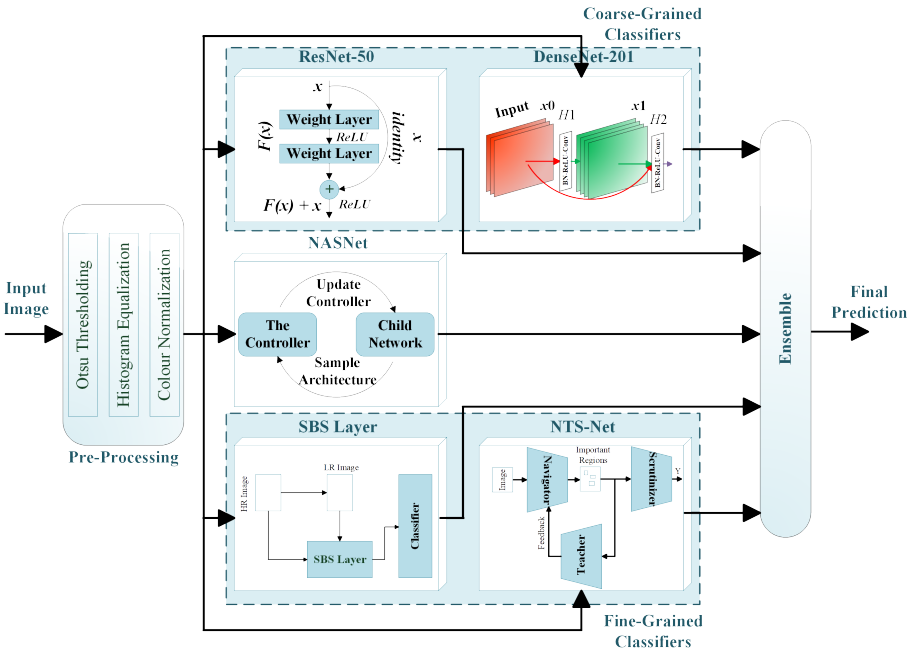

Fig. 2: System Overview: Combining Coarse-Grained and Fine-Grained Classifiers

(a) Original Image before Preprocessing (b) After Contrast Enhancement (c) After Local Average Colour Subtraction

Fig. 3: Effects of preprocessing steps on retinal fundus images

Experimental Setup From EyePACS train set, we randomly selected 30000 images for training and rest of the 4469 images were used for validation. Test set of EyePACS was used for reporting results on this dataset. From Messidor, we used 800 images for training and 400 images from Lariboisière Hospital for testing (as done by Lam et al. [6]). We employed a broad range of hyper parameters during training. All networks are initialized with pre-trained weights and fine-tuned on ophthalmology datasets. To evaluate these models on EyePACS and Messidor datasets under similar problem settings, we first parallelized DR grades of both datasets using criteria given in Figure 4.

3 Results and Analysis

From section 1.1 we observe that previous works on EyePACS and Messidor have used disparate train and test splits and different classification tasks for example Quaternary, Ternary and Binary (rDR vs nrDR and nDR vs aDR). Furthermore, different researchers use different performance metrics to evaluate

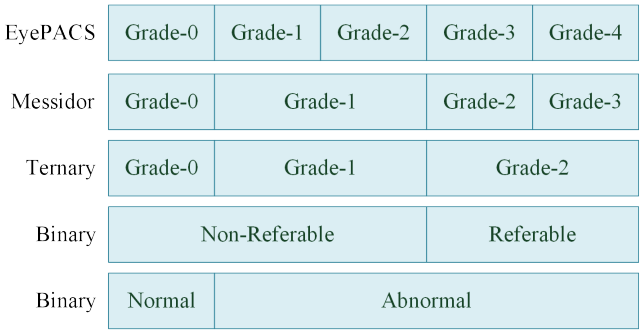

Fig. 4: Conversion of EyePACS grades to Quaternary, Ternary and Binary Classification

their method. Therefore, in such scenario comparison of any two works in not directly possible [1]. However, we conducted extensive experiments to perform all four classification tasks mentioned above and report comprehensive results to allow a rough comparison with some of the published state-of-the-art results on these datasets.

3.1 Results of Binary Classification

As discussed above, many previous works focus primarily on binary classification as nDR vs aDR or rDR vs nrDR grading. The criteria to convert 4 or 5 grades into binary grades is given in Figure 4. For our binary classification, the number of images used for training, validation and testing from EyePACS and Messidor are given in Table 3 and 4. It can be seen from the tables that there is extensive class imbalance between both classes. Table 5 provides detailed performance metrics for all classification tasks including nDR vs aDR classification. Our results are competitive to that of Wang et al. in terms of accuracy and for all other metrics we outperform them. It should be noted here that Wang et al. performed 10-fold cross validation and although their sensitivity of 96 is higher than our 89.75, it is calculated at 50% specificity while ours is at 90% specificity.

Table 3: Class Distribution for Normal vs Abnormal Classification

Grade	EyePACS			Messidor		
	Train	Validate	Test	Train	Validate	Test
Normal	22668	2744	24741	346	49	151
Abnormal	7332	1725	8682	354	51	249
Total	30000	4469	33423	700	100	400

Table 4: Class Distribution for Referable vs Non-Referable Classification

Grade	EyePACS			Messidor		
	Train	Validate	Test	Train	Validate	Test
Non-Referable	28825	4177	31937	453	65	181
Referable	1175	292	1486	247	35	219
Total	30000	4469	33423	700	100	400

Results of rDR vs nrDR classification can also be found in Table 5. All networks performed significantly better for this task than for normal vs abnormal classification on EyePACS dataset reaching maximum accuracy around 96% with 99.44% AUC using SBS layer architecture. For Messidor dataset, both NTS-Net and SBS Layer stand out from traditional classifiers. NTS-Net outperforms all other methods in all metrics, whereas ensemble of all methods gives sub-optimal performance than individual fine-grained methods. This can happen when majority of classifiers used for ensembling have a skewed performance towards downside and only a few give standout results.

3.2 Results of Multi-Class Classification

The complexity of classification task was gradually increased from binary to ternary and quaternary classification. Table 6 and 7 show the class distribution in train, validation and test splits for this multi-class setting. For ternary classification we used the criterion used by [6], as shown in Figure 4.

Table 5: Detailed Performance Metrics for Various Classification Settings

Model	Accuracy (%)		AUC (%)		Sensitivity (%)		Specificity (%)	
	EyePACS	Messidor	EyePACS	Messidor	EyePACS	Messidor	EyePACS	Messidor
Results of Binary (Normal vs Abnormal) Classification								
NTS-Net	**88.19**	88.00	92.72	95.20	88	88.00	72	87.51
SBS Layer	80.11	89.50	86.20	95.17	80	89.50	54	**92.07**
ResNet-50	82.86	87.75	89.46	95.06	83	87.75	75	90.49
DenseNet-201	82.66	87.75	89.69	95.89	83	87.75	**77**	88.14
NASNet	82.19	87.25	88.49	95.04	82	87.25	73	89.14
Ensemble	87.74	**89.75**	**93.44**	**96.50**	**88**	**89.75**	75	91.44
Vo et. al	N/A	87.10	N/A	87.00	N/A	88.2	N/A	85.7
Wang et. al	N/A	90.50	N/A	92.10	N/A	96	N/A	50
Soud et. al	N/A	N/A	N/A	89.90	N/A	N/A	N/A	N/A
Results of Binary (Referable vs Non-Referable) Classification								
NTS-Net	94.93	**93.25**	99.10	**96.56**	95	**93**	75	**94**
SBS Layer	**95.89**	88.75	**99.44**	94.90	**96**	89	67	90
ResNet-50	95.08	86.75	98.97	94.95	95	87	81	89
DenseNet-201	94.70	89.25	99.05	95.33	95	89	82	91
NASNet	91.98	87.50	97.45	95.16	92	88	**85**	89
Ensemble	95.34	89.25	99.23	96.45	95	89	81	91
Lam et. al	N/A	74.5	N/A	N/A	N/A	N/A	N/A	N/A
Vo et. al	N/A	89.70	N/A	89.10	N/A	89.3	N/A	90
Wang et. al	N/A	91.10	N/A	95.70	N/A	97.8	N/A	50
Seoud et. al	N/A	74.5	N/A	91.60	N/A	N/A	N/A	N/A
Results of Ternary Classification								
NTS-Net	84.43	84.50	94.89	94.61	84	85	72	**94**
SBS Layer	76.93	84.50	90.95	94.12	77	85	50	91
ResNet-50	81.23	80.50	93.51	93.79	81	81	74	92
DenseNet-201	79.20	80.25	92.87	94.25	79	80	**77**	93
NASNet	78.95	81.75	91.93	94.00	79	82	71	89
Ensemble	**84.94**	**85.25**	**95.28**	**95.40**	**85**	**85**	73	92
Lam et. al	N/A	68.8	N/A	N/A	N/A	N/A	N/A	N/A
Results of Quaternary Classification								
NTS-Net	82.53	74.50	95.72	91.84	83	75	**76**	**92**
SBS Layer	82.00	65.00	95.69	88.43	82	65	67	88
ResNet-50	81.82	70.25	95.53	91.31	82	70	71	89
DenseNet-201	79.38	74.00	95.04	92.26	79	74	75	91
NASNet	73.73	71.75	92.06	90.84	74	72	74	86
Emsemble	**83.42**	**76.25**	**96.31**	**92.99**	**83**	**76**	73	91
Lam et. al	N/A	57.2	N/A	N/A	N/A	N/A	N/A	N/A

Table 6: Class Distribution for 4-Class Classification

Grade	EyePACS			Messidor		
	Train	Validate	Test	Train	Validate	Test
0	22668	2744	24741	346	49	151
1	6157	1433	7196	107	16	30
2	685	166	753	155	22	70
3	490	126	733	92	13	149
Total	30000	4469	33423	700	100	400

Table 7: Class Distribution for 3-Class Classification

Grade	EyePACS			Messidor		
	Train	Validate	Test	Train	Validate	Test
0	22668	2744	24741	346	49	151
1	6157	1433	7196	107	16	30
2	1175	292	1486	247	35	219
Total	30000	4469	33423	700	100	400

Performance of individual networks and their ensemble for ternary and quaternary classification is given in Table 5. Ensemble of all models gave better performance in this case. We also observe that the performance of NTS-Net is higher than all other individual networks. Our accuracies for both ternary and quaternary classification are superior than accuracies reported by Lam et al. [6]. Figure 5 provides a detailed overview of classification performance of ensemble.

4 Conclusion

Diabetic Retinopathy detection using retinal fundus images is a fine-grained classification task. The biomarkers of this disease on retinal images are usually very small in size, especially for early stages, and are scattered all across the image. The ratio of pathologically important region to the whole input volume is therefore minuscule. Due to this reason traditional deep CNNs usually struggle to identify regions of interest and do not learn discriminatory features well. This

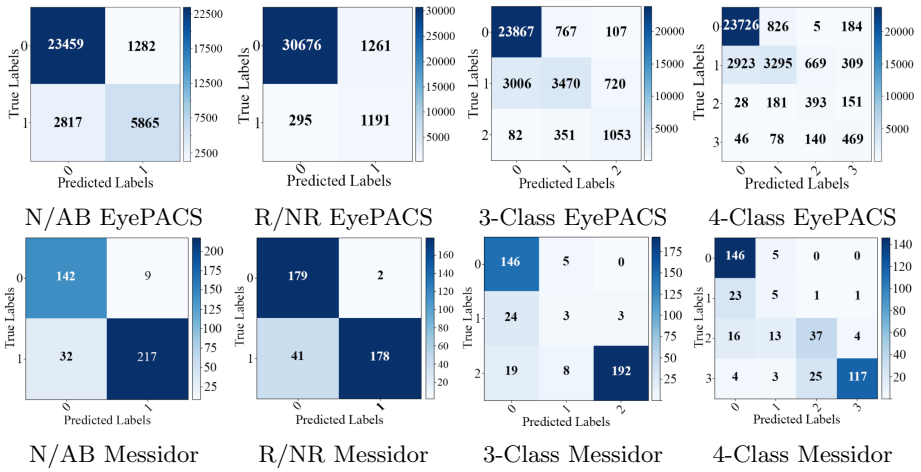

Fig. 5: Confusion Matrices for EyePACS and Messidor for all Classification Tasks. N/AB refers to Normal vs Abnormal; R/NR refers to Referable vs Non-Referable

problem of small and distributed visual artefacts coupled with unavailability of large publicly available high quality dataset with reasonable class imbalance makes DR detection particularly challenging for deep CNN models. However, fine-grained classification networks have high potential to provide standardized and large scale initial screening of diabetic retinopathy and help in prevention and better management of this disease. These networks are equipped with specialized algorithms to discover the important region from the image and pay particular heed to learn characterizing features from those regions.

We achieved superior performance for diabetic retinopathy detection on binary, ternary and quaternary classification tasks than many previously reported results. However, due to hugely different experimental setups and choice of performance metrics, it is unfair to draw a direct comparison with any of the cited research. Nevertheless, we have provided a wide spectrum of performance metrics and detailed experimental setup for comparison by any future work.

References

1. Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M.: Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative ophthalmology & visual science **57**(13), 5200–5206 (2016)
2. Akram, M.U., Khalid, S., Tariq, A., Khan, S.A., Azam, F.: Detection and classification of retinal lesions for grading of diabetic retinopathy. Computers in biology and medicine **45**, 161–171 (2014)
3. American Academy of Ophthalmology: International clinical diabetic retinopathy disease severity scale (2002), http://www.icoph.org/dynamic/attachments/resources/diabetic-retinopathy-detail.pdf
4. Amin, J., Sharif, M., Yasmin, M., Ali, H., Fernandes, S.L.: A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions. Journal of Computational Science **19**, 153–164 (2017)
5. Aujih, A., Izhar, L., Mériaudeau, F., Shapiai, M.I.: Analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification. In: 2018 International conference on intelligent and advanced system (ICIAS). pp. 1–6. IEEE (2018)
6. Carson Lam, D.Y., Guo, M., Lindsey, T.: Automated detection of diabetic retinopathy using deep learning. AMIA Summits on Translational Science Proceedings **2017**, 147 (2018)
7. Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M., Campilho, A.: End-to-end adversarial retinal image synthesis. IEEE transactions on medical imaging **37**(3), 781–791 (2018)
8. Decencière, 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. Image Analysis & Stereology **33**(3), 231–234 (2014)
9. Guan, M.Y., Gulshan, V., Dai, A.M., Hinton, G.E.: Who said what: Modeling individual labelers improves classification. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)

10. 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 **316**(22), 2402–2410 (2016)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
13. Lee, C.Y., Gallagher, P.W., Tu, Z.: Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In: Artificial intelligence and statistics. pp. 464–472 (2016)
14. Liu, C., Zoph, B., Shlens, J., Hua, W., Li, L., Fei-Fei, L., Yuille, A.L., Huang, J., Murphy, K.: Progressive neural architecture search. CoRR **abs/1712.00559** (2017), http://arxiv.org/abs/1712.00559
15. Quellec, G., Charrière, K., Boudi, Y., Cochener, B., Lamard, M.: Deep image mining for diabetic retinopathy screening. Medical image analysis **39**, 178–193 (2017)
16. Rakhlin, A.: Diabetic retinopathy detection through integration of deep learning classification framework. bioRxiv p. 225508 (2018)
17. Ramachandran, N., Hong, S.C., Sime, M.J., Wilson, G.A.: Diabetic retinopathy screening using deep neural network. Clinical & experimental ophthalmology **46**(4), 412–416 (2018)
18. Recasens, A., Kellnhofer, P., Stent, S., Matusik, W., Torralba, A.: Learning to zoom: a saliency-based sampling layer for neural networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 51–66 (2018)
19. Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J.P.: Red lesion detection using dynamic shape features for diabetic retinopathy screening. IEEE transactions on medical imaging **35**(4), 1116–1126 (2016)
20. Somkuwar, A.C., Patil, T.G., Patankar, S.S., Kulkarni, J.V.: Intensity features based classification of hard exudates in retinal images. In: India Conference (INDICON), 2015 Annual IEEE. pp. 1–5. IEEE (2015)
21. Vo, H.H., Verma, A.: New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. In: Multimedia (ISM), 2016 IEEE International Symposium on. pp. 209–215. IEEE (2016)
22. Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H., Wang, X.: Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 267–275. Springer (2017)
23. Welikala, R., Dehmeshki, J., Hoppe, A., Tah, V., Mann, S., Williamson, T.H., Barman, S.: Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. Computer methods and programs in biomedicine **114**(3), 247–261 (2014)
24. Yang, Z., Luo, T., Wang, D., Hu, Z., Gao, J., Wang, L.: Learning to navigate for fine-grained classification. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 420–435 (2018)
25. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8697–8710 (2018)