

Leveraging Multimodal Deep Learning Architecture with Retina Lesion Information to Detect Diabetic Retinopathy

Vincent S. Tseng^{1,2}, Ching-Long Chen³, Chang-Min Liang³, Ming-Cheng Tai³, Jung-Tzu Liu⁴, Po-Yi Wu⁴, Ming-Shan Deng⁴, Ya-Wen Lee⁴, Teng-Yi Huang⁴, and Yi-Hao Chen³

¹ Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

² Institute of Data Science and Engineering, National Chiao Tung University, Hsinchu, Taiwan

³ Department of Ophthalmology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

⁴ Computational Intelligence Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

Correspondence: Yi-Hao Chen, Department of Ophthalmology, Tri-Service General Hospital, No. 325, Sec. 2, Cheng-Kong Rd., Taipei 114, Taiwan. e-mail: doc30879@mail.ndmctsgh.edu.tw

Received: December 9, 2019

Accepted: May 28, 2020

Published: July 16, 2020

Keywords: diabetic retinopathy; fusion architecture; convolutional neural network; object detection; fundus image

Citation: Tseng VS, Chen C-L, Liang C-M, Tai M-C, Liu J-T, Wu P-Y, Deng M-S, Lee Y-W, Huang T-Y, Chen Y-H. Leveraging multimodal deep learning architecture with retina lesion information to detect diabetic retinopathy. *Trans Vis Sci Tech.* 2020;9(2):41, <https://doi.org/10.1167/tvst.9.2.41>

Purpose: To improve disease severity classification from fundus images using a hybrid architecture with symptom awareness for diabetic retinopathy (DR).

Methods: We used 26,699 fundus images of 17,834 diabetic patients from three Taiwanese hospitals collected in 2007 to 2018 for DR severity classification. Thirty-seven ophthalmologists verified the images using lesion annotation and severity classification as the ground truth. Two deep learning fusion architectures were proposed: late fusion, which combines lesion and severity classification models in parallel using a postprocessing procedure, and two-stage early fusion, which combines lesion detection and classification models sequentially and mimics the decision-making process of ophthalmologists. Messidor-2 was used with 1748 images to evaluate and benchmark the performance of the architecture. The primary evaluation metrics were classification accuracy, weighted κ statistic, and area under the receiver operating characteristic curve (AUC).

Results: For hospital data, a hybrid architecture achieved a good detection rate, with accuracy and weighted κ of 84.29% and 84.01%, respectively, for five-class DR grading. It also classified the images of early stage DR more accurately than conventional algorithms. The Messidor-2 model achieved an AUC of 97.09% in referral DR detection compared to AUC of 85% to 99% for state-of-the-art algorithms that learned from a larger database.

Conclusions: Our hybrid architectures strengthened and extracted characteristics from DR images, while improving the performance of DR grading, thereby increasing the robustness and confidence of the architectures for general use.

Translational Relevance: The proposed fusion architectures can enable faster and more accurate diagnosis of various DR pathologies than that obtained in current manual clinical practice.

Introduction

Diabetic retinopathy (DR) is a sight-threatening disease; however, timely diagnosis in the early stage can reduce the occurrence of vision loss or blindness by spurring timely medical intervention and management of glucose levels and blood pressure.¹⁻⁴ Long-term diabetes is likely to cause DR that can impair

the microvascular transport of blood and nutrients to the retina, causing it to leak or swell and eventually lead to blindness. Taiwan's National Health Insurance system recommends annual fundus examination for diabetic patients to detect DR. The examination rate is low because most patients are unaware of their condition until they experience vision reduction.^{5,6} To increase adherence to the examination, a one-stop

service consisting of primary care and retinal imaging has been established.⁷

However, several other issues remain to be addressed in DR detection. The first issue is expertise: a well-trained ophthalmologist is required for DR grading and lesion type assessment.^{8,9} The second issue is intergrader reliability: human interpretation of imaging varies among ophthalmologists.^{10,11} The third issue is manpower: the compound annual growth rate (CAGR) of the number of eye doctors (CAGR: 2.60%) is lower than that of the diabetic population in Taiwan (CAGR: 4.78%).^{12,13} Thus, there is an urgent and critical need for an artificial intelligence–based approach to support decision making.¹⁴ Therefore, developing a robust and automated grading system for DR that gives a prompt response is required to support frontline clinicians who are not experts in the ophthalmology field. This would reduce the clinicians' workload and alleviate the personnel insufficiency associated with a large number of diabetic patients.

The DR severity level determination is based on the observed findings from the fundus image. The International Clinical Diabetic Retinopathy Disease Severity (ICDR) Scale has been widely used to identify patients with signs related to the types of DR lesions, such as microaneurysms (MA), hemorrhages (H), and exudates (EX).^{1,15} According to the signs and distribution of lesions of the ICDR scale definition, the DR severity can be divided into five levels: no apparent retinopathy, mild nonproliferative DR (NPDR), moderate NPDR, severe NPDR, and proliferative DR (PDR).¹ A patient need not be referred to an ophthalmologist if his or her eye could be graded as nonreferable DR (less than moderate NPDR); the patient could be referred only for referable DR (moderate/severe NPDR and PDR).¹ The early signs of DR are MA, H, EX, and so on.^{9,16} MA is the first clinical sign of DR and the only characteristic in mild NPDR. Therefore, MA recognition is critical in the clinical management of DR and patient education. Hence, in addition to a convolutional neural network (CNN)–based grading model, we focus on lesion information as a complementary feature to improve the DR severity classification.

Previous algorithms incorporating lesion information have shown promising results,^{16,17} but their inference speed is hindered by the patch-based method.^{18,19} Hence, we propose two CNN-based fusion architectures, instead of using lesion patches as the inputs, to support DR grading efficiently. To understand better the signs of lesions and which types and distributions affect the DR severity, we explored whether the proposed architectures can increase the robustness and interpretability of DR severity classification. Two architectures are proposed: a late fusion method to

combine two deep learning architectures by a postprocessing procedure and a two-stage early fusion method to exploit lesion localization at pixel level for DR classification. Assuming that the extracted neighborhood context of lesions enhances the classification performance, the lesion detection or localization may support clinical diagnosis, especially for subtle lesion detection in the early stages of DR. As such, we aimed to identify the DR severity of Taiwanese diabetic patients using fundus images from 2007 to 2018 with added lesion information via an improved hybrid recognition method.

Methods

This section presents detailed information on the collected database and proposes two different architectures for fusing both lesion information and a grading network for DR classification. First, a late fusion architecture combines the grading model and lesion-classification model via a postprocessing procedure. Second, a two-stage early fusion architecture highlights the suspicious DR lesions and produces fully weighted lesion images in the first stage. Then, raw images and fully weighted images are trained jointly in the second stage for DR grading.

Database

This study used two data sets: a private data set from three Taiwan hospitals and a public data set, Messidor-2. For the private data set, we used 26,699 fundus images obtained from 17,834 patients between 2007 and 2018 at Tri-Service General Hospital, Chung Shan Medical University Hospital, and China Medical University Hospital. The hospitals' institutional review boards and the Industrial Technology Research Institute approved this study, and the research followed the tenets of the Declaration of Helsinki. The need for informed consent was waived owing to the retrospective nature of the study. A variety of ophthalmoscopes were used with 45° fields of view. A group of board-certified ophthalmologists independently graded the images based on the ICDR scale,¹ and they annotated the corresponding lesions. The private data set was randomly split into three independent data sets based on patient IDs: training set (22,617 images), validation set (2039 images), and testing set (2043 images). The distributions of the five-class DR severity and four-type DR lesion are shown in [Figure 1](#). A new distribution of Messidor-2,^{20,21} with 1748 images (78.26% nonreferable DR and 21.74% referable DR), was used for the testing as well.

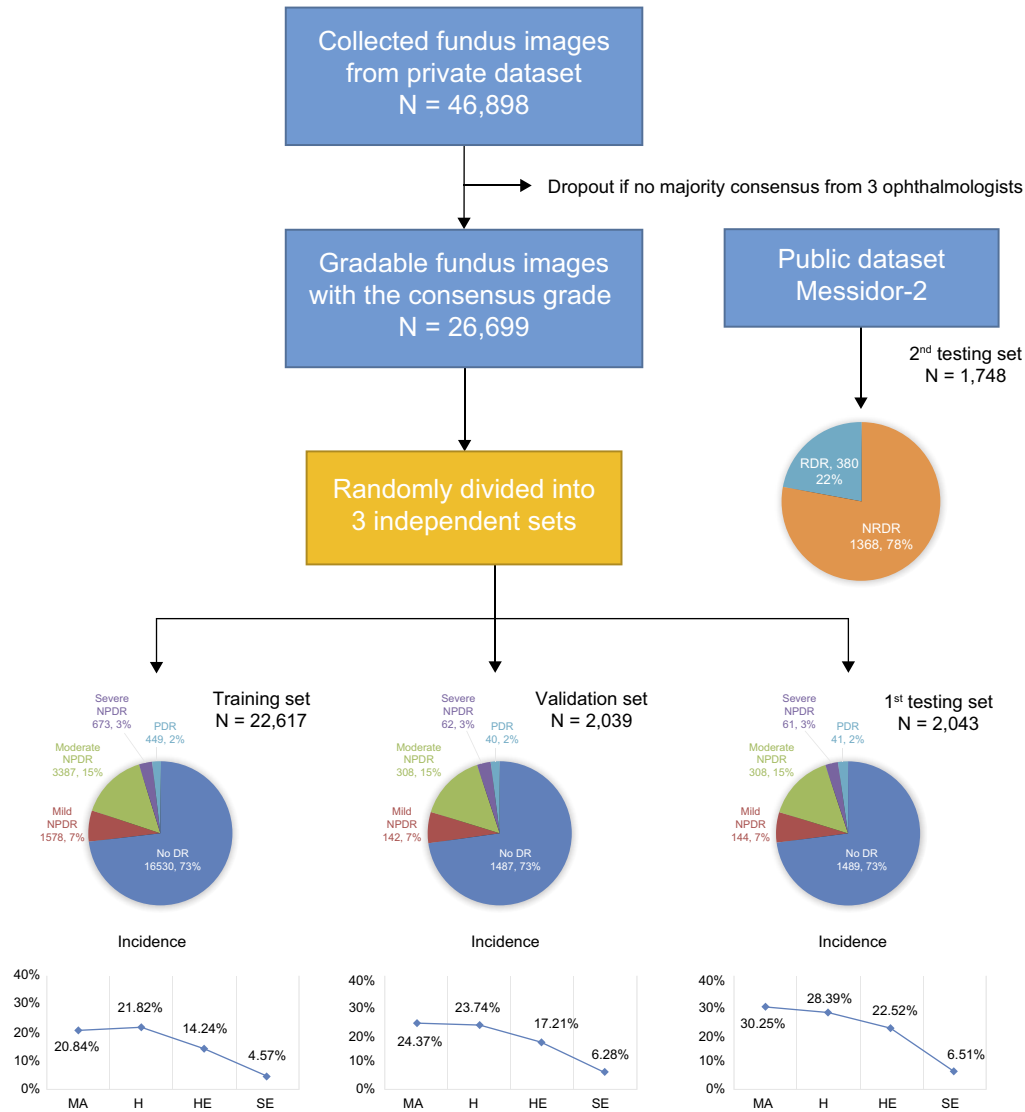


Figure 1. Workflow diagram showing distribution of DR severity level and the incidence rate of DR lesions in a different data set.

Ground Truth

For the private data set, the ground truth (GT) of disease severity for each image was based on the majority consensus of the three ophthalmologists. If there was an ungradable image or an image without a majority consensus for a five-class classification, that image was removed to minimize grading bias. The percentage of such removals was 43%. A total of 26,699 images were used after dropout (Fig. 1).

The GT of lesion location for each image was based on the following rules: (1) bounding boxes for the image labeled by two ophthalmologists are compared. If the same symptom is marked and the intersection over union (IoU) > 25%, then the intersection area is taken as the GT. (2) Otherwise, the marked symptoms are

retained as the GT. (3) The GT obtained from the previous steps is compared with the image marked by the third ophthalmologist, and then the GT is updated.

Figure 2 shows the annotated lesion combination process between two ophthalmologists. Based on the rules, the final GT distribution of DR lesion types by DR severity level could be determined. As seen in Figure 3, lesions are marked in the majority level of no DR. It implies that one of the graders marked lesion(s) and the other two graders marked no lesion in the same image. It is worth noting that the number of lesions in severe NPDR and PDR is less than that in moderate NPDR. This arises because the invasion of the area is usually greater in the severe levels of DR, and the relative number of lesions may decrease. Furthermore, the signs of neovascularization should be

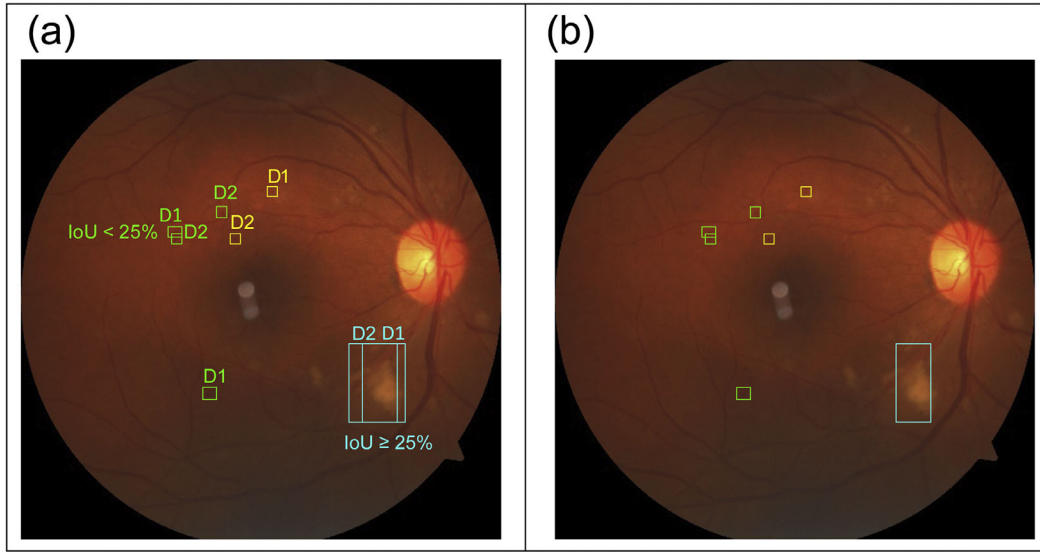


Figure 2. Lesion location GT production process. (a) Lesion annotated by two ophthalmologists (D1 and D2). (b) Rule-based combination results.

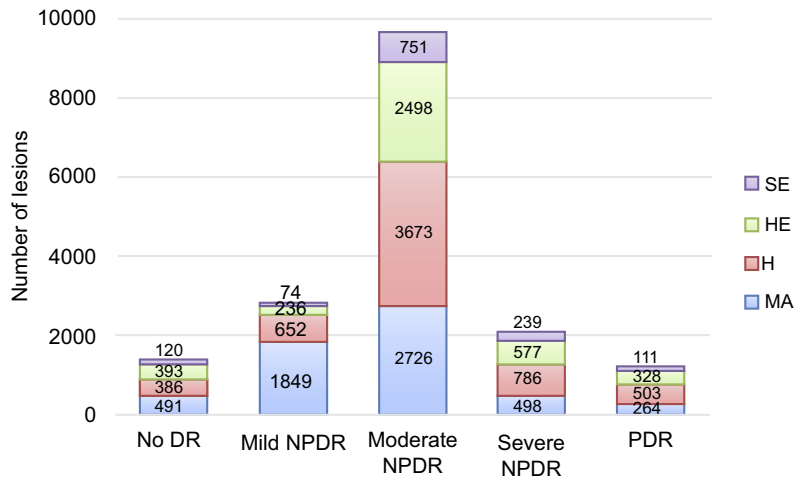


Figure 3. Distribution of DR lesion types by DR severity level.

taken into consideration in the judgment of PDR for a complete research.

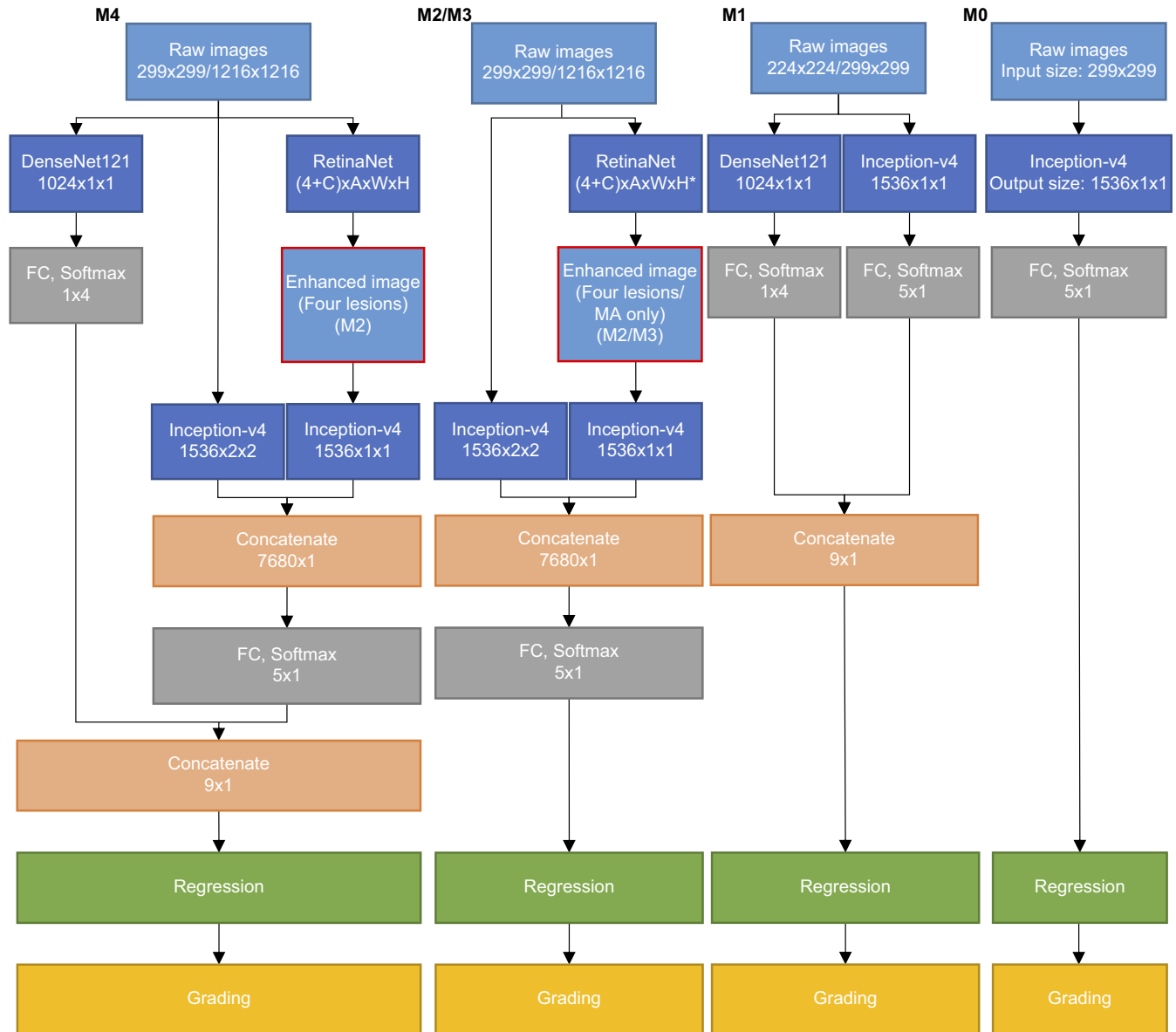
For the public data set, Messidor-2, the grades made available by Abramoff²⁰ were adopted in this study.

Late Fusion

To prepare useful information in the training process, image preprocessing was conducted in which the nonretinal background was cropped from raw images. As can be seen in Figure 4, we developed a late fusion model (M1) in which the grading model (baseline model, M0) and the four lesion type-

classification models were trained independently with the cropped images. With images of size 299×299 as the inputs, we trained a CNN model using the Inception-v4 architecture²² for grading and another four CNN models using the DenseNet architecture²³ with images of size 224×224 as the inputs for binary lesion classification.

The lesion-classification models (with or without a lesion type of MA, H, hard exudates [HE], or soft exudates [SE]) were used for feature extraction. The lesion-classification features were deemed as supplementary information by the late fusion architecture. As the features from the softmax regression were the



*C: number of lesion class; A: number of anchor; W: kernel weight; H: kernel height

Figure 4. Workflow of the baseline model (M0) and four fusion models (M1–M4).

final outputs produced from the heterogeneous models (grading model and four binary lesion-classification models), a postprocessing method combined all the features with an ordinal ridge regression model to classify the disease severity (Fig. 4).

Two-Stage Early Fusion

Inspired by previous work,¹⁶ we also developed a two-stage early fusion architecture (M2/M3) in which two different types of input images are used for grading DR. As can be seen in Figure 4, we used the raw input

images instead of lesion patches for training an object-detection model based on RetinaNet²⁴ with images of size 1216×1216 in the first stage. Our object-detection model was trained to enhance the four major symptoms of suspicious DR regions in a full image (Fig. 5). In the second stage, a classification model using Inception-v4 with lesion-enhanced images and raw images of size 299×299 was simultaneously trained for severity classification. Both features were concatenated before the fully connected fusion layer.

Specifically, we replaced the raw RGB pixels with new pixels to highlight the potential DR lesions in the first stage. To enhance the suspicious DR lesions

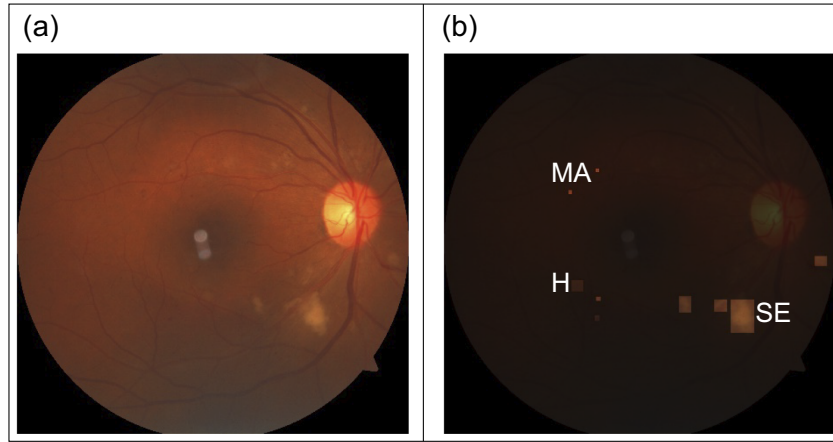


Figure 5. Input images: (a) raw image and (b) enhanced image with highlighted lesion locations (MA, H, and SE).

from the predicted regions, first, we divided the original image RGB matrix by 4; second, we multiplied it by the lesion type based on the predicted annotation; and third, we multiplied it by a function $f(x)$ based on the confidence level information (c) of the detected lesion. The confidence level of the detector may suppress the degree of enhancement. Therefore, two enhancement strategies were used: strategy 1 ($S1$) using the confidence level ($x, x \in [0, 1]$) as additional information and strategy 2 ($S2$) without using the confidence level. The formula to calculate the new RGB pixel is shown in Equation (1):

$$\text{New RGB pixel} = \text{raw RGB pixel}/4 \times \text{lesion type} \times f(x), \quad (1)$$

where lesion type is 1 (for no DR), 2 (for H), 3 (for HE/SE), or 4 (for MA), and $f(x)$ is defined as $f(x) = \begin{cases} c & \text{if } x \in S1 \\ 1 & \text{if } x \in S2 \end{cases}$.

As Figure 6 shows, in $S1$, if an observed confidence level of the four main symptoms with a value of 0.5 and a raw pixel value of 255 existed for each symptom, then the new pixel values were updated as 127.5, 63.75, and 95.625, respectively. Alternatively, using $S2$ in $M2$, the new weighted pixel values of the four main symptoms were increased to 255, 127.5, and 191.25. The differences among the pixel values of different DR lesions are, therefore, elevated without information suppression, using $S2$.

Furthermore, for early DR detection purposes in model $M3$, we focused on MA detection alone and modified Equation (1) to obtain Equation (2):

$$\text{New RGB pixel} = \text{raw RGB pixel}/2 \times \text{lesion type} \times f(x) \quad (2)$$

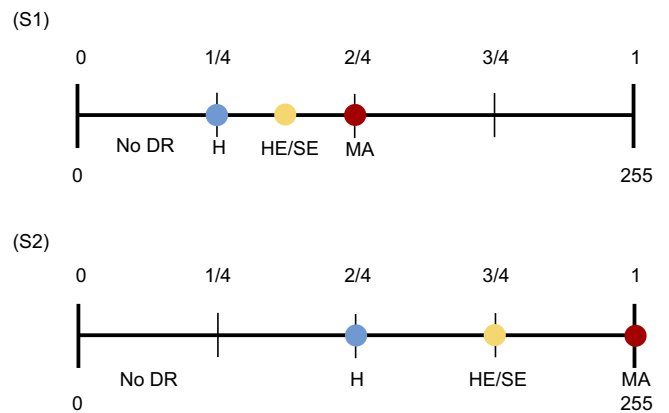


Figure 6. Strategies of the potential DR lesions extraction. Blue dots: H; yellow dots: HE or SE; red dots: MA.

where lesion type can be 1 (for no DR) or 2 (for MA), and $f(x) = 1$.

Image artifacts may influence the performance of MA detection because of the presence of dust or dirt. The morphology of these artifacts is similar to MA in terms of color and size. Hence, we filtered the images through an object-detection model to remove dust or dirt particles before producing the MA-enhanced images.

Finally, model $M4$ combines the binary lesion type information and features from the enhanced image for performance enhancement.

Data Analysis

We analyzed the performance of both the binary lesion type-classification model and the referable/nonreferable DR model for image-level recognition by calculating accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and

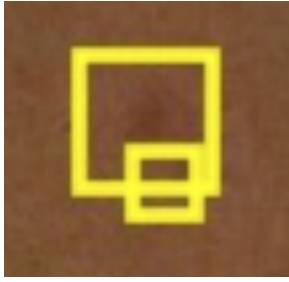


Figure 7. Closeup of MA. Example of IoU smaller than 0.15. The larger bounding box is produced by GT; the smaller bounding box is produced by a prediction model.

specificity. Note that in the binary lesion type classification, “true positive” denotes one of the locations of a predicted lesion having an IoU $>15\%$ compared to the GT location, “true negative” denotes both GT and prediction without any lesion detection, “false positive” denotes GT without any lesion detection but with prediction, and “false negative” denotes GT with at least one location but no prediction or any prediction location having an IoU $\leq 15\%$. The IoU should be larger than 15% to include the lesions because the pixel size of MA usually is within 3×3 , and the bounding box size of GT is around 7×7 .

Figure 7 shows an example of IoU that approximates 15%. The threshold definition of IoU is reasonable and rigorous because several studies published in the literature used at least one pixel overlaps with a GT^{25,26} that an image is considered to contain a target lesion. Moreover, the accuracy and weighted κ with Fleiss-Cohen κ coefficient weights²⁷ were calculated to evaluate the performances of the fusion architectures in five-class disease severity classification.

The benchmark data set, Messidor-2, was used to identify the performance of a hybrid model, which has the best results from five-class classification for binary classification (nonreferable DR versus referable DR), based on the calculation of accuracy, AUC, sensitivity, and specificity.

Results

For the private data set of 2043 images, the lesion-classification model detected the DR symptoms with an AUC greater than 81% for each symptom (Table 1). The sensitivity in detecting one of the symptoms was greater than 65% and the specificity in detecting the absence of the symptoms correctly was greater than 80%. This classification model detected the true negatives more than the true positives.

We also explored the effectiveness of four fusion models for DR grading. For the two- and five-class severity classification, a comparison between the baseline model (M0) and the proposed fusion models (M1–M4) is summarized in Table 2. The performance of M0 in terms of accuracy and weighted κ was 81.60% and 80.09%, respectively. The late fusion model M1 integrated the four major DR symptoms, slightly increasing the accuracy and weighted κ . The results of the early fusion models, M2 and M3, were similar; however, M2 decreased the misclassification rate at a severity level of mild NPDR and maintained the rate at moderate NPDR (data not shown). For the early detection of referrals, in M4, we combined the features from the lesion-classification (M1) and early fusion (M2) models with a regression model to obtain an output

Table 1. Performance of Binary Lesion Type-Classification Model at the Image Level

Lesion Type	Accuracy (%)	AUC (%)	Sensitivity (%)	Specificity (%)
MA	77.04 ^a	81.32	69.90	80.14
H	87.08	90.06	79.14	90.23
HE	79.59	82.26	64.57	83.95
SE	81.79	87.90	77.44	82.09

^aData are means.

Table 2. Performance Comparison of the Baseline Model (M0) and the Proposed Fusion Models (M1–M4)

Model	Five-Class		Two-Class			
	Accuracy	Weighted κ	Accuracy	AUC	Sensitivity	Specificity
M0	81.60	80.09	92.12	94.19	80.98	94.92
M1	81.69	81.19	92.22	95.08	82.20	94.73
M2	84.24	83.86	92.27	95.06	90.49	92.71
M3	85.12	84.43	91.09	94.21	90.98	91.12
M4	84.29	84.01	92.95	95.51	86.83	94.49

Table 3. Performance Comparison on Messidor-2 in Detecting Referable DR

Benchmark	Year	Training Data Set		Training with		Approach	Accuracy (%)	AUC (%)	Sensitivity (%)	Specificity (%)
		Mydiatic/	Nonmydiatic	Mydiatic/	Nonmydiatic					
Abramoff et al. ²⁸	2016	1,250,000 images (EyeCheck project and the University of Iowa)	Unknown	Unknown	Unknown	CNN	—	98.0	96.8	87.0
Gulshan et al. ¹⁰	2016	128,175 images (EyePACS)	Mixed	Mixed	Mixed	Inception-v3 CNN	—	99.0	87	98.5
Gargeya and Leng ^{29,a}	2017	75,137 images (EyePACS)	Mixed	Mixed	Mixed	CNN + gradient boosting classifier	—	94	93	87
Pires et al. ³⁰	2019	35,126 images (part of Kaggle data set)	Mixed	Mixed	Mixed	Similar to VGG-16	—	98.2	—	—
Voets et al. ²¹	2019	88,702 images (part of Kaggle data set)	Mixed	Mixed	Mixed	Inception-v3 CNN	84.21	85.30	68.70	88.50
Li et al. ^{31,b}	2019	19,233 images (Chinese hospitals)	Unknown	Unknown	Unknown	Inception-v3 CNN	93.49	99.05	96.93	93.45
Zago et al. ³²	2020	28 images with 262,144 patches per image (DiaretDB1)	Mydiatic	Mydiatic	Mydiatic	Patch-based CNN	—	94.4	90.0	87.0
Proposed model M4	2020	22,617 images (Taiwanese hospitals)	Mixed	Mixed	Mixed	Fusion CNN architecture	91.99	97.09	93.68	91.52

^aAny DR.^bOnly 800 images were selected from Messidor-2 to create the testing set.

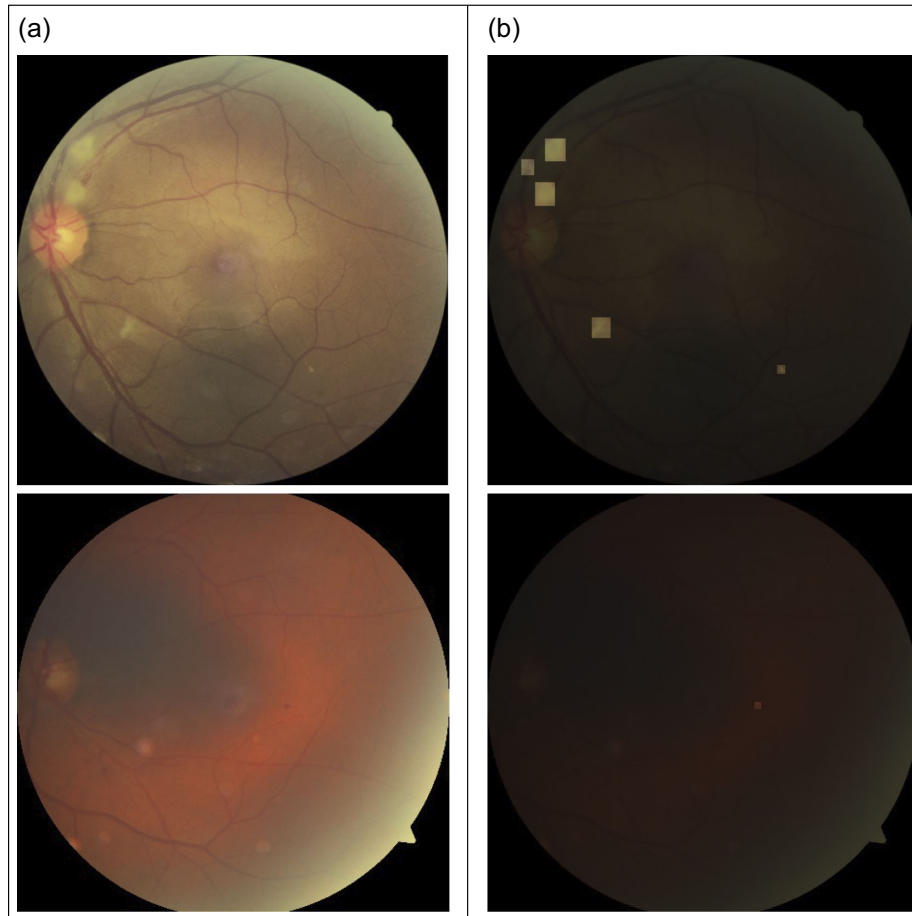


Figure 8. (a) Raw images. (b) Enhanced images.

accuracy of 92.95% and AUC of 95.51%, which is better than that of the other models. We also found that S2 yielded better results than S1 (data not shown).

M4 is the best-performing model as it produced the highest AUC when Messidor-2 was used to benchmark the performance of the state-of-the-art algorithms.^{10,21,28-32} In Table 3, M4 with an AUC of 97.09% has similar results to those presented in previous works. M4 also achieved a comparable sensitivity in detecting referable DR of 93.68% and a specificity in detecting nonreferable DR of 91.52%.

Discussion

A previous study achieved a weighted κ of 84% from a large training data set (1.6 million images).¹¹ We provide good baseline (M0) results from training on a smaller data set (22,000 images) with a weighted κ of 80%. To reduce the gap of training sample size, two-stage early fusion architectures enhanced the

performance of DR grading and achieved a similarly weighted κ of 84%, indicating that the lesion detection assistance was useful. The M4 hybrid model combines the lesion type classification and early fusion information, producing the best results in terms of sensitivity and specificity for detecting referral DR. Note that the incidence of soft exudates is relatively low in DR images compared to the other lesion types, and some of the small lesion features may disappear in the last convolutional layer. Thus, the lesion detection information is required to directly provide complementary information for the classifier.

As can be seen in the upper panel of Figure 8, soft exudates were highlighted by the enhancement algorithm; a small hemorrhage was also highlighted in the lower panel. Both enhanced images assisted M4 to classify the image more correctly to referable DR than M0 (original prediction class is nonreferral). Furthermore, we used a public data set, Messidor-2, obtained from France, to validate the proposed model in practical use. M4 also achieved performance on Messidor-2 comparable to that of the benchmark algorithms

without using the lesion information. In summary, these results validate that M4 performed equally well on both the private and public data sets in improving the overall performance of DR grading.

The proposed strategy mimics the evaluation process of ophthalmologists, in which the fundus image is inspected to identify suspicious entities (lesion types/locations) and then classified. This hybrid process combines candidate lesion features and whole-image deep learning features, which increases the overall performance for DR grading. Moreover, although background pigmentation varies across races and ethnicities and may hinder diagnosis, the DR signs are immutable.^{7,33} Our architectures were trained without using a transfer-learning model; they were trained based solely on the Asian fundus images and obtained a robust performance on both test data sets (AUC of 95.51% for Asian, 97.09% for Messidor-2). Hence, the proposed architectures can be combined with well-trained DR signs to become more highly applicable to different ethnicities. This result is similar to the findings in Li et al.³⁴ Furthermore, both nonmydriatic and mydriatic images were used for training to demonstrate a generalized application of the proposed architectures. Instead of using a time-consuming patch-based method, early fusion efficiently decreased the inference time for lesion detection in supporting DR grading.

Improving the misclassification rate in the early stage of DR is essential for clinical management and preventing patient vision loss in the future. A minor visual change between the mild and moderate severity stages assessed with a fundus photograph or optical coherence tomography, such as MA, intraretinal hemorrhages, or small hard drusen, may be overestimated or underestimated even by experienced ophthalmologists.^{11,28} For example, the pixels of MA are less than 0.002% of the image. Furthermore, the image artifacts are sometimes similar to MA. Consequently, the intergrader variability is well known, with a lower κ ,¹¹ which affects the performance of the CNN model as well. Accordingly, we developed the fusion architectures and combined the lesion information using the CNN model for DR grading. This may compensate for the information loss during the computation of the convolutional layers of the CNN model.

A limitation of our study is that the fusion architectures excluded information on neovascularization, which is an important feature in the class of PDR. This feature was not trained because sparse data were marked as neovascularization. In addition, the performance improvement of the late fusion architecture was unclear. This finding was unexpected and suggests that

there may have been overlapping features between the baseline and the lesion-classification models.

Future work will include adjusting different weighting methods or modifying the losses from both the image enhancement classifier and the raw image classifier by using a controlled hyperparameter. Moreover, longitudinal image data make DR prediction more accurate and objective; this has some potential that should be explored further.

In conclusion, we have developed fusion architectures that combine lesion information with disease severity classification. The M4 hybrid model performed well on Messidor-2 when compared with state-of-the-art algorithms without lesion detection information. Thus, we believe that M4 will assist frontline health care providers in efficiently highlighting lesion information and classifying DR severity and can be considered a representative model for general use.

Acknowledgments

The authors thank Tri-Service General Hospital, Chung Shan Medical University Hospital, and China Medical University Hospital in Taiwan for providing invaluable in situ data for this study.

Supported by the Industrial Technology Research Institute, Hsinchu, Taiwan, for the fellowship of “Decision Support Technology of Fundus Image in Diabetes Mellitus” (grant J367B82210). The sponsor or funding organization participated in the design of the study, conducting the study, data collection, data management, data analysis, interpretation of the data, preparation, review, and approval of the manuscript.

Disclosure: **V.S. Tseng**, None; **C.-L. Chen**, None; **C.-M. Liang**, None; **M.-C. Tai**, None; **J.-T. Liu**, None; **P.-Y. Wu**, None; **M.-S. Deng**, None; **Y.-W. Lee**, None; **T.-Y. Huang**, None; **Y.-H. Chen**, None

References

1. American Academy of Ophthalmology Retina/Vitreous Panel. *Preferred Practice Pattern Guidelines. Diabetic Retinopathy*. San Francisco, CA: American Academy of Ophthalmology; 2017.
2. Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye Vis*. 2015;2:1–25.
3. Leasher JL, Bourne RR, Flaxman SR, et al. Global estimates on the number of people blind or visually

- impaired by diabetic retinopathy: a meta-analysis from 1990 to 2010. *Diabetes Care*. 2016;39:1643–1649.
4. Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol*. 2019;7:140–149.
 5. Yu NC, Chen IC. A decade of diabetes care in Taiwan. *Diabetes Res Clin Pract*. 2014;106:S305–S308.
 6. Shah K, Gandhi A, Natarajan S. Diabetic retinopathy awareness and associations with multiple comorbidities: insights from DIAMOND study. *Indian J Endocrinol Metab*. 2018;22:30–35.
 7. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1:39.
 8. Lin GM, Chen MJ, Yeh CH, et al. Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy. *J Ophthalmol*. 2018;2159702:1–6.
 9. Besenczi R, Tóth J, Hajdu A. A review on automatic analysis techniques for color fundus photographs. *Comput Struct Biotechnol J*. 2016;14:371–384.
 10. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–2410.
 11. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125:1264–1272.
 12. Ministry of Health and Welfare. Number of medical staff in medical institutions over the years. Available at: <https://iiqsw.mohw.gov.tw/InteractiveIntro.aspx?TID=93D49587A7935C04>. Accessed February 28, 2019.
 13. Ministry of Health and Welfare. National health insurance medical statistics. Available at: <https://dep.mohw.gov.tw/DOS/np-1918-113.html>. Accessed February 28, 2019.
 14. Lafta R, Zhang J, Tao X, et al. An intelligent recommender system based on predictive analysis in telehealthcare environment. *Web Intell*. 2016;14:325–336.
 15. Wilkinson C, Ferris FI, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110:1677–1682.
 16. Yang Y, Li T, Li W, Wu H, Fan W, Zhang W. Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, (Eds.). *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*. Vol. 10435. Quebec City, Canada: Springer; 2017:533–540.
 17. Wang Z, Yin Y, Shi J, Fang W, Li H, Wang X. Zoom-in-Net: deep mining lesions for diabetic retinopathy detection. *Lect Notes Comput Sci*. 2017;10435 LNCS:267–275.
 18. Sabokrou M, Fayyaz M, Fathy M, Moayed Z, Klette R. Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput Vis Image Underst*. 2018;172:88–97.
 19. Lam C, Yu C, Huang L, Rubin D. Retinal lesion detection with deep learning using image patches. *Investig Ophthalmol Vis Sci*. 2018;59:590–596.
 20. Abramoff MD. Messidor-2 dataset. Available at: <https://medicine.uiowa.edu/eye/abramoff>. Accessed February 28, 2019.
 21. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PLoS One*. 2019;14:1–11.
 22. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Thirty-first AAAI conference on artificial intelligence*. 2017;4278–4284.
 23. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017;4700–4708.
 24. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Machine Intell*. 2018;42:318–327, DOI: 10.1109/TPAMI.2018.2858826.
 25. Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. *Med Image Anal*. 2018;39:178–193.
 26. Chudzik P, Majumdar S, Calivá F, Al-Diri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. *Comput Methods Programs Biomed*. 2018;158:185–192.
 27. Joseph LF, Jacob C. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33:613–619.
 28. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, Niemeijer M. Improved automated detection of diabetic retinopathy on

- a publicly available dataset through integration of deep learning. *Investig Ophthalmol Vis Sci*. 2016;57:5200–5206.
29. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124:962–969.
 30. Pires R, Avila S, Wainer S, Valle E, Abràmoff MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. *Artif Intell Med*. 2019;96:93–106.
 31. Li F, Liu Z, Chen H, Jiang M, Zhang X, Wu Z. Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. *Transl Vis Sci Technol*. 2019;8:4.
 32. Zago GT, Andreão RV, Dorizzi B, Salles EOT. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Comput Biol Med*. 2020;116:103537.
 33. Raman R, Srinivasan S, Virmani S, Sivaprasad S, Rao C, Rajalakshmi R. Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye*. 2019;33:97–109.
 34. Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. *Diabetes Care*. 2018;41:2509–2516.