Scientific Reports

https://doi.org/10.1038/s41598-026-43573-x

Article in Press

BigEye: a clinically interpretable deep learning framework for diabetic retinopathy detection and stage prediction

Received: 4 November 2025

Accepted: 5 March 2026

Published online: 08 March 2026

Cite this article as: Gill H.M., Salem D.H., Omoru O.B. *et al.* BigEye: a clinically interpretable deep learning framework for diabetic retinopathy detection and stage prediction. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-43573-x

Hunter Mathias Gill, Doaa Hassan Salem, Okiemute Beatrice Omoru, Frank Dash Bogan, Jeffrey Xiao Liu, Michael Happe, Amir Reza Hajrasouliha & Sarath Chandra Janga

BigEye: a Clinically Interpretable Deep Learning Framework for Diabetic Retinopathy Detection and Stage Prediction

AUTHORS:

Hunter Mathias Gill [1]

Doaa Hassan Salem, Ph.D. [1,5]

Okiemute Beatrice Omoru [1]

Frank Dash Bogan [2]

Jeffrey Xiao Liu [2]

Michael Happe [2]

Amir Reza Hajrasouliha, M.D. [2]

Sarath Chandra Janga, Ph.D. [1, 3, 4 *]

AFFILIATIONS:

1. Department of Biomedical Engineering & Informatics, Luddy School of Informatics, Computing and Engineering, Indiana University Indianapolis, 535 West Michigan Street, Indianapolis, Indiana 46202

2. Department of Ophthalmology, Eugene & Marilyn Glick Eye Institute, Indiana University School of Medicine, 1160 W. Michigan Street Indianapolis, Indiana, 46202

3. Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, 975 West Walnut Street, Indianapolis, Indiana, 46202

4. Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), 410 West 10th Street, Indianapolis, Indiana, 46202

5. Computers and Systems Department, National Telecommunication Institute, 5,ش Mahmoud El-Meligy Cairo, Egypt

* corresponding author

Correspondence should be addressed to:

Sarath Chandra Janga, Ph.D.

535 West Michigan Street

Indianapolis, IN 46202

scjanga@iu.edu

ABSTRACT

Diabetic Retinopathy (DR) is a major cause of vision loss and blindness in diabetic individuals. DR is conventionally diagnosed by assessing retinal lesion findings from fundus photographs taken during exams and applying a scale like International Classification of Diabetic Retinopathy (ICDR). The expected rise in future DR cases highlights the need for deep learning models capable of identifying relevant lesions and delivering explainable results. To this end we present BigEye, a novel framework that uses extracted lesion features to predict ICDR stage. A dataset of fundus images from a local hospital and a public dataset, annotated with segmentation masks and DR stages, is assembled to train a DeepLabV3+ model on six retinal lesions. Lesion quantities and pixel area features are integrated by a classifier model evaluated through 10-fold nested cross validation (0.77 ± 0.07 precision, 0.71 ± 0.06 recall, 0.72 ± 0.07 F1 score, 0.95 ± 0.02 ROC-AUC, 0.83 ± 0.03 accuracy). A Shapely Additive Explanations (SHAP) value analysis notably shows close alignment between discriminative lesions for each DR stage and corresponding ICDR stage criteria. These results demonstrate that BigEye is well suited for providing explainable ICDR stage predictions grounded in clinical knowledge.

Keywords: Diabetic Retinopathy, DR, Retinal Lesions, DR Staging, Explainable AI

INTRODUCTION

Diabetic Retinopathy (DR)

Diabetic retinopathy (DR) is a vision-threatening complication of diabetes mellitus (DM) and a leading cause of blindness across the world [1] [2]. Meta-analyses of global DR prevalence indicate approximately 20% of individuals with DM are affected by DR, highlighting the scope and multifactorial nature of diabetes-mediated vision decline. The duration of diabetes is a significant predictor of DR with 20-30% of DM cohorts developing DR within 10 years [3] [4] [5] [6] [7] [8]; other factors like age [9] [10], ethnicity [11], HbA1C levels [8], and comorbidities such as high blood pressure [9] or renal disease [12] are also significant contributors to DR risk. Symptoms of DR including blurred vision and dark spots reflect underlying damage to the retina.

The retina, a network of photoreceptor and ganglion cells at the back of the eye, enables vision by transducing incoming light into electrical impulses and transmitting those impulses to the visual cortex of the brain [13]. The high metabolic demand of transduction is met by an extensive retinal microvascular network of arterioles, venules, and capillaries [14] [15]. While the microvasculature ordinarily maintains retinal health by providing a constant supply of oxygenated and nutrient-rich blood, this role is subverted in DR through microscopic changes to vessel structures.

The characteristic microvascular deterioration observed in DR stems from the breakdown of the inner blood-retina barrier (iBRB). The iBRB is the interface between the inner retinal layers and the microvascular network and controls the movement of substances between them. This barrier is composed of retinal capillary endothelial cells joined together by tight junctions and reinforced by pericytes [16]. DR is initiated when chronic hyperglycemia-induced disturbances in molecular pathways result in

iBRB dysfunction via pericyte dropout [17] [18] [19] **(Supplementary Fig. 1)**. The loss of retinal capillary pericytes precipitates further damage to the retina.

The Significance of Retinal Lesions in DR

Injuries to the retina manifest as a set of aberrant histological structures called retinal lesions **(Fig. 1)**. Retinal lesions indicate neurovascular damage in one or more retinal layers and contribute to impaired visual acuity in DR patients [20] [21]. Many lesions are visible through digital photographs taken by specialized fundus cameras. These fundus images are routinely obtained during ophthalmological exams and are gold standard modalities for DR detection [22]. The following sections describe the roles of retinal lesions in DR pathophysiology, their appearance in fundus images, and their impacts on vision.

Microaneurysm lesions are one of the first clinical signs of microvascular damage in DR [23]. Microaneurysms develop as small, focal outpouchings in pericyte-deficient areas of capillary walls [24] and present as small dots in fundus images. Microaneurysms increase capillary permeability and act as origin points for fluid egression [24]. High numbers of microaneurysms are associated with worsening DR and emergence of other retinal lesions [25]

Hemorrhages are areas of bleeding from retinal blood vessels. Symptoms of vision impairment including blurred vision, dark spots, and floaters are attributable to bleeding in different retinal layers. Flame hemorrhages emerge in the nerve fiber layer (NFL), while dot and blot hemorrhages develop from rupturing microaneurysms in deep retinal capillaries [26] [27].

Exudates are outer plexiform layer clusters of lipid-filled macrophages and extracellular proteins and lipids that are deposited through capillary leakages [28]. They

appear in fundus images as focal aggregations or circinate rings of yellow-white, waxy plaques. Exudates are associated with decreased retinal sensitivity [29] and facilitate irreversible central vision loss when present in the fovea.

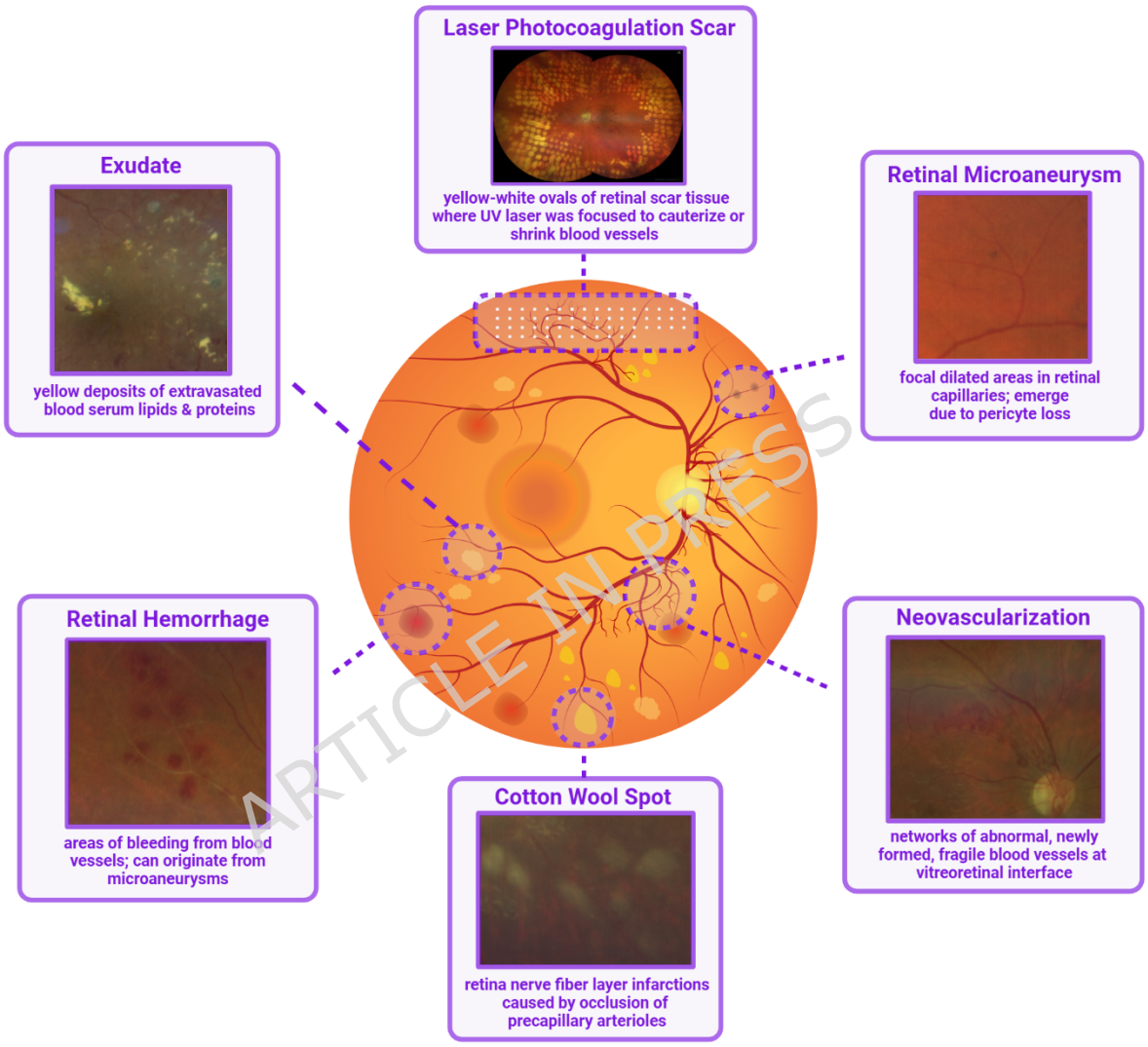


Fig. 1

Cotton wool spots are infarcts in the nerve fiber layer resulting from retinal arteriole occlusion and consequent ischemia. The ischemic conditions facilitate focal axoplasmic flow stasis, leading to distinguishing buildups of cytoid bodies (accumulations of mitochondria and other sub-cellular material) [30]. In fundus images cotton wool spots appear as small, wispy, white patches. Cotton wool spots are associated with loss of retinal ganglion cells and concomitant scotoma development [30,31] .

Neovascularization describes the abnormal growth of new retinal blood vessels [32]. Hypoxic conditions secondary to retinal blood vessel occlusions and leakages stimulate the production of vascular endothelial growth factor (VEGF) [33]. VEGF promotes the formation of new vessels at the vitreoretinal interface by mediating endothelial cell proliferation and migration; these irregular new blood vessels are fragile and prone to bleeding into the vitreous humor [34]. Fibrovascular scar tissue forms alongside vessels that bleed extensively [32]. Fibrovascular scars pull on the retina and may eventually separate it from the retinal pigment epithelium, leading to tractional retinal detachment, a serious condition that can rapidly result in blindness [35 36].

Not all retinal changes are attributable to DR pathophysiology, as treatment modalities can also leave marks on the retina. Laser photocoagulation is an eye surgery in which a laser beam is directed onto certain areas of the retina in order to ablate or destroy abnormal tissues [3738]. The destruction of ischemic retinal tissue through focused laser beams reduces VEGF production and increases the diffusion of oxygen from the choroid to the inner retina [39], impeding the development of additional retinal lesions. Laser photocoagulation is performed in either concentrated focal spots or in a pattern scanning laser (PASCAL) grid (*37*) and is often combined

with intravitreous injection of anti-VEGF agents [40] [41] [42] [43]. Areas affected by laser treatment undergo protein denaturation and appear in fundus images as yellow-white ovals; older laser scars also feature black pigmentation near the center.

Retinal Lesions and DR Stages

Retinal lesions emerge at different junctures in DR. Information with respect to the presence, quantities, and locations of certain lesions can be leveraged to systematically describe the progression of the disease. DR is broadly divided into non-proliferative DR (NPDR, a.k.a. background retinopathy) and proliferative DR (PDR) categories by the absence or presence of neovascularization, respectively. The International Classification of Diabetic Retinopathy (ICDR) scale extends these categories by establishing criteria **(Table 1)** for five DR stages: (1) no apparent retinopathy (no DR), (2) mild NPDR, (3) moderate NPDR, (4) severe NPDR, (5) PDR [44]. The widespread use of ICDR in DR monitoring and treatment strategies highlights the value of methods that deliver accurate stage predictions.

With the high prevalence of DR among diabetics, early detection is crucial as timely intervention can prevent vision loss. Advances in artificial intelligence have galvanized the development of specialized deep learning (DL) models trained to identify signs of DR. DL models capable of achieving parity with ophthalmologists in DR detection and staging tasks show considerable promise in contexts like clinical decision support [45], telemedicine [46] and DR screening programs. To date, the United States Food and Drug Administration (FDA) has approved three DL models for point-of-care clinical DR screening: (1) IDx-DR, (2) EyeArt, and (3) AEYE Diagnostic Screening (AEYE-DS) [46]. Similar DL models leverage image-level features or clinically relevant retinal lesions

to predict DR stages. The approved models and those with similar architectures work by identifying pixel-level patterns, segmentation of clinically relevant image features, or a combination of both.

ICDR Stage	Criteria
No DR	no apparent retinal lesions
mild NPDR	microaneurysms in any fundus region
moderate NPDR	microaneurysms AND exudates, hemorrhages, and / or cotton wool spots in any fundus region
severe NPDR	One of the following: (1) more than 20 microaneurysms in all four fundus quadrants, (2) venuous beading in two or more quadrants, (3) intraretinal microvascular abnormality (IRMA) in one or more quadrants
PDR	previously observed lesions AND vitreous/preretinal hemorrhage, neovascularization, and/or pan-retinal laser photocoagulation scars

Table 1

Several studies demonstrate the utility of ImageNet transfer learning (TL) - based classification models in predicting DR stages [47], particularly in cases where available data is limited or imbalanced [48]. One study by Narayanan et al. assessed AlexNet, VGG16, ResNet, and InceptionV3 models for DR stage classification on the APTOS-2019 dataset. 10-fold cross- validation of the models showed decent accuracies (Alexnet 71.8±6.1%, VGG16 71.7±5.3%, ResNet 72.1±5.0%, InceptionV3 73.2±4.3%); however, the authors found the models yielded better results when used as feature extractors for a SVM-based classifier (Alexnet 80.5±6.2%, VGG16 82.3±5.7%, Resnet 82.3±5.7%, InceptionV3 83.7±6.0%) [49]. Other studies similarly employ TL networks as feature extraction blocks. Work by Chilukoti et.al. examined a range of VGG-16, ResNet-101, and EfficientNet architectures on the EyePACS, APTOS 2019, and MESSIDOR datasets. They demonstrated strong performances with average accuracies ranging from 0.721 (ResNet-152 and EfficientNet-B5) to 0.92 (EfficientNet-B3_En) [50]. Another group applied a ResNet-50 classifier on a fundus image subset from the Danish OPTIMISM clinical study and reported stage – specific accuracies ranging from 80% - 96% [51]. Further studies indicate the utility of transformer architectures in enhancing TL performance. Experiments by Gu et.al show that a vision transformer-based TL model achieves ~ 91% accuracy on the DDR dataset [52] while another work based on the FunSwin transformer reports similar results on MESSIDOR [53].

Although TL models generally achieve excellent results in DR stage prediction, their interpretability is restricted by an inability to directly incorporate retinal

lesion information. In contrast, segmentation models like fully connected network (FCN), U-Net, BiSeNet, and DeepLabV3+ are capable of extracting lesions from fundus images. Some studies apply these models for binary segmentation of specific retinal lesions such as exudates [54] or laser photocoagulation scars [55] whereas others perform multiclass segmentation to capture a range of lesions. The scale of different retinal lesions presents a complication to multiclass segmentation models, as some lesion types are several magnitudes of order larger in area than others (e.g. hemorrhages or laser photocoagulation scars versus microaneurysms). This challenge is addressed through specialized model units, weights, and loss functions. A BiSeNet model proposed by Gao et al. introduces a ghost convolution unit to improve segmentation accuracy of small lesions like microaneurysms [56]. A separate work introduces a multiscale feature fused (MSFF) block to the U-Net architecture to better capture fundus image features at different scales. The modification improves the intersection-over-union (IoU) scores for all considered lesions [57].

Data derived from segmentation predictions permit finer control in feature engineering-based approaches to DR stage prediction. One study applied a fully connected network (FCN) to segment areas of neovascularization; the combined fundus images and segmentation probability maps were then used to predict NPDR and PDR categories with 87.7% accuracy [58]. A separate analysis featured a similar protocol, using fundus images with overlaid microaneurysm, hemorrhage, exudate, cotton wool spot, intraretinal

microvascular abnormalities (IRMA), and photocoagulation scar segments as inputs for a DR stage classification model (70.4% accuracy across ICDR stages) [59]. A hybrid approach previously developed by our group also demonstrated that combining TL-extracted image features with microaneurysm, exudate and hemorrhage lesion data boosts DR classification performance [60]. The DeepDR joint learning model developed by Dai et al. employs a Mask- RCNN subnetwork to segment hemorrhages, exudates, and cotton wool spots (IoU values 0.97, 0.97, and 0.94 respectively) and uses the segmentation probability maps to train a DR stage classification subnetwork (0.95 ROC-AUC for overall stage prediction) [61].

RESULTS

BigEye Leverages Retinal Lesion Information to Predict DR Stage

Diabetic retinopathy (DR) is a progressive eye disease characterized by the breakdown of the inner blood-retina barrier and concomitant vision loss and blindness. At-risk individuals are monitored through non-invasive fundus photography that directly images the retina. Fundus imaging provides a snapshot of eye health and is useful for detecting retinal lesions (areas of abnormal tissue resulting from microvascular damage) that act as indicators of DR. Certain retinal lesions are associated with worsening DR; accordingly, information regarding different retinal lesions is incorporated by systems like the International Classification of Diabetic Retinopathy (ICDR) scale to formulate DR stages (**Table 1**). Although DR staging is traditionally done by

ophthalmologists, the projected rise in diabetes patients highlights the need for deep learning (DL) models that can process fundus images and predict DR stages using a similar mechanism to ICDR criteria. Transfer learning (TL)-based classification models perform well in DR staging; however, they are regarded as “black boxes” and may be difficult to interpret. In contrast, segmentation-based approaches can be built around clinical knowledge by extracting retinal lesions that are significant hallmarks of DR progression.

BigEye (https://github.iu.com/Janga-Lab/BigEye/) is a DL framework that uses retinal lesion measurements to classify DR stage. The framework (**Fig. 2**) employs a model chaining approach based on two main components: (1) a DeepLabV3+ segmentation model that extracts microaneurysm, hemorrhage, exudate, cotton wool spot, neovascularization, and laser photocoagulation scar lesions from fundus image inputs, and (2) a classification model that uses lesion quantities and pixel areas computed from the predicted segmentation masks to output DR stage. This transparent approach facilitates classification explainability as model decisions are exclusively based on the presence and numbers of retinal lesions. The importance of model trustworthiness and interpretability in ophthalmological healthcare [62] reflects the benefit of incorporating practical retinal lesion features in DR stage predictions.

Composition of Fundus Image Dataset

BigEye is developed and evaluated on a dataset of regular (133° FOV) and widefield (200° FOV) color fundus images (**Fig. 3**). The dataset (n = 602) represents five ICDR categories: (1) no DR (n= 316), (2) mild NPDR (n = 134),

(3) moderate NPDR (n = 62), severe NPDR (n = 44), and PDR (n = 46) and is a composite of two separate sources: (1) an internal Indiana University Health (IUH) dataset and (2) the public e-ophtha dataset [62]. The deidentified IUH samples (n = 164) were obtained from a retrospective study. After filtering, the IUH dataset

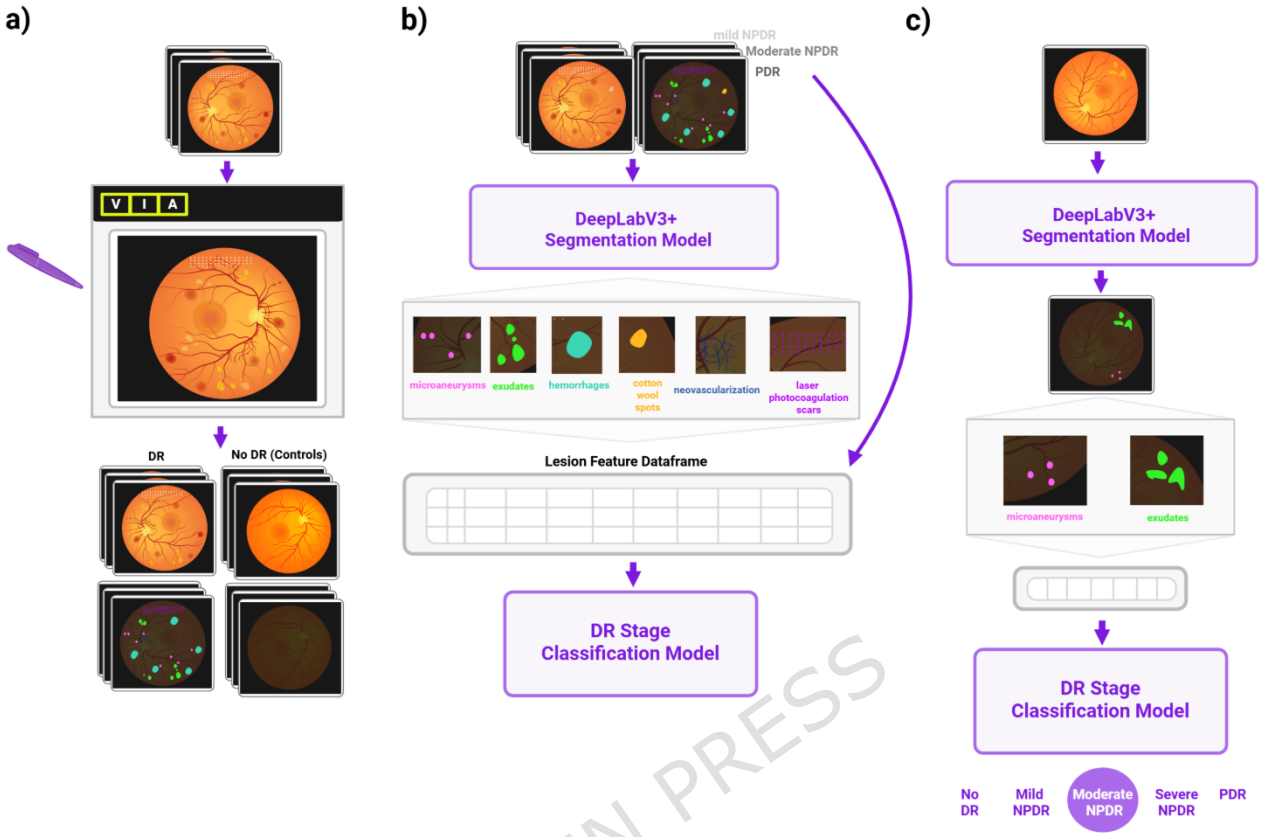


Fig. 2

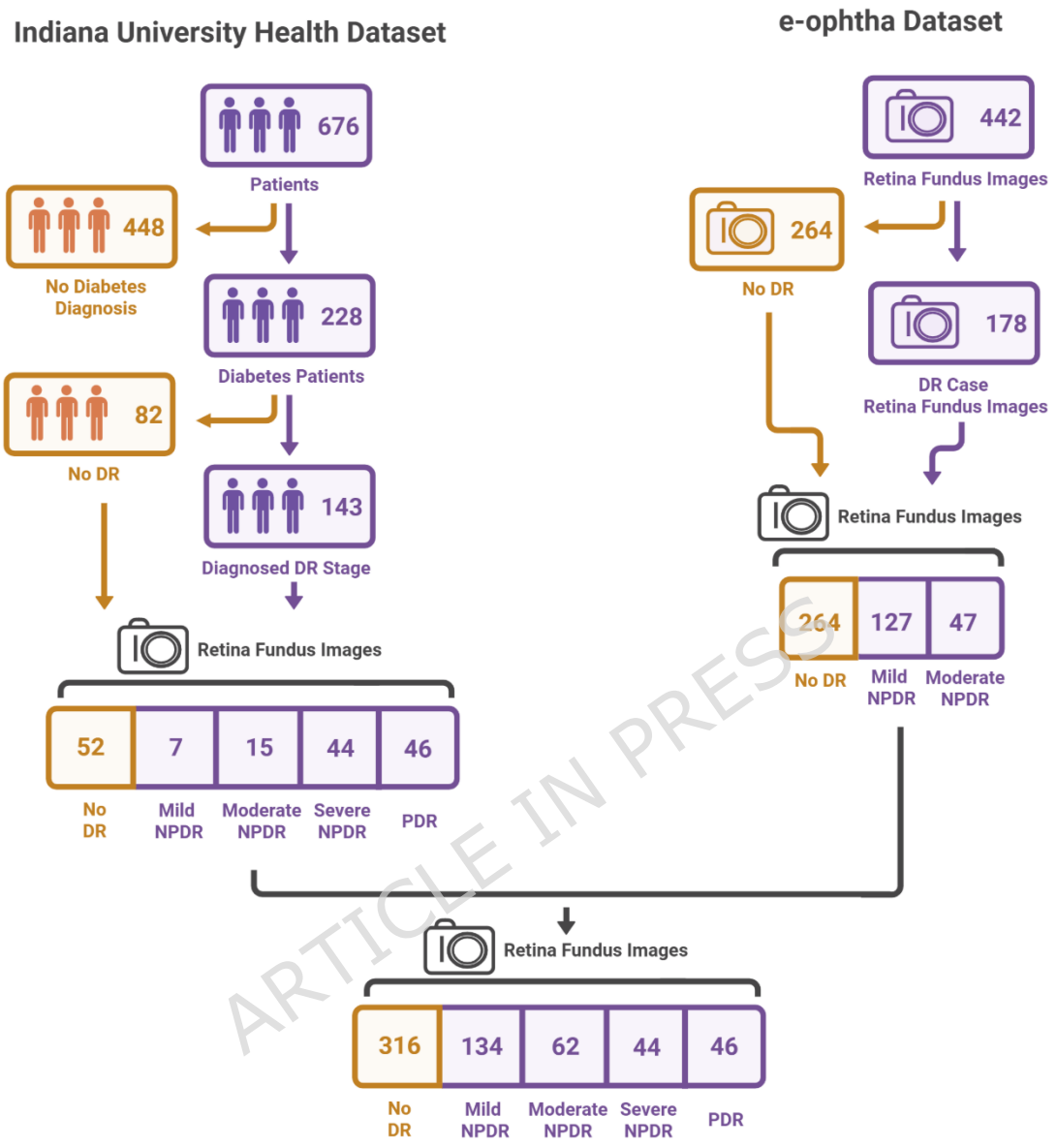


Fig. 3

amounts to 52 controls (no DR), 7 mild NPDR, 15 moderate NPDR, 44 severe NPDR, and 46 PDR. All e-ophtha samples (n = 438) originate from the French OPHDIAT® telemedicine network [62] and represent 264 control (no DR), 127 mild NPDR, and 47 moderate NPDR images.

After collection, all fundus images are annotated for DR lesions. The VGG Image Annotator (VIA) tool [63] is used to create one multiclass segmentation mask for all IUH images. Exudate and microaneurysms segmentation masks for the e-ophtha samples are downloaded directly and modified where necessary (merging binary lesion annotations, adding extra annotations with VIA) to create one segmentation mask per image. The combined fundus images and segmentation masks are used to train the BigEye DeepLabV3+ retinal lesion segmentation model.

BigEye Segmentation Model Demonstrates Robust Performance in Retinal Lesion Detection

The BigEye segmentation model serves as an integral part of the BigEye workflow by providing retinal lesion segmentations that are processed downstream for DR stage prediction. The DeepLabV3+ model achieves good average performances across retinal lesions, with a cumulative test intersection-over-union (IoU) value of 0.56 ± 0.11 (**Table 2**). We note that the micro-averaged precision, recall, and F1 scores are higher than ROC-AUC. Metrics like precision, recall, and F1 are evaluated at a single classification threshold (e.g. 0.5). At this threshold, the DeepLabV3+ model correctly predicts nearly all the background pixel class. Since the background class is

the most numerous, the standard precision, recall, and F1 scores are inflated to be high values before even considering the lesion classes. In contrast, the ROC-AUC score measures the segmentation model's discriminative capability across all decision thresholds, including those where the model is less confident, leading to a score of ~0.70. To provide a more balanced perspective on model performance, we also provide macro-averaged versions of the precision, recall, and F1 score metrics. In macro-averaging, the metric is computed separately for each class before taking their unweighted average, giving equal weight to minority lesion classes regardless of how few pixels they contain. The macro-averaged precision (0.71 ± 0.08), recall (0.55± 0.11), and F1 score (0.62 ± 0.09) values are more closely aligned to the ROC-AUC score and together offer a more balanced assessment of model performance on minority retinal lesion classes. The precision, recall, F1 score, accuracy, and ROC-AUC metrics stabilize near 50 epochs of training, while IoU begins to level off at epoch 125 (**Supplementary Fig. 2**). Training and validation loss curves do not significantly diverge, indicating model robustness and generalizability. A lesion-wise IoU score breakdown shows notable scores with little training-validation gap for laser photocoagulation scars, cotton wool spots, and the background (non-lesion portions of fundus images), while the microaneurysm and hemorrhage scores are lower with larger gaps between training and validation set performances (**Supplementary Fig. 2**). Predictions taken at various time points indicate increasing skill during model training (**Supplementary Fig. 3**).

	Subset		
Metric	**Training**	**Validation**	**Test**
Precision	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
Precision (Macro)	0.89 ± 0.01	0.85 ± 0.08	0.71 ± 0.08
Recall	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
Recall (Macro)	0.87 ± 0.01	0.72 ± 0.07	0.55± 0.11
F1 Score	0.99±0.00	0.98 ± 0.01	0.98 ± 0.01
F1 Score (macro)	0.89 ± 0.01	0.78 ± 0.07	0.62 ± 0.09
Accuracy	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
ROC-AUC	0.69 ± 0.01	0.71 ± 0.01	0.69 ± 0.01
IoU (total)	0.81 ± 0.01	0.68 ± 0.06	0.56 ± 0.11
Loss	0.03 ± 0.00	0.08 ± 0.04	0.10 ± 0.09

Table 2

BigEye Classification Model Shows Notable DR Stage Prediction Performance

The BigEye classification model uses tabulated retinal lesion quantities and pixel areas to predict DR stages. Nested 10-fold cross-validation of the BigEye dataset is used to compare performances of six classification models: (1) support vector machine (SVM), (2) random forest (RF), (3) simple neural network (NN), (4) extreme gradient boosting (XGB), (5) light gradient boosting machine (LGBM), and (6) TabNet (**Table 3**). Accuracy scores on the test set show that the tree-based approaches (0.83 ± 0.05 RF, 0.84 ± 0.04 XGB, 0.83 ± 0.03 LGBM) outperformed the SVM (0.80 ± 0.04) and neural network-based classifiers (0.80 ± 0.04 NN, 0.80 ± 0.04 TabNet). Given its accuracy score and relatively small train – validation – test set loss gap, the LGBM model is nominated as the official BigEye classification model given its overall reliable performance. LGBM performance is similar across different folds (**Supplementary Fig. 4**), and the model correctly classifies most examples of DR stages (with moderate NPDR showing the overall weakest performance at 56% correctly predicted cases (**Fig. 4**).

Classification Model SHAP Analysis Reflects DR Clinical Insights

A Shapely additive explanations (SHAP) value analysis is conducted on the LGBM model to determine retinal lesion feature importances for different DR stages (**Fig. 5**). For control (no DR) samples, high values for all lesion features had a large negative impact (~ 1 SHAP value) on model performance, while low values had a moderate positive impact (-1 – -2). In mild NPDR samples,

microaneurysm quantity is the most significant predictor with positive SHAP values. Higher microaneurysm quantities correlate with greater SHAP value, while lower microaneurysm quantities are associated with lower positive or small negative SHAP values.

High quantities for other lesion values like cotton wool spots, exudates, laser photocoagulation scars are negative predictors for mild NPDR, while those for hemorrhages show a more ambivalent effect. Moderate NPDR samples show that high microaneurysm and exudate counts have a strong positive impact (+2). Lower numbers of cotton wool spots positively impact model performance on moderate NPDR (0.5) while larger numbers have a large negative impact (-0.5). High cotton wool spot and hemorrhage counts are strong indicators of severe NPDR (4 and 2, respectively), while high predicted quantities of microaneurysms, laser photocoagulation scars, exudates, and neovascularization are negatively associated (-0.5 – -1). Finally, high quantities of laser photocoagulation scars have a strong positive impact on PDR.

	Training					
Classifier	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**ROC-AUC**	**Loss**
SVM	0.82 ± 0.01	0.76 ± 0.02	0.78 ± 0.02	0.83 ± 0.00	0.93 ± 0.00	NA
RF	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	NA
NN	0.82 ± 0.09	0.57 ± 0.19	0.66 ± 0.19	0.75 ± 0.04	0.93 ± 0.04	0.78 ± 0.25
XGB	0.95 ± 0.01	0.92 ± 0.01	0.94 ± 0.01	0.94 ± 0.00	0.99 ± 0.00	0.25 ± 0.03
LGBM	0.86 ± 0.02	0.82 ± 0.02	0.84 ± 0.02	0.88 ± 0.00	0.97 ± 0.00	0.39 ± 0.02
TabNet	0.85 ± 0.02	0.75 ± 0.03	0.79 ± 0.02	0.80 ± 0.02	0.96 ± 0.01	0.57 ± 0.05
	Validation					
Classifier	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**ROC-AUC**	**Loss**
SVM	0.77 ± 0.10	0.70 ± 0.08	0.71 ± 0.09	0.81 ± 0.05	0.91 ± 0.03	NA
RF	0.78 ± 0.07	0.74 ± 0.06	0.75 ± 0.07	0.83 ± 0.04	0.95 ± 0.02	NA
NN	0.87 ± 0.06	0.62 ± 0.21	0.70 ± 0.20	0.82 ± 0.04	0.95 ± 0.03	0.67 ± 0.22
XGB	0.81 ± 0.08	0.78 ± 0.07	0.77 ± 0.07	0.85 ± 0.04	0.96 ± 0.02	0.47 ± 0.10
LGBM	0.78 ± 0.08	0.72 ± 0.07	0.73 ± 0.08	0.83 ± 0.04	0.95 ± 0.02	0.53 ± 0.09
TabNet	0.88 ± 0.05	0.72 ± 0.06	0.79 ± 0.05	0.83 ± 0.04	0.96 ± 0.02	0.64 ± 0.10
	Test					
Classifier	**Precision**	**Recall**	**F1 Score**	**Accuracy**	**ROC-AUC**	**Loss**
SVM	0.76 ± 0.10	0.70 ± 0.07	0.71 ± 0.08	0.80 ± 0.04	0.91 ± 0.04	NA
RF	0.79 ± 0.07	0.76 ± 0.07	0.76 ± 0.07	0.83 ± 0.05	0.95 ± 0.02	NA
NN	0.86 ± 0.04	0.74 ± 0.05	0.79 ± 0.04	0.80 ± 0.04	0.96 ± 0.01	0.55 ± 0.10
XGB	0.79 ± 0.07	0.76 ± 0.06	0.77 ± 0.07	0.84 ± 0.04	0.96 ± 0.01	0.49 ± 0.08
LGBM	0.77 ± 0.07	0.71 ± 0.06	0.72 ± 0.07	0.83 ± 0.03	0.95 ± 0.02	0.54 ± 0.07
TabNet	0.86 ± 0.04	0.75 ± 0.04	0.80 ± 0.04	0.80 ± 0.04	0.95 ± 0.02	0.63 ± 0.10

Table 3

Fig. 4

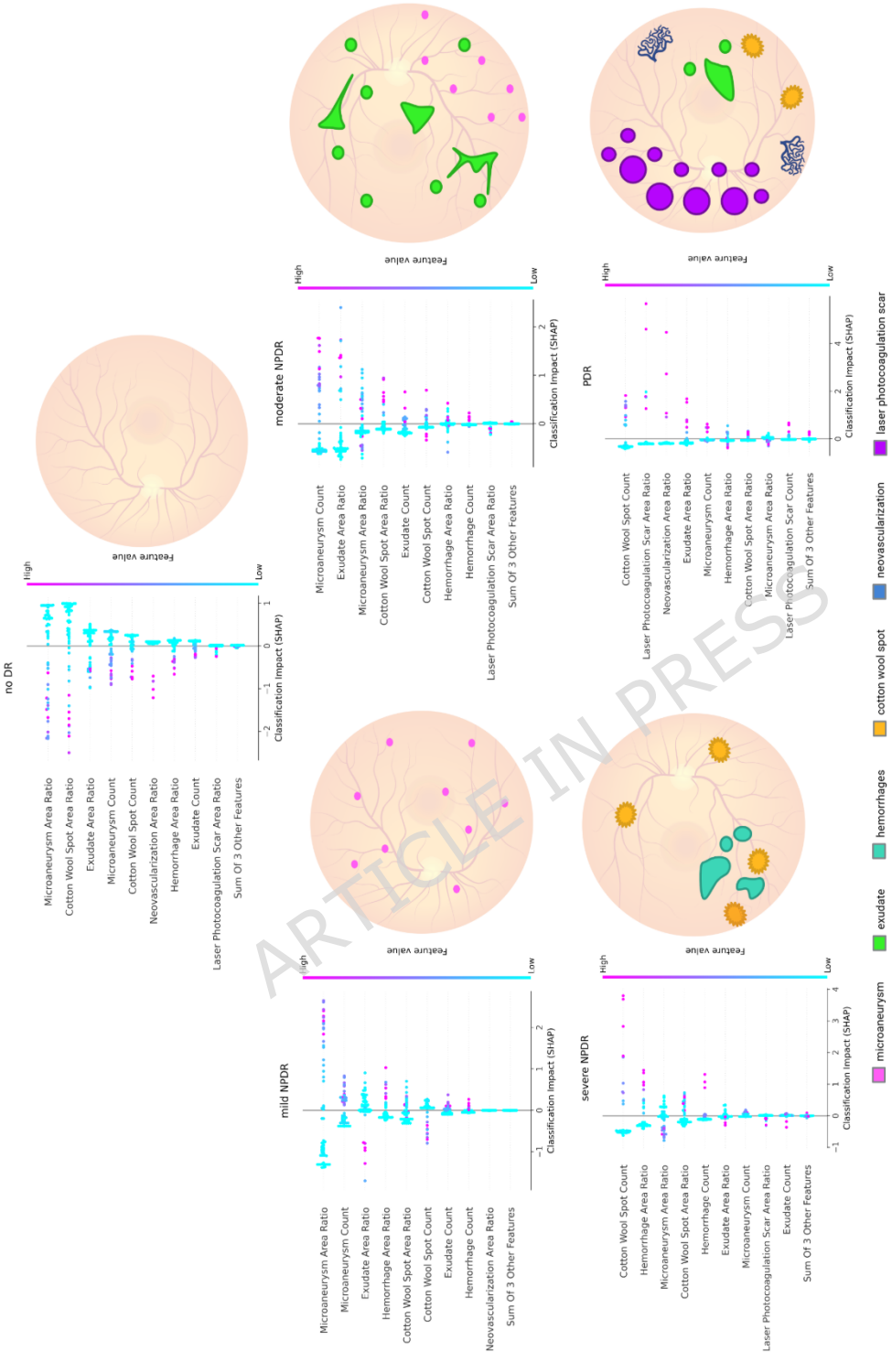


Fig. 5

DISCUSSION

In recent years, the idea of leveraging machine learning to support ophthalmological care has shifted from a distant concept to a real-world paradigm. Several models have received approval from the United States Food and Drug Administration to perform DR screening in clinical settings and have shown practical successes in DR stage prediction for diverse patient populations [64–66]. Further value is expected from comparable approaches to DR stage classification, with potential benefits including fewer false positive and false negative rates [67,68], quicker turnaround times for DR detection and stage identification [69], ability to reach more patients through telehealth [46,70], and reduced burden on ophthalmologists [71,72]. If healthcare-oriented models are to offer such benefits, it is crucial for them to embed a high degree of explainability.

Explainable DL models produce outputs that are traceable (attributable to distinct features), plausible (overlap at least partially with established knowledge), and interpretable (easy for target users to understand) [73,74]. Using expert knowledge to refine model inputs, as well as post- model, attribution-based visualizations provided by tools like SHAP, helps to address the three points above [75].

The BigEye framework integrates several aspects of explainable DL. Domain expertise is applied extensively in segmentation and classification feature engineering techniques. The DeepLabV3+ model is trained to segment six

retinal lesions important in DR pathophysiology. The classification model is trained on lesion count and pixel area features, which are likewise informed by literature indicating that the quantities and surface areas of lesions scale with DR severity [76]. SHAP analysis of the classification model also highlights similarities between important features for each DR stage and the corresponding ICDR stage criteria (**Fig. 5**). High quantities of all lesion types are negatively correlated with No DR, indicating that images within this class correctly have no or few predicted lesions. The mild NPDR class is strongly correlated with high microaneurysm counts while no other lesions contribute positively, reflecting the ICDR criteria that only microaneurysms are present in mild NPDR. High quantities of exudates and microaneurysms are correctly shown to be predictors of moderate NPDR. Higher numbers of hemorrhages and cotton wool spots are strong predictors for severe NPDR. Although cotton wool spots can appear as early as moderate NPDR, they are more often encountered in severe NPDR due to increased hypoxic conditions [77]. Finally, the SHAP analysis indicates that high counts of laser photocoagulation scars and neovascularization lesions are correlated with PDR. Neovascularization lesions distinguish PDR from all other stages according to ICDR criteria and laser photocoagulation scars are evidence of treatment received by individuals with advanced, vision-threatening DR.

A few limitations are present in this study. The BigEye dataset is small (n = 602) and underrepresents later stages of DR like severe NPDR (n = 42) and PDR (n = 46). We also note that although intraretinal microvascular

abnormalities (IRMA) lesions are also used to help determine ICDR stage, there were no such lesions present in the fundus images included in the BigEye dataset. Another drawback is that segmentation performance for microaneurysms was notably lower than other lesion types. Two factors help to explain this issue: (1) microaneurysm lesions are very small; resizing the images to smaller dimensions reduces resolution and further shrinks the microaneurysms and (2) microaneurysms are similar in size and color to dot hemorrhages, a separate lesion type. Finally, while our approach boosts explainability by integrating discrete retinal lesion signals, it is still sensitive to cases where relevant retinal lesions are not detected by the segmentation model or where regions of an image are mistakenly interpreted as signs of DR (e.g. false positives). The confusion matrix provided in **Fig. 4** helps to illustrate this point. Of all control / no DR cases, 5% are incorrectly classified as mild NPDR due to some image regions being mistaken as microaneurysms. It is also observed that fractions of all true DR cases (12% of mild NPDR, 15% of moderate NPDR, 7% of severe NPDR, 7% of PDR) are incorrectly classified as no DR cases due to the segmentation model missing relevant retinal lesion signals. Moreover, fuzziness in some feature values leads to misclassifications between “neighboring” DR stages. For example, 13% of PDR cases are misclassified as severe NPDR, likely due to similar values of retinal lesion features shared between the two stages (e.g. cotton wool spots, hemorrhages). This issue is most obvious for moderate NPDR, where 25% of misclassified

cases are split between the two neighboring stages (19% mild NPDR, 6% severe NPDR).

There are a few methods that would likely reduce the odds of misclassifications between neighboring DR stages. Since DR stages are partially described by the number of lesions in each fundus image quadrant (inferior, superior, nasal, temporal), a framework that can identify fundus images as belonging to the left or right eye can break the image down into the correct quadrants and create more fine-grained retinal lesion features (e.g. counts of microaneurysms in the temporal quadrant). Other strategies may also help more broadly with DR stage prediction performance. An approach that directly incorporates low-level transfer learning features, similar to our technique in the DR-Detector framework [60], may enhance performance by pre-training on a larger set of fundus images and capturing a broader set of simple pixel features for each lesion than would otherwise be detected by a segmentation model. Multi-modal techniques are also known to achieve good DR stage prediction accuracy by integrating lab measurements (blood pressure, HbA1C levels), lifestyle data (diabetes duration, smoking status), and sociodemographic information (age, gender) [78].

METHODS

Experiment Design

This retrospective observational study investigated whether retinal lesion measurements could be used alone to accurately predict DR stages according

to ICDR criteria. To approach this aim, a diverse dataset of regular (133° FOV) and widefield montage (200° FOV) images was assembled from samples retrospectively collected from a local hospital (collection approved by IU IRB #14402) and from e-ophtha , a public fundus image dataset [62]. Subject experts assigned DR stage annotations and generated segmentation masks for six retinal lesion types: (1) microaneurysms, (2) hemorrhages, (3) exudates, (4) cotton wool spots, (5) neovascularization and (6) laser photocoagulation scars. The fundus images, segmentation masks, and weight maps generated from segmentation masks were used by a DeepLabV3+ [79] segmentation model to extract segmentations for the six retinal lesion types. Quantity and pixel area features for each lesion type were extracted from predicted segmentation masks. The lesion features were used to evaluate six DR stage classification models ((1) SVM, (2) RF, (3) NN, (4) XGB, (5) LGBM, and (6) TabNet) through nested 10-fold cross-validation. The most performant classification model was assessed through test accuracy scores. Finally, SHAP beeswarm plots were created post-model to visualize feature importances for each DR stage.

IRB Approval

This study was approved by the Institutional Review Board (IRB) of Indiana University (IU IRB #14402) under an “exempt” category and the requirement for informed consent was waived. All analyses were performed on deidentified data in accordance with relevant guidelines and regulations.

Dataset

Indiana University Health

676 de-identified patient records were acquired from the Indiana University Health (IUH) EHR system as part of a retrospective observational study. Patient records were separated into non-diabetes (n = 448) and diabetes (n = 228) categories based on HbA1C measurements. IU Glick Eye Institute ophthalmologists identified cases of DR (n = 143) and assigned DR stages according to corresponding patient record fundus images. One fundus image from either the left or right eye was used per IUH patient. Images from non-DR diabetes and DR patient records were combined to create a dataset (n = 164) spanning all five ICDR categories: (1) no DR (n = 52), (2) mild NPDR (n = 7), (3) moderate NPDR (n = 15), (4) severe NPDR (n = 44), and (5) PDR (n = 46). Dataset images were macula-centered color retina fundus images obtained from the left and/or right eyes at either 133° FOV (regular) or 200° FOV (widefield montage). Images were taken with a Zeiss Clarus 500/700 retinal camera.

A trained medical student used the VGG Image Annotation (VIA) tool [63] to annotate fundus images for six retinal lesions: (1) microaneurysms, (2) exudates, (3) hemorrhages, (4) cotton wool spots, (5) neovascularization, and (6) laser photocoagulation scars. The annotation guidelines were to trace lesions as close as possible to the actual boundaries. However, less restrictive annotations were approved if the lesions were (1) very small and difficult to annotate at even the highest VIA zoom level (e.g. microaneurysms), or (2)

blurry or out-of-focus. All VIA annotations were reviewed and approved by a board-certified IU Glick Eye Insititute ophthalmologist. Lesion annotation coordinates were combined to create one multiclass segmentation mask for each fundus image. Each retinal lesion in a segmentation mask was assigned to a unique pixel value. Blank, background-only segmentation masks were generated for control (non-DR) images. A schematic overview of VIA-based lesion annotation is shown in **Supplementary Fig. 5**.

e-ophtha

e-ophtha is a public dataset of retina fundus images collected from the Ophthalmology Diabetes Telemedicine (OPHDIAT®) DR screening network [62]. The network covers 16 screening centers based across France. One or both fundus images for each patient are used, as reported in the original e-ophtha dataset. The dataset included non-DR (n = 264) and DR (n = 178) images. All e-ophtha images were macula-centered color retina fundus images taken at 45° FOV using non-mydriatic fundus cameras (CR-DGI, Canon, Tokyo, Japan or TRC-NW6, Topcon, Rotterdam, Netherlands) [62].

The e-ophtha dataset categorizes fundus images as non-DR or DR. IU Glick ophthalmologists applied ICDR criteria to further annotate fundus images according to DR stage. The revised e- ophtha data (n = 438) represented three classes: (1) non-DR (n = 264), mild NPDR (n = 127), and (3) moderate NPDR (n = 47). e-ophtha provides binary segmentation masks for some lesions (microaneurysms and exudates). A trained medical student used VIA to

annotate DR fundus images for the previously unaccounted- for lesions (see **Supplementary Fig. 5**); the annotations were later merged with existing microaneurysm and exudate segmentation masks to create one multiclass segmentation mask for each fundus image. All VIA annotations were reviewed and approved by a board-certified IU Glick Eye Institute ophthalmologist. Each retinal lesion in a segmentation mask was assigned to a unique pixel value. Blank, background-only segmentation masks were generated for control (non-DR) images.

Final Dataset

The IU Glick Eye Institute and e-ophtha datasets were merged to produce a final composite dataset covering all five ICDR categories: (1) no DR (n = 316), (2) mild NPDR (n = 134), (3) moderate NPDR (n= 62), severe NPDR (n = 44), and PDR (n = 46).

Metrics

The BigEye segmentation model was evaluated in terms of precision, macro precision, recall, macro recall, F1 score, macro F1 score, accuracy, Receiver Operating Characteristic Area Under Curve (ROC-AUC), and Intersection over Union (IoU) and uses categorical cross-entropy as its loss function. The BigEye classification models used the same metrics listed above (except IoU) and the same loss function (where applicable). A threshold value of 0.5 was applied for all precision, recall, and F1 score metrics. The metric equations are provided below:

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{macro precision} = \frac{\sum_{i=1}^{N} \text{precision}(i)}{N}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{macro recall} = \frac{\sum_{i=1}^{N} \text{recall}(i)}{N}$$

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{macro F1 score} = \frac{\sum_{i=1}^{N} \text{F1 score}(i)}{N}$$

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{ROC - AUC} = \int_{0}^{1} TPR(FPR)\,dFPR$$

$$\text{IoU} = \frac{A \cap A'}{A \cup A'}$$

where:

TP = True Positives

FP = False Positives

FN = False Negatives

TN = True Negatives

N = Number of Classes

A = Area of Ground Truth Pixels

A’ = Area of Predicted Pixels

DeepLabV3+ Segmentation Model

Dataset Split and Preprocessing

The dataset was split through a nested 5-fold cross validation strategy. RGB retina fundus images undergo a Contrast Limited Adaptive Histogram Equalization (CLAHE) [80] operation. CLAHE enhances local contrast within the image and increases the visibility of some retinal lesions while limiting the amplification of noise. CLAHE was performed separately on each channel before merging again (**Supplementary Fig. 6**).

Masks were used to generate weight maps. The weight maps were passed along with the images and masks during training and were incorporated by the loss function to penalize the misclassification of underrepresented pixels more heavily than others (**Supplementary Fig. 7**). This concept is particularly important in retinal lesion segmentation, where background (non-lesion) pixels vastly outnumber lesion pixels (**Supplementary Table 1**). Masks were transformed from categorical representations to one-hot encoded representations**.**

Augmentation

Three augmentation steps were applied after image preprocessing: (1) random horizontal flips, (2) random vertical flips, and (3) random zoom (-20%

- +10%). A seed value was set to apply the same augmentation operations on corresponding images, segmentation masks, and weight maps to prevent them from becoming disjoint.

Architecture

The semantic segmentation model is the encoder-decoder based DeepLabV3+ network (**Supplementary Fig. 8**). The encoder block captures local feature information as it gradually down-samples an input into a low dimension feature map. The encoder is comprised of two major sub-blocks: (1) the backbone model and (2) the atrous spatial pyramid pooling (ASPP) module. The backbone model receives the network input layer and applies atrous convolutions. In atrous convolution, the atrous rate parameter r controls the stride used for sampling the input; increasing the r value in turn increases the kernel's receptive field and captures a larger context than regular convolution layers with an identical number of parameters. The backbone model outputs a shallow feature map that is passed to the decoder and ASPP module separately. The ASPP module applies multiple atrous convolution layers with different rates in parallel and pools them together to extract dense, multi-scale information from the input feature map. The ASPP module output is passed to a (1x1) convolution block before proceeding to the decoder.

The decoder block is used to upsample low-resolution feature maps to the inputs' original height and width. The decoder concatenates backbone feature maps and upsampled ASPP feature maps before applying a series of upsampling and convolution layers to gradually increase the output resolution

to the specified size. The output softmax layer projects the probabilities distributions for all object classes in the segmentation dataset.

The BigEye DeepLabV3+ implementation accepts input tensors with size (n, 512, 512, 3). The encoder backbone model is a ResNet50 network with ImageNet pretrained weights [80]. The ResNet50 output layer connects to the ASPP module, which consists of five layers: (1) (1x1) atrous convolution with rate 1, (2) (3x3) atrous convolution with rate 6, (3) (3x3) atrous convolution with rate 12, (4) (3x3) atrous convolution with rate 18, (5) pooling. All convolution layers use 256 filters with batch normalization. Final encoder features are created by concatenating ASPP outputs before passing them to a (1x1) atrous convolution layer. The outputs from the (1x1) layer undergo (4x) bilinear upsampling. Shallow features from the encoder backbone are passed through another (1x1) atrous convolution layer before being concatenated with the upsampled features. The concatenation output is passed through a (3x3) atrous convolution layer before undergoing (4x) bilinear upsampling. The upsampling output is passed to a softmax layer.

Model Hyperparameters

The DeepLabV3+ model was trained using categorical cross-entropy loss and an Adam optimizer with a constant learning rate of 1e-4. Training was set to 250 epochs with a batch size of 16. A checkpoint callback assessed model performance at the end of each epoch and saved the model weight with the highest test set IoU score.

Classification Models

Feature Engineering

The trained DeepLabV3+ segmentation model was applied to predict retinal lesions for all fundus images in the dataset. Two feature types were extracted from the predictions: lesion counts (quantities) and lesion pixel areas (**Supplementary Fig. 9**). Count features were obtained from predictions by taking the sums of all canny edges for a particular lesion type. Pixel area features were obtained by dividing the pixel area of a particular lesion type by the pixel area of the fundus. The fundus pixel area was obtained by thresholding the image to extract the lighter, retina fundus portion of the image from the darker background. All computations were performed on the unseen outer-fold test set images at the end of fold model training; the values were then averaged across all folds. All features were tabulated and organized into a spreadsheet.

Architectures

Six different classifier model architectures were investigated: (1) support vector machine (SVM), (2) random forest (RF), (3) simple neural network (NN), (4) extreme gradient boosting (XGB), (5) light gradient boosting machine (LGBM), and (6) TabNet.

Hyperparameters

The SVM was linear kernel based. The RF used 100 estimators and Gini impurity criterion. The NN model used a 0.2 dropout rate, employed an

exponential decay learning rate scheduler that began at 1e-3 and ended at 1e-4, and was trained for 500 epochs with a batch size of 32. The XGB model was set to a max depth of 4, reg_alpha to 0.5, reg_lambda colsample_bytree to 0.8, learning rate to 0.1, n_estimators to 100, and early stopping rounds to 5. The LGBM parameters were set as boosting type to gbdt, num_leaves to 16, learning_rate to 0.02, feature fraction to 0.7, bagging fraction to 0.7, bagging freq to 0.5, and early stopping rounds to 5. TabNet decision and attention dims were set to 16, n_steps to 5, n_shared_glus and n_dependent_glus to 2, relaxation factor to 1.3, epsilon to 1e-15, momentum to 0.98, mask_type to softmax, and lambda_sparse to 1e-3.

Model Selection

The LGBM model (accuracy 0.83 ± 0.03) was used for all downstream DR stage predictions.

SHAP Analysis

The LGBM model was used to construct post-model beeswarm plots for all ICDR stages using the Python shap library [81].

3.2.4.8 Hardware

All model training and analyses were done on the Indiana University Big Red 200 High Performance Computing cluster using GPU-accelerated nodes with 64-core, 2.0 GHz EPYC 7713 CPUs, NVIDIA A100 GPUs, and 500 GB memory.

Acknowledgements

The authors thank Indiana University Information Technology Services Research Technologies (UITS RT) for maintaining the Big Red 200 cluster used in this study. We also thank Indiana University Health and the French Research Agency's ANR-TECSAN-TELEOPHTA and OPHDIAT© projects for the public data used in this study. Figures were created in part with BioRender (https://www.biorender.com).

Author Contributions

H.M.G., D.H.S., A.R.H., and S.C.J contributed to study conceptualization and design. M.H. and A.R.H. performed data curation. M.H. performed initial initial fundus image annotation, and A.R.H. reviewed and approved all annotations. H.M.G, D.H.S, and O.B.O. contributed to result organization and visualization. H.M.G and D.H.S. wrote the BigEye codebase. M.H., F.D.B., J.X.L., and A.R.H provided clinical insight. H.M.G., D.H.S., F.D.B., and S.C.J performed manuscript preparation. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

e-ophtha original data: https://www.adcis.net/en/third-party/e-ophtha/.

IU Health data: Some fundus images used in this study contain protected health information from the Indiana University Health system and cannot be made publicly available in accordance with patient privacy regulations. Data access requests may be directed to the Indiana University Data Management

Council (iudata@iu.edu), subject to ethical approval and execution of appropriate data use agreements.

BigEye source code and model weights: https://github.com/Janga-Lab/BigEye/.

Competing Interests Statement

The authors declare no competing interests.

Funding

This study was supported by NIH / NEI grant 5T35EY031282-05, the Indiana University Indianapolis Institute of Integrative Artificial Intelligence (iAI), and the Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. The funding providers had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

FIGURE CAPTIONS

Fig. 1 The Major Signs of Diabetic Retinopathy (DR) in Retina Fundus Images. It is important to note that laser photocoagulation scars occur as a result of specialized eye surgery and do not develop organically. All other retinal lesions emerge at different points in DR progression and can be analyzed systematically to determine DR stage. Created in BioRender. Gill, H. (2026) https://BioRender.com/y13v130

Fig. 2 Summary of BigEye Framework Training and Prediction. a) Color retina fundus images are collected from a private cohort of Indiana University

Health EHR records and the public e-ophtha database. The boundaries of retinal lesions are annotated to create a comprehensive fundus image segmentation dataset. The annotated DR fundus images are balanced by non-DR control images, which are important in reducing false positives during model training. **b)** The dataset is used to train a DeepLabV3+ model to segment six different lesion types: (1) microaneurysms, (2) exudates, (3) hemorrhages, (4) cotton wool spots, (5) neovascularization, (6) laser photocoagulation scars. Measurements from the predicted lesions are matched with corresponding DR stages and used to train a DR stage classification model. **c)** The same general approach for model training is applied to unseen retina fundus images. Here, microaneurysm and exudate features are segmented and combined into a dataframe, which is then used by the classification model to predict the “Mild NPDR” stage. Created in BioRender. Gill, H. (2026) https://BioRender.com/lg2sbrr

Fig. 3 BigEye Dataset Composition. The BigEye framework is trained on images from a cohort of private Indiana University Health records, as well as the public e-ophtha dataset. The combined dataset is roughly a 50:50 split of control to DR cases and features regular and widefield fundus images captured with different camera systems, enhancing image diversity. Created in BioRender. Gill, H. (2026) https://BioRender.com/n96p373

Fig. 4 DR Stage Confusion Matrix. Confusion Matrix of the LGBM model’s performance in classifying DR stage. Created in BioRender. Gill, H. (2026) https://BioRender.com/l70p820

Fig. 5 SHAP Analysis of DR Stages Shapely additive explanations (SHAP) values for the most informative retinal lesion features for each DR stage. Positive SHAP values show which lesion features contribute towards a certain DR stage, while negative values are tied to lesion features that count against prediction of that stage. Absolute SHAP values indicate the strength of the impact towards or against a DR stage, with values at or near 0 having less impact. Created in BioRender. Gill, H. (2026) https://BioRender.com/y86kv4a

TABLE CAPTIONS

Table 1: International Classification of Diabetic Retinopathy (ICDR) Severity Scale

Table 2: Performance Metrics for BigEye Lesion Segmentation Model

Table 3: Performance Metrics for BigEye DR Stage Classification Models

REFERENCES

1. Chong, D. D., Das, N. & Singh, R. P. Diabetic retinopathy: Screening, prevention, and treatment. *Cleve. Clin. J. Med.* **91**, 503–510 (2024).

2. Kaur, A., Kumar, R. & Sharma, A. Diabetic Retinopathy Leading to Blindness- A Review. *Curr. Diabetes Rev.* **20**, e240124225997 (2024).

3. Henricsson, M. *et al.* The incidence of retinopathy 10 years after diagnosis in young adult people with diabetes: results from the nationwide population-based Diabetes Incidence Study in Sweden (DISS). *Diabetes Care* **26**, 349–354 (2003).

4. Berrabeh, S. *et al.* Prevalence and Risk Factors of Retinopathy in Type 1 Diabetes: A Cross-Sectional Study. *Cureus* **15**, e47993 (2023).

5. Romero-Aroca, P. *et al.* Ten-year incidence of diabetic retinopathy and macular edema. Risk factors in a sample of people with type 1 diabetes. *Diabetes Res. Clin. Pract.* **94**, 126–132 (2011).

6. Romero-Aroca, P. *et al.* Differences in incidence of diabetic retinopathy between type 1 and 2 diabetes mellitus: a nine-year follow-up study. *Br. J. Ophthalmol.* **101**, 1346–1351 (2017).

7. Voigt, M. *et al.* Prevalence and Progression Rate of Diabetic Retinopathy in Type 2 Diabetes Patients in Correlation with the Duration of Diabetes. *Exp. Clin. Endocrinol. Diabetes* **126**, 570–576 (2018).

8. Yin, L., Zhang, D., Ren, Q., Su, X. & Sun, Z. Prevalence and risk factors of diabetic retinopathy in diabetic patients: A community based cross-sectional study. *Medicine (Baltimore)* **99**, e19236 (2020).

9. Aljehani, E. A., Alhawiti, A. E. & Mohamad, R. M. Prevalence and determinants of diabetic retinopathy among type 2 diabetic patients in saudi arabia: A systematic review. *Cureus* **15**, e42771 (2023).

10. Song, P., Yu, J., Chan, K. Y., Theodoratou, E. & Rudan, I. Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis. *J. Glob. Health* **8**, 010803 (2018).

11. Lundeen, E. A. *et al.* Prevalence of diabetic retinopathy in the US in 2021. *JAMA Ophthalmol.* **141**, 747–754 (2023).

12. Perais, J. *et al.* Prognostic factors for the development and progression of proliferative diabetic retinopathy in people with diabetic retinopathy. *Cochrane Database Syst. Rev.* **2**, CD013775 (2023).

13. Grossniklaus, H. E., Geisert, E. E. & Nickerson, J. M. Introduction to the retina. *Prog. Mol. Biol. Transl. Sci.* **134**, 383–396 (2015).

14. Singh, C. Metabolism and vascular retinopathies: current perspectives and future directions. *Diagnostics (Basel)* **12**, (2022).

15. Country, M. W. Retinal metabolism: A comparative look at energetics in the retina. *Brain Res.* **1672**, 50–57 (2017).

16. Hudson, N. & Campbell, M. Tight junctions of the neurovascular unit. *Front. Mol. Neurosci.* **14**, 752781 (2021).

17. Rudraraju, M., Narayanan, S. P. & Somanath, P. R. Regulation of blood-retinal barrier cell-junctions in diabetic retinopathy. *Pharmacol. Res.* **161**, 105115 (2020).

18. Eshaq, R. S., Aldalati, A. M. Z., Alexander, J. S. & Harris, N. R. Diabetic retinopathy: Breaking the barrier. *Pathophysiology* **24**, 229–241 (2017).

19. Park, D. Y. *et al.* Plastic roles of pericytes in the blood-retinal barrier. *Nat. Commun.* **8**, 15296 (2017).

20. Willis, J. R. *et al.* Vision-Related Functional Burden of Diabetic Retinopathy Across Severity Levels in the United States. *JAMA Ophthalmol.* **135**, 926–932 (2017).

21. Curtis, T. M., Gardiner, T. A. & Stitt, A. W. Microvascular lesions of diabetic retinopathy: clues towards understanding pathogenesis? *Eye* **23**, 1496–1508 (2009).

22. Horie, S. & Ohno-Matsui, K. Progress of Imaging in Diabetic Retinopathy-From the Past to the Present. *Diagnostics (Basel)* **12**, (2022).

23. Takamura, Y., Yamada, Y. & Inatani, M. Role of microaneurysms in the pathogenesis and therapy of diabetic macular edema: A descriptive review. *Medicina (Kaunas)* **59**, (2023).

24. An, D., Tan, B., Yu, D.-Y. & Balaratnasingam, C. Differentiating microaneurysm pathophysiology in diabetic retinopathy through objective analysis of capillary nonperfusion, inflammation, and pericytes. *Diabetes* **71**, 733–746 (2022).

25. Santos, A. R. *et al.* Microaneurysm Turnover in Mild Non-Proliferative Diabetic Retinopathy is Associated with Progression and Development of Vision-Threatening Complications: A 5-Year Longitudinal Study. *J. Clin. Med.* **10**, (2021).

26. Shukla, U. V., Gurnani, B. & Kaufman, E. J. Intraocular Hemorrhage. in *StatPearls* (StatPearls Publishing, 2025).

27. Kanukollu, V. M. & Ahmad, S. S. Retinal Hemorrhage. in *StatPearls* (StatPearls Publishing, 2025).

28. Chew, E. Y. *et al.* Association of elevated serum lipid levels with retinal hard exudate in diabetic retinopathy. Early Treatment Diabetic Retinopathy Study (ETDRS) Report 22. *Arch. Ophthalmol.* **114**, 1079–1084 (1996).

29. Raman, R., Nittala, M. G., Gella, L., Pal, S. S. & Sharma, T. Retinal Sensitivity over Hard Exudates in Diabetic Retinopathy. *J. Ophthalmic Vis. Res.* **10**, 160–164 (2015).

30. Chui, T. Y. P., Thibos, L. N., Bradley, A. & Burns, S. A. The mechanisms of vision loss associated with a cotton wool spot. *Vision Res.* **49**, 2826–2834 (2009).

31. Kim, J. S. *et al.* The microperimetry of resolved cotton-wool spots in eyes of patients with hypertension and diabetes mellitus. *Arch. Ophthalmol.* **129**, 879–884 (2011).

32. Chaudhary, S., Zaveri, J. & Becker, N. Proliferative diabetic retinopathy (PDR). *Dis. Mon.* **67**, 101140 (2021).

33. Lange, J., Hadziahmetovic, M., Zhang, J. & Li, W. Region-specific ischemia, neovascularization and macular oedema in treatment-naïve proliferative diabetic retinopathy. *Clin. Experiment. Ophthalmol.* **46**, 757–766 (2018).

34. Agrawal, M. *et al.* Pattern and distribution of neovascularization in proliferative diabetic retinopathy on fundus fluorescein angiography: A growing paradigm. *Med. J. Armed Forces India* **79**, 207–212 (2023).

35. Iyer, S. S. R., Regan, K. A., Burnham, J. M. & Chen, C. J. Surgical management of diabetic tractional retinal detachments. *Surv. Ophthalmol.* **64**, 780–809 (2019).

36. Stewart, M. W., Browning, D. J. & Landers, M. B. Current management of diabetic tractional retinal detachments. *Indian J. Ophthalmol.* **66**, 1751–1762 (2018).

37. Muqit, M. M. K. *et al.* Pascal panretinal laser ablation and regression analysis in proliferative diabetic retinopathy: Manchester Pascal Study Report 4. *Eye* **25**, 1447–1456 (2011).

38. Pande, G. S. & Tidake, P. Laser treatment modalities for diabetic retinopathy. *Cureus* **14**, e30024 (2022).

39. Stefánsson, E. The therapeutic effects of retinal laser treatment and vitrectomy. A theory based on oxygen and vascular physiology. *Acta Ophthalmol. Scand.* **79**, 435–440 (2001).

40. Wang, X. *et al.* Panretinal photocoagulation plus intravitreal conbercept for diabetic retinopathy in real world: a retrospective study. *BMC Ophthalmol.* **23**, 400 (2023).

41. Zhan, H.-Q., Zhou, J.-L., Zhang, J., Wu, D. & Gu, C.-Y. Conbercept combined with laser photocoagulation in the treatment of diabetic macular edema and its influence on intraocular cytokines. *World J. Diabetes* **14**, 1271–1279 (2023).

42. Lin, Z., Deng, A., Hou, N., Gao, L. & Zhi, X. Advances in targeted retinal photocoagulation in the treatment of diabetic retinopathy. *Front Endocrinol (Lausanne)* **14**, 1108394 (2023).

43. Gawęcki, M., Kiciński, K., Bianco, L. & Battaglia Parodi, M. Regression of Neovascularization after Panretinal Photocoagulation Combined with Anti-VEGF Injection for Proliferative Diabetic Retinopathy-A Review. *Diagnostics (Basel)* **14**, (2023).

44. Cleland, C. Comparing the International Clinical Diabetic Retinopathy (ICDR) severity scale. *Community Eye Health* **36**, 10 (2023).

45. Romero-Aroca, P. *et al.* A clinical decision support system for diabetic retinopathy screening: creating a clinical support application. *Telemed. J. E Health* **25**, 31–40 (2019).

46. Nakayama, L. F. *et al.* Artificial intelligence for telemedicine diabetic retinopathy screening: a review. *Ann. Med.* **55**, 2258149 (2023).

47. Mohanty, C. *et al.* Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors* **23**, (2023).

48. Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z. & Jabbar, A. Transfer Learning-Based Model for Diabetic Retinopathy Diagnosis Using Retinal Images. *Brain Sci.* **12**, (2022).

49. Narayanan, B. N., Hardie, R. C., De Silva, M. S. & Kueterman, N. K. Hybrid machine learning architecture for automated detection and grading of retinal images for diabetic retinopathy. *J Med Imaging (Bellingham)* **7**, 034501 (2020).

50. Chilukoti, S. V., Shan, L., Tida, V. S., Maida, A. S. & Hei, X. A reliable diabetic retinopathy grading via transfer learning and ensemble learning with quadratic weighted kappa metric. *BMC Med. Inform. Decis. Mak.* **24**, 37 (2024).

51. Larsen, T. J. *et al.* The use of artificial intelligence to assess diabetic eye disease among the Greenlandic population. *Int. J. Circumpolar Health* **83**, 2314802 (2024).

52. Gu, Z. *et al.* Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention. *Comput. Intell. Neurosci.* **2023**, 1305583 (2023).

53. Yao, Z. *et al.* FunSwin: A deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images. *Front. Physiol.* **13**, 961386 (2022).

54. Manan, M. A. *et al.* Semantic segmentation of retinal exudates using a residual encoder-decoder architecture in diabetic retinopathy. *Microsc. Res. Tech.* **86**, 1443–1460 (2023).

55. Jiang, Y. *et al.* Segmentation of Laser Marks of Diabetic Retinopathy in the Fundus Photographs Using Lightweight U-Net. *J. Diabetes Res.* **2021**, 8766517 (2021).

56. Gao, W., Fan, B., Fang, Y. & Song, N. Lightweight and multi-lesion segmentation model for diabetic retinopathy based on the fusion of mixed attention and ghost feature mapping. *Comput. Biol. Med.* **169**, 107854 (2024).

57. Xu, Y. *et al.* FFU-Net: Feature Fusion U-Net for Lesion Segmentation of Diabetic Retinopathy. *Biomed Res. Int.* **2021**, 6644071 (2021).

58. Alam, M., Zhao, E. J., Lam, C. K. & Rubin, D. L. Segmentation-Assisted Fully Convolutional Neural Network Enhances Deep Learning Performance to Identify Proliferative Diabetic Retinopathy. *J. Clin. Med.* **12**, (2023).

59. Andersen, J. K. H., Hubel, M. S., Rasmussen, M. L., Grauslund, J. & Savarimuthu, T. R. Automatic Detection of Abnormalities and Grading of Diabetic Retinopathy in 6-Field Retinal Images: Integration of Segmentation Into Classification. *Transl. Vis. Sci. Technol.* **11**, 19 (2022).

60. Hassan, D. *et al.* Combining transfer learning with retinal lesion features for accurate detection of diabetic retinopathy. *Front Med (Lausanne)* **9**, 1050436 (2022).

61. Dai, L. *et al.* A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **12**, 3242 (2021).

62. Decencière, E. *et al.* TeleOphta: Machine learning and image processing methods for teleophthalmology. *IRBM* **34**, 196–203 (2013).

63. Dutta, A. & Zisserman, A. The VIA annotation software for images, audio and video. in *Proceedings of the 27th ACM International Conference on Multimedia - MM '19* 2276–2279 (ACM Press, 2019). doi:10.1145/3343031.3350535.

64. Riotto, E. *et al.* Accuracy of Autonomous Artificial Intelligence-Based Diabetic Retinopathy Screening in Real-Life Clinical Practice. *J. Clin. Med.* **13**, (2024).

65. Rajalakshmi, R., Subashini, R., Anjana, R. M. & Mohan, V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye* **32**, 1138–1144 (2018).

66. Karabeg, M. *et al.* A pilot cost-analysis study comparing AI-based EyeArt® and ophthalmologist assessment of diabetic retinopathy in minority women in Oslo, Norway. *Int. J. Retina Vitreous* **10**, 40 (2024).

67. Raumviboonsuk, P. *et al.* Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *npj Digital Med.* **2**, 25 (2019).

68. Lee, A. Y. *et al.* Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems. *Diabetes Care* **44**, 1168–1175 (2021).

69. Kalavar, M. *et al.* Applications of artificial intelligence for the detection, management, and treatment of diabetic retinopathy. *Int. Ophthalmol. Clin.* **60**, 127–145 (2020).

70. Kong, M. & Song, S. J. Artificial intelligence applications in diabetic retinopathy: what we have now and what to expect in the future. *Endocrinol Metab (Seoul)* **39**, 416–424 (2024).

71. Bora, A. *et al.* Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit. Health* **3**, e10–e19 (2021).

72. Vujosevic, S., Limoli, C., Luzi, L. & Nucci, P. Digital innovations for retinal care in diabetic retinopathy. *Acta Diabetol.* **59**, 1521–1530 (2022).

73. Bienefeld, N. *et al.* Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *npj Digital Med.* **6**, 94 (2023).

74. Amann, J. *et al.* Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **20**, 310 (2020).

75. Singh, A., Sengupta, S. & Lakshminarayanan, V. Explainable deep learning models in medical image analysis. *J. Imaging* **6**, (2020).

76. Esmaeilkhanian, H. *et al.* The relationship of diabetic retinopathy severity scales with frequency and surface area of diabetic retinopathy lesions. *Graefes Arch Clin Exp Ophthalmol* **261**, 3165–3176 (2023).

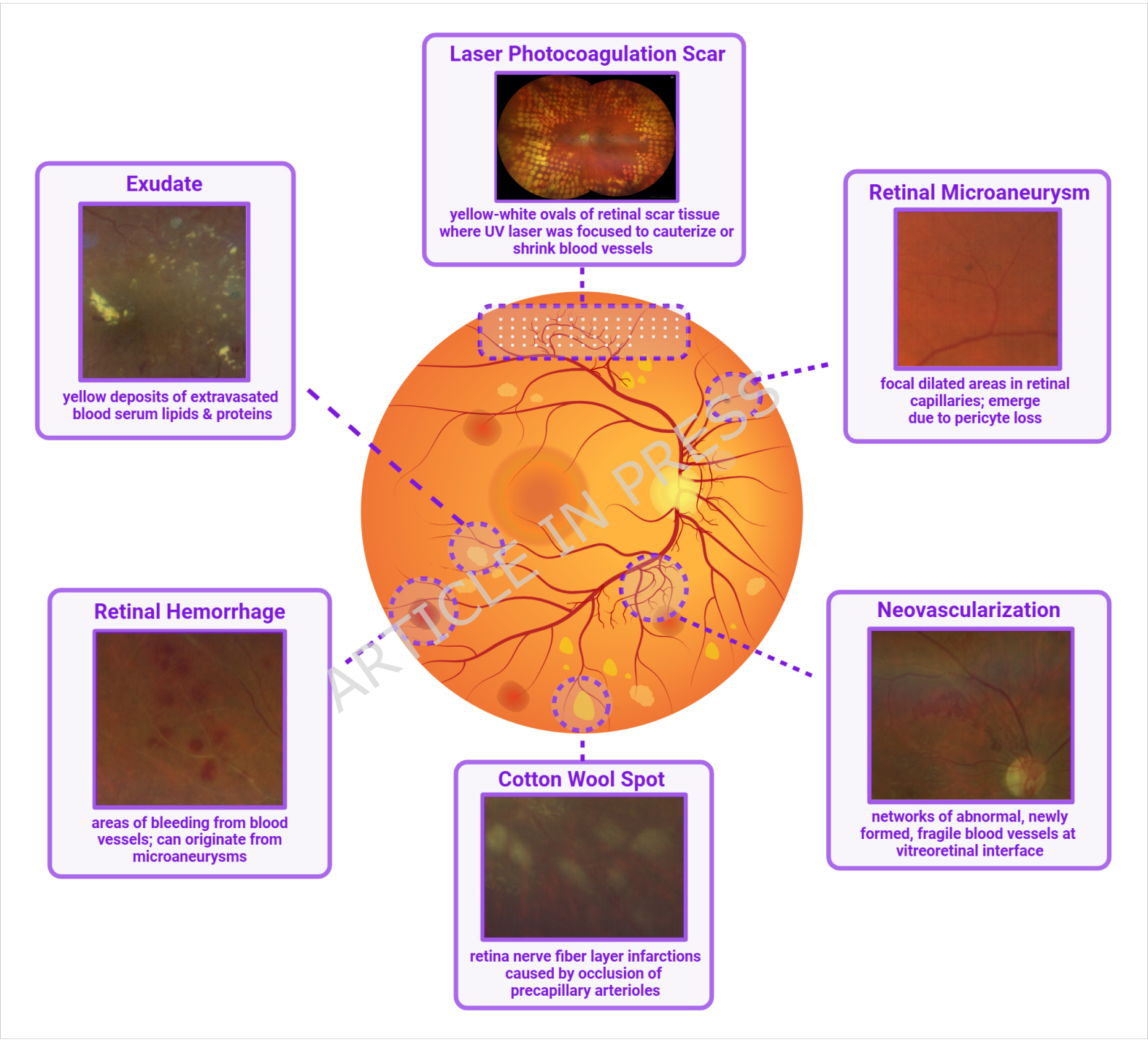
77. Rahimi, M. *et al.* Inner retinal oxygen delivery and metabolism in progressive stages of diabetic retinopathy. *Sci. Rep.* **14**, 4414 (2024).

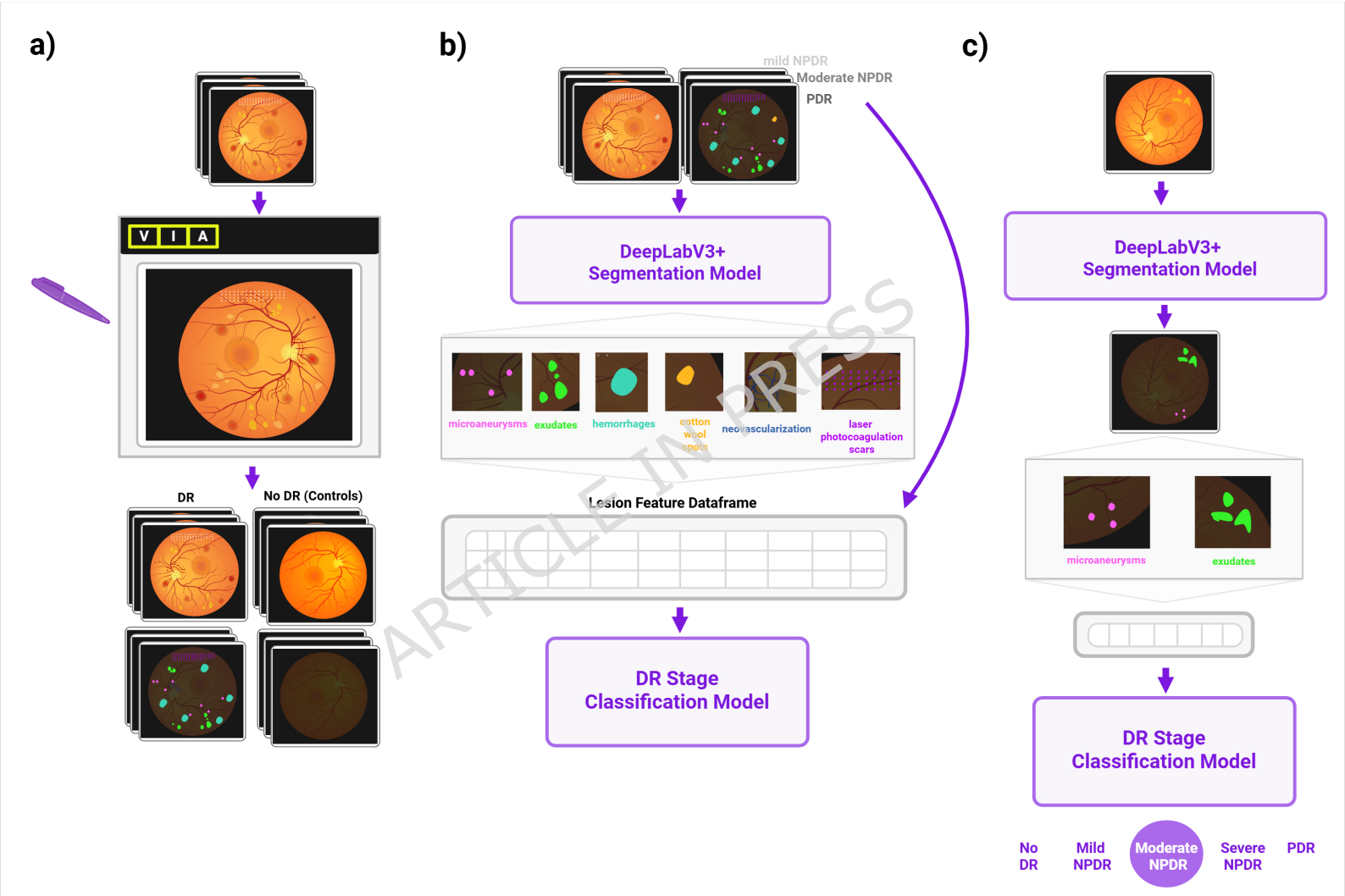
78. Dai, L. *et al.* A deep learning system for predicting time to progression of diabetic retinopathy. *Nat. Med.* **30**, 584–594 (2024).

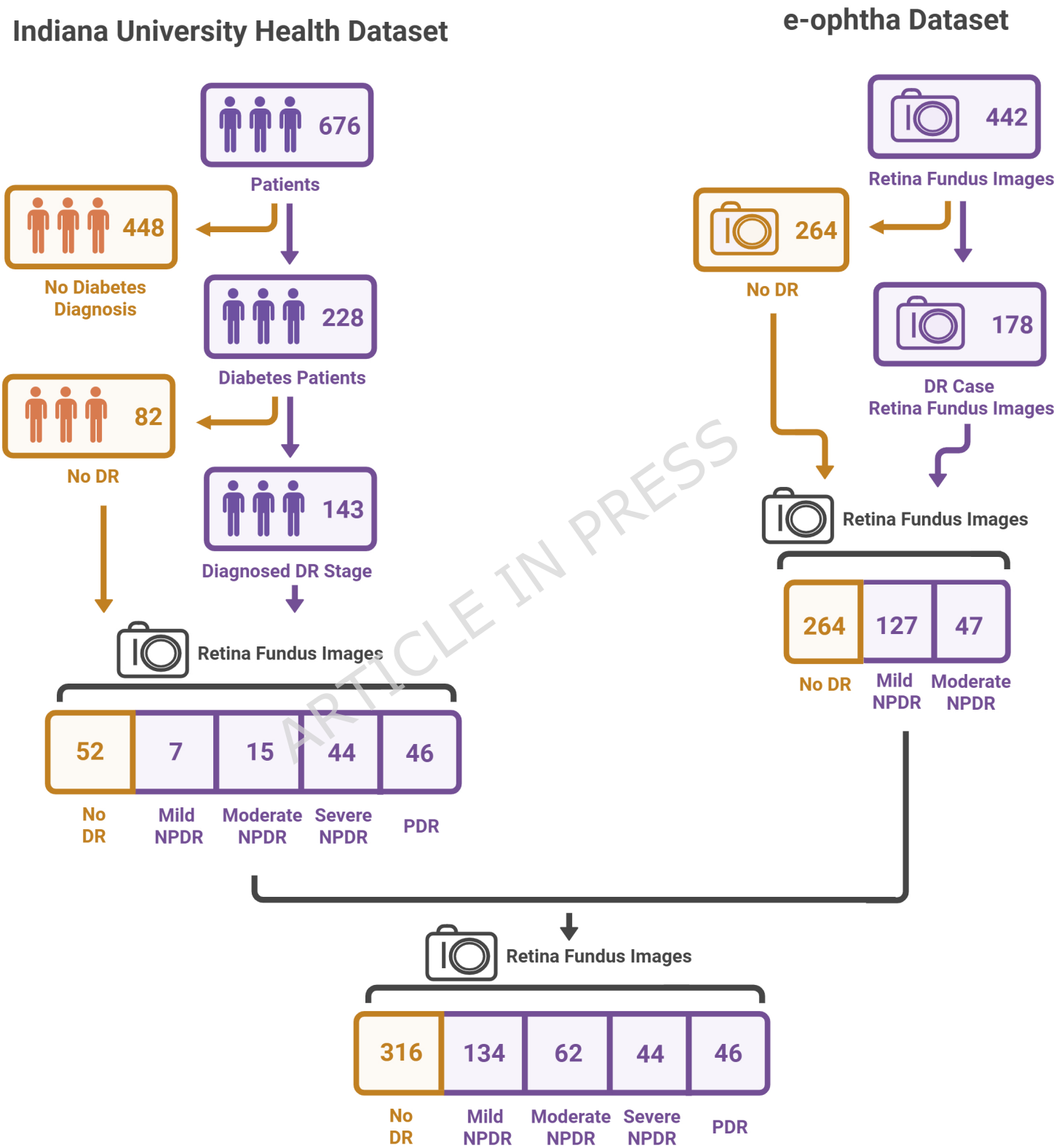
79. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *arXiv* (2018) doi:10.48550/arxiv.1802.02611.

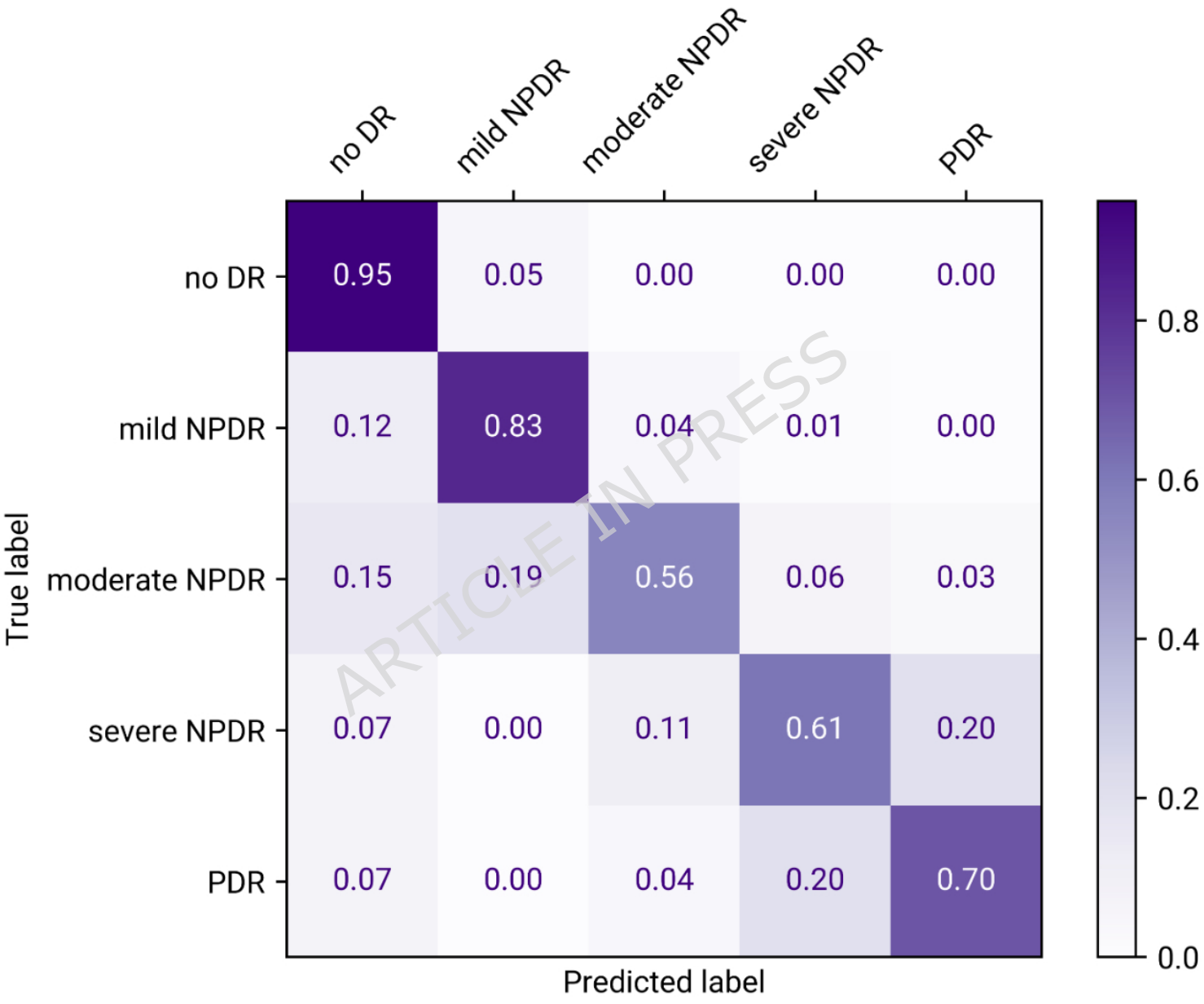
80. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. https://www.cse.unr.edu/~bebis/CS474/StudentPaperPresentations/1994%20-%20CLAHE.pdf.

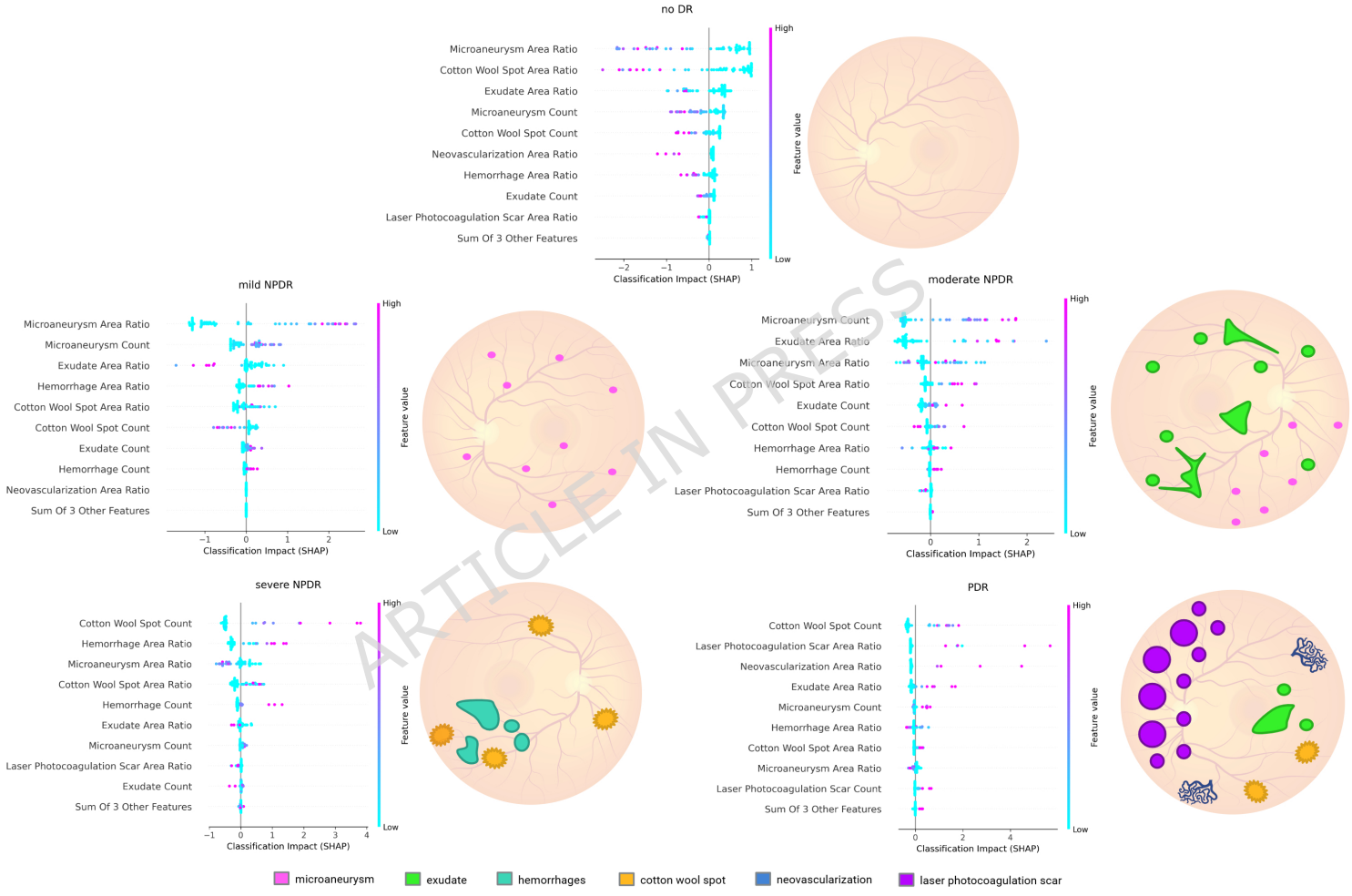
81. Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* (2017) doi:10.48550/arxiv.1705.07874.











ICDR Stage
No DR
mild NPDR
moderate NPDR
severe NPDR
PDR

Criteria
no apparent retinal lesions
microaneurysms in any fundus region
icroaneruysms AND exudates, hemorrhages, and / or cotton wool spots in any fundus regio
all four fundus quadrants, (2) venuous beading in two or more quadrants, (3) intraretinal mi
lesions AND vitreous/preretinal hemorrhage, neovascularization, and/or panretinal laser phc

ɔn
icrovascular abnormality (IRMA) in one or more qua
ɔtocoagulation scars

	Subset		
Metric	**Training**	**Validation**	**Test**
Precision	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
Precision (Macro)	0.89 ± 0.01	0.85 ± 0.08	0.71 ± 0.08
Recall	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
Recall (Macro)	0.87 ± 0.01	0.72 ± 0.07	0.55± 0.11
F1 Score	0.99±0.00	0.98 ± 0.01	0.98 ± 0.01
F1 Score (macro)	0.89 ± 0.01	0.78 ± 0.01	0.62 ± 0.09
Accuracy	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01
ROC-AUC	0.69 ± 0.01	0.71 ± 0.01	0.69 ± 0.01
IoU (total)	0.81 ± 0.01	0.68 ± 0.06	0.56 ± 0.11
Loss	0.03 ± 0.00	0.08 ± 0.04	0.10 ± 0.09

	Train		
Classifier	**Precision**	**Recall**	**F1 Score**
SVM	0.82 ± 0.01	0.76 ± 0.02	0.78 ± 0.02
RF	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
NN	0.82 ± 0.09	0.57 ± 0.19	0.66 ± 0.19
XGB	0.95 ± 0.01	0.92 ± 0.01	0.94 ± 0.01
LGBM	0.86 ± 0.02	0.82 ± 0.02	0.84 ± 0.02
TabNet	0.85 ± 0.02	0.75 ± 0.03	0.79 ± 0.02

	Valida		
Classifier	**Precision**	**Recall**	**F1 Score**
SVM	0.77 ± 0.10	0.70 ± 0.08	0.71 ± 0.09
RF	0.78 ± 0.07	0.74 ± 0.06	0.75 ± 0.07
NN	0.87 ± 0.06	0.62 ± 0.21	0.70 ± 0.20
XGB	0.81 ± 0.08	0.78 ± 0.07	0.77 ± 0.07
LGBM	0.78 ± 0.08	0.72 ± 0.07	0.73 ± 0.08
TabNet	0.88 ± 0.05	0.72 ± 0.06	0.79 ± 0.05

	Tes		
Classifier	**Precision**	**Recall**	**F1 Score**
SVM	0.76 ± 0.10	0.70 ± 0.07	0.71 ± 0.08
RF	0.79 ± 0.07	0.76 ± 0.07	0.76 ± 0.07
NN	0.86 ± 0.04	0.74 ± 0.05	0.79 ± 0.04
XGB	0.79 ± 0.07	0.76 ± 0.06	0.77 ± 0.07
LGBM	0.77 ± 0.07	0.71 ± 0.06	0.72 ± 0.07
TabNet	0.86 ± 0.04	0.75 ± 0.04	0.80 ± 0.04

ing		
Accuracy	**ROC-AUC**	**Loss**
0.83 ± 0.00	0.93 ± 0.00	NA
0.99 ± 0.00	0.99 ± 0.00	NA
0.75 ± 0.04	0.93 ± 0.04	0.78 ± 0.25
0.94 ± 0.00	0.99 ± 0.00	0.25 ± 0.03
0.88 ± 0.00	0.97 ± 0.00	0.39 ± 0.02
0.80 ± 0.02	0.96 ± 0.01	0.57 ± 0.05

ation		
Accuracy	**ROC-AUC**	**Loss**
0.81 ± 0.05	0.91 ± 0.03	NA
0.83 ± 0.04	0.95 ± 0.02	NA
0.82 ± 0.04	0.95 ± 0.03	0.67 ± 0.22
0.85 ± 0.04	0.96 ± 0.02	0.47 ± 0.10
0.83 ± 0.04	0.95 ± 0.02	0.53 ± 0.09
0.83 ± 0.04	0.96 ± 0.02	0.64 ± 0.10

st		
Accuracy	**ROC-AUC**	**Loss**
0.80 ± 0.04	0.91 ± 0.04	NA
0.83 ± 0.05	0.95 ± 0.02	NA
0.80 ± 0.04	0.96 ± 0.01	0.55 ± 0.10
0.84 ± 0.04	0.96 ± 0.01	0.49 ± 0.08
0.83 ± 0.03	0.95 ± 0.02	0.54 ± 0.07
0.80 ± 0.04	0.95 ± 0.02	0.63 ± 0.10