Check for updates

Combining transfer learning with retinal lesion features for accurate detection of diabetic retinopathy

Doaa Hassan[1,2†], Hunter Mathias Gill[1†], Michael Happe[3], Ashay D. Bhatwadekar[3], Amir R. Hajrasouliha[3] and Sarath Chandra Janga[1,4,5]*

[1]Department of BioHealth Informatics, School of Informatics and Computing, Indiana University Purdue University, Indianapolis, IN, United States, [2]Computers and Systems Department, National Telecommunication Institute, Cairo, Egypt, [3]Department of Ophthalmology, Glick Eye Institute, Indiana University School of Medicine, Indianapolis, IN, United States, [4]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Medical Research and Library Building, Indianapolis, IN, United States, [5]Centre for Computational Biology and Bioinformatics, Indiana University School of Medicine, 5021 Health Information and Translational Sciences (HITS), Indianapolis, IN, United States

Diabetic retinopathy (DR) is a late microvascular complication of Diabetes Mellitus (DM) that could lead to permanent blindness in patients, without early detection. Although adequate management of DM *via* regular eye examination can preserve vision in in 98% of the DR cases, DR screening and diagnoses based on clinical lesion features devised by expert clinicians; are costly, time-consuming and not sufficiently accurate. This raises the requirements for Artificial Intelligent (AI) systems which can accurately detect DR automatically and thus preventing DR before affecting vision. Hence, such systems can help clinician experts in certain cases and aid ophthalmologists in rapid diagnoses. To address such requirements, several approaches have been proposed in the literature that use Machine Learning (ML) and Deep Learning (DL) techniques to develop such systems. However, these approaches ignore the highly valuable clinical lesion features that could contribute significantly to the accurate detection of DR. Therefore, in this study we introduce a framework called DR-detector that employs the Extreme Gradient Boosting (XGBoost) ML model trained *via* the combination of the features extracted by the pretrained convolutional neural networks commonly known as transfer learning (TL) models and the clinical retinal lesion features for accurate detection of DR. The retinal lesion features are extracted *via* image segmentation technique using the UNET DL model and captures exudates (EXs), microaneurysms (MAs), and hemorrhages (HEMs) that are relevant lesions for DR detection. The feature combination approach implemented in DR-detector has been applied to two common TL models in the literature namely VGG-16 and ResNet-50. We trained the DR-detector model using a training dataset comprising of 1,840 color fundus images collected from e-ophtha, retinal lesions and APTOS 2019 Kaggle datasets of which 920 images are healthy. To validate the DR-detector model, we test the model on external dataset that consists of 81 healthy images collected from High-Resolution Fundus (HRF) dataset and MESSIDOR-2 datasets and 81 images with DR signs collected from Indian

Diabetic Retinopathy Image Dataset (IDRID) dataset annotated for DR by expert. The experimental results show that the DR-detector model achieves a testing accuracy of 100% in detecting DR after training it with the combination of ResNet-50 and lesion features and 99.38% accuracy after training it with the combination of VGG-16 and lesion features. More importantly, the results also show a higher contribution of specific lesion features toward the performance of the DR-detector model. For instance, using only the hemorrhages feature to train the model, our model achieves an accuracy of 99.38 in detecting DR, which is higher than the accuracy when training the model with the combination of all lesion features (89%) and equal to the accuracy when training the model with the combination of all lesions and VGG-16 features together. This highlights the possibility of using only the clinical features, such as lesions that are clinically interpretable, to build the next generation of robust artificial intelligence (AI) systems with great clinical interpretability for DR detection. The code of the DR-detector framework is available on GitHub at https://github.com/Janga-Lab/DR-detector and can be readily employed for detecting DR from retinal image datasets.

Introduction

Diabetic Retinopathy (DR) is a microvascular disorder associated with long-term diabetes mellitus and is one of the leading causes of preventable vision loss across the worldwide (1). DR manifests in individuals diagnosed with Type 1 Diabetes (T1D) or Type 2 Diabetes (T2D). Roughly one-third of diabetic patients are affected by DR (2, 3), and the likelihood of developing DR scales with the length of diabetes duration (4).

The progression of DR in T1D and T2D is characterized by damage to the retina. The retina is a multilayered network of rod and cone photoreceptor cells integrated with bipolar and ganglion cells that enable vision by encoding information gained from light as nerve impulses (5). The retina is supplied with oxygen and nutrients by an extensive vascular system. In T1D and T2D, high blood glucose levels contribute to pro-inflammatory changes that increase the permeability of the blood-retina barrier, leading to leakage of fluids and blood into the retina (6). High blood glucose can also block small retinal capillaries, impeding the delivery of nutrients and contributing to further damage (7).

Although adequate management of DM *via* regular eye examination can preserve vision in DR in many cases, DR screening and diagnoses currently involve highly trained and qualified medical professionals at a high cost. Thus, there is a continuous need for the development of automatic approaches for DR detection as a cheaper alternative to the time-consuming manual DR diagnosis by trained clinicians. A promising application of these approaches is Computer Assisted Diagnosis (CAD) support for detection of DR. An advantage of such CAD applications is that they offset the burden on medical professionals like expert ophthalmologists and fill their absence in addition to preventing DR before affecting vision. This consideration is critical, considering that the global burden of DR is expected to expand to 700 million cases by the 2040s (8). Many DR detection techniques suitable for CAD utilize Machine Learning (ML), Deep Learning (DL) algorithms and various previously pretrained DL models commonly known as Transfer Learning (TL) models.

The TL models have been successfully used for automated binary and multi-class classification of color fundus retinal images for DR detection (9–12). These algorithms have shown a great performance in the automatic detection of DR in non-clinical setups when the dataset is very small and might cause chances for underfitting or high generalization error. In such cases, TL is preferred over standard DL techniques. Recently, the focus has been shifted to TL feature-based models, where common TL algorithms are used for extracting many important local (textural) features from retinal images for detection of DR and predicting its severity level through convolving with a sliding window and forming a filter. For example, features extracted from AlexNet TL model (13) were passed to the Support Vector Machine (SVM) ML model to enhance the efficiency of the DR classification system, where SVM model achieved accuracies of 97.93 and 95.26% in five-class DR classification with linear discriminant analysis (LDA) feature selection and principle component analysis (PCA) dimensional reduction, respectively (14), when training the model and testing

on Kaggle dataset[1]. As an extension to the same direction, features from the final layers of VGG-19 TL model (15) were collected and aggregated to get a deeper representation of retinal images, and these dense features were reduced by PCA and singular value decomposition (SVD) (16), where it was fed to a deep Neural network (DNN) model that achieved accuracies of 97.96, and 98.34% in DR severity classification with PCA and SVD, respectively when training the model and testing on Kaggle dataset. Yaqoob et al. (17) introduced a feature representation extracted by ResNet-50 TL model (18) that was fed to Random Forest (RF) classifier for binary and multiclassification of DR. This approach achieved an accuracy of 96% when it was applied on a dataset comprised of two DR categories for detecting DR and accuracy of 75.09% when it was applied as on five DR category dataset for predicting the severity of DR.

Bodapati et al. (19) introduced a DR classification model that aggregates features extracted from multiple convolution blocks of TL models to enhance feature representation and hence improve DR detection. The model has compared various methodologies of pooling and feature aggregation and it was concluded that averaging pooling with simple fusion approaches using Deep Neural Networks (DNN) led to an improved performance. The authors of this work claimed that their approach for blending features from the convolution layers of the same TL model is simpler and better than the simple concatenation of features extracted from various TL models at a different scale that was presented early in (20). The latter approach introduced a multi-modal blended TL feature representation for extracting deep features from penultimate layers of multiple TL models and blending them using different pooling approaches to obtain the final DR image representation.

However, besides the local features presented by various TL models or the fusion of those features, the global image features have been playing an important role in DR detection. Those features are represented by the contour and structural features that describe retinal lesions like exudates (EXs), microaneurysms (MAs), and hemorrhages (HEMs), where the presence of DR disease is characterized by detecting one or more of these lesions. Thus, those global features (lesion features) are considered as good signs of retinal image lesions and hence can be successfully applied to improve the final accuracy of a TL-feature based DR screening system. Therefore, there were some attempts that used image segmentation techniques for extracting/detecting retinal lesion features that can be used for DR detection and staging (21, 22). In such image segmentation-based methods, a label is assigned to every pixel of an image based on pixel characteristics (23). The labels are encoded in a segmentation mask with equal dimensions to the image. In binary segmentation tasks, each mask pixel represents either the foreground (corresponds to an area-of-interest in the image;

value = 1) or the background (corresponds to all non-area-of-interest; value = 0). Thus, binary segmentation tasks are useful for extracting notable areas from biomedical images.

With the introduction of deep learning (DL), especially convolutional neural network (CNN), the DL based methods have resulted in an outstanding performance in medical image segmentation (24). UNET-architecture CNN models are one of DL models that achieve remarkable performance in medical image segmentation tasks (25). For the task of retinal image segmentation for DR detection, there were multiple research reports that demonstrate the use of UNET for segmentation of leakage-prone blood vessels (26, 27). UNET models have also been successfully developed for the segmentation of MAs (28–31), EXs (32–35), and HEMs (36).

Although TL and segmentation features provide robust information for DR detection, they were not used together in the literature for a two-class (binary) DR-detection. Only a few existing research attempts utilized the fusion of both types of features for improving the performance of predicting the severity levels of DR (37, 38). For example, Harangi et al. (37) proposed a framework that combines AlexNet TL-based features with image-level features that reflect the intensity, shape, and texture of the structures of the image and lesion-specific features associated with MAs and EXs. This combination of features was passed through an additional fully connected layer followed by a softmax function that achieved an accuracy of 90.07% in predicting the class probabilities corresponding to 5 classes for DR that express its severity levels. In Bogacsovics et al. (38), the same idea was extended to several commonly used TL models for local image feature extraction other than AlexNet. Next, the results of concatenating the TL features of those models with the hand-crafted features (image level and lesion features) were objectively compared to demonstrate the best concatenation framework that improves the accuracy of predicting DR grades. Next, the best concatenation was passed through an additional fully connected layer then a softmax function to predict the five class probabilities of DR severity. However, both approaches in (37, 38) were not tested for binary classification of DR to report the presence of the disease. Also, the performance of the lesion feature extraction was not explicitly investigated as well as the impact of those features on DR detection. Moreover, both approaches combined the image level features with TL and lesion features for training the DR severity levels predictors, increasing the curse of dimensionality.

Therefore, in this study we propose a framework called DR-detector (Figure 1) that employs the XGboost ML model (39) for an accurate detection of DR. The model is trained with a combination of the TL features, and three clinical lesion features that capture EXs, MAs, and HEMs. Such a combination is used to get a better representation of retinal image features that can be used to decide about the presence of DR disease. Thus, we seek the power of TL model to extract features that accurately capture the local textural retinal images while simultaneously
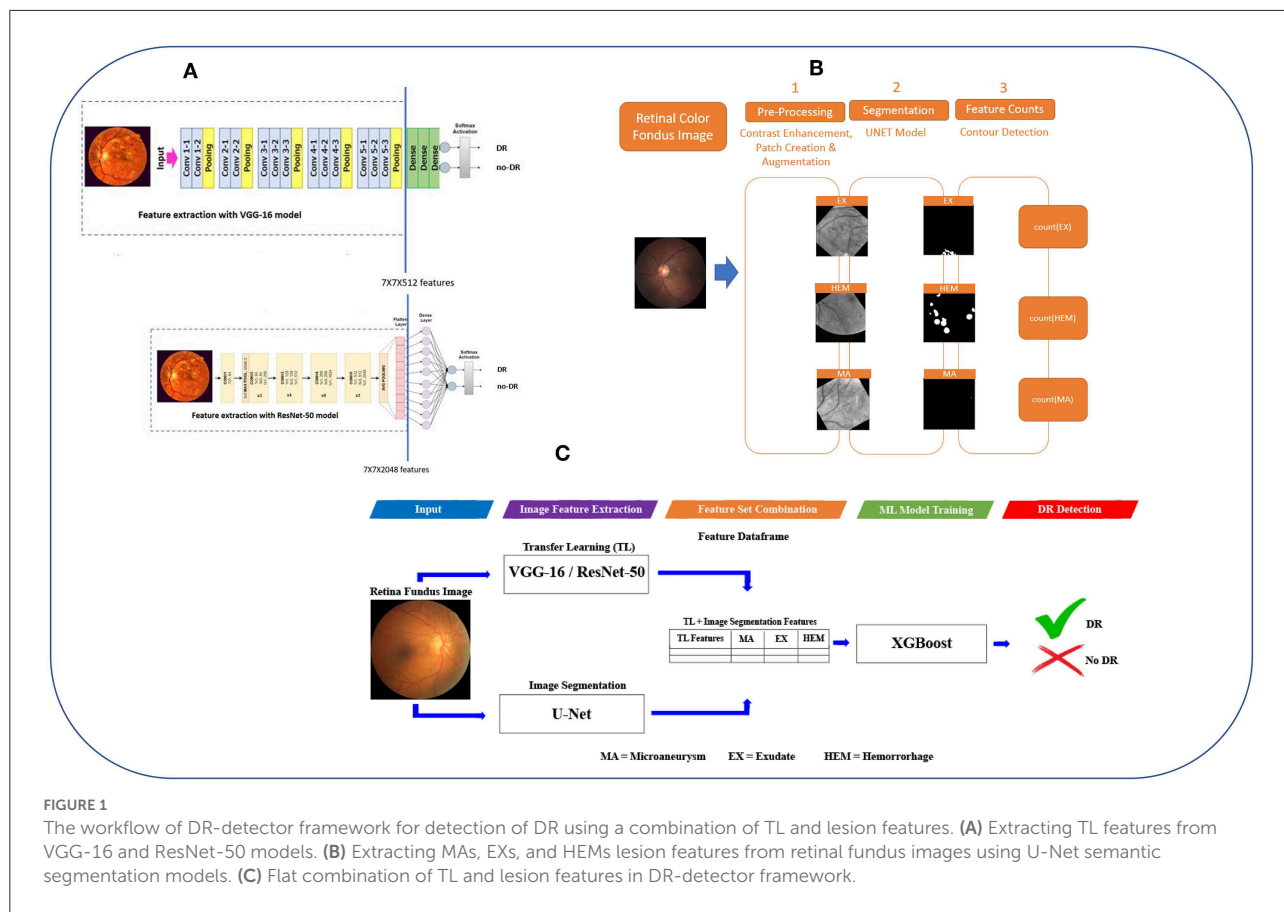
FIGURE 1
The workflow of DR-detector framework for detection of DR using a combination of TL and lesion features. **(A)** Extracting TL features from VGG-16 and ResNet-50 models. **(B)** Extracting MAs, EXs, and HEMs lesion features from retinal fundus images using U-Net semantic segmentation models. **(C)** Flat combination of TL and lesion features in DR-detector framework.

taking advantage of the power of lesion features to represent the global features of the retinal images that would result in improving the performance of the DR-detector model and its interpretability for clinical use. We tested the DR-detector model on an external dataset of retinal images for detection of DR. We have applied our proposed framework to two common TL models in the literature namely VGG-16 and ResNet-50 models (Figure 1A). The experimental results show that the DR-detector model achieves an accuracy of 100% in detecting DR when testing it on an external dataset after training it with the combination of Resnt-50 and lesion features and 99.38% accuracy after training it with the combination of VGG-16 and lesion features. The results also show a higher contribution of some lesion features toward the performance of the model over other lesion features. For instance, using the hemorrhages feature to train the model, our model achieves an accuracy of 99.38 in detecting DR which is higher than the accuracy of the model when training it with the combination of all lesion features (89%) and equal to the accuracy when training it with the combination of all lesions and VGG-16 features together. Thus, we arrived at two major conclusions. First, the extracted relevant lesion features can complement the textural features

extracted by the TL model to improve the performance of DR-detector model and its interpretability for clinical use in detecting the presence of DR disease. Second, the contribution of lesion features to the performance of DR-detector model varies from one lesion feature to another. This highlights the possibility of using only the lesion features for training the next generation of robust and accurate AI models with clinical interpretability for DR detection.

Materials and methods

Approach pipeline

The main objective of this work is to develop a robust and efficient framework called DR-detector for automatic detection of DR. Thus, we employed ML model namely the XGBoost in this framework to achieve this objective. This model is trained with a combination of two types of extracted features. The first type is deep convolutional features extracted using a TL model (VGG16-model or Resnet50) pre-trained previously on ImageNet dataset (40) (Figure 1A). Those deep features were

known as the most descriptive and discriminate features that ultimately improve the performance of DR recognition (16). The second type of features are three clinical lesion features that capture the EXs, MAs, and HEMs and are extracted using image segmentation *via* U-Net DL model (Figure 1B). Those lesion features were found to be the most common pathological signs of DR in the literature (41). Next the performance of DR-detector model is tested on external dataset of fundus retinal images after training it with the combination of TL and lesion features (Figure 1C). In summary, the proposed pipeline of the DR-detector framework (Figure 1) has five different modules including TL feature extraction (Figure 1A), lesion feature extraction (Figure 1B), and feature combination, model training and model evaluation (Figure 1C).

Datasets

Training dataset

We conduct our experiments on 1,840 color fundus images. 920 images of them have DR signs, and the remaining are healthy images. Those images were used for training the DR-detector model. The images with DR signs were collected from two public-available datasets, namely the e-ophtha[2], and retinal lesions (42, 43) with binary masks for extracting and quantifying the EXs, Mas, and HEMs lesion features. The healthy images were collected from APTOS 2019 Blindness Detection Kaggle competition training dataset (44). The Binary masks for healthy eye images are generated by creating all-black images with identical dimensions.

The images in the training dataset have various levels of DR on a scale of 0 to 4 [0 - No DR (healthy), 1-Mild, 2-Moderate, 3-Severe, 4-Proliferative DR] to indicate different DR severity levels. However, the data is imbalanced as it consists of 920 healthy images of DR-level 0, 120 images of DR-level 1, 681 images of DR-level 2, 64 images of DR level 3, 55 images of DR level 4. Since we found that there is not enough data from DR classes of DR-levels (1–4) that we can include in the training dataset to balance it, we decided to go with the binary classification of DR for automatic detection of this disease. Therefore, we adopt the training data set for binary classification problem by merging all images of DR signs of 1–4 into a single positive class of 920 images labeled as DR and the remaining 920 images are labeled as healthy and assigned to the negative class as shown in Table 1.

Testing dataset

For testing the DR-detector model, we have used a dataset of 162 color fundus images, where 81 of them are annotated as DR affected, and the remaining are from healthy individuals. Those images were collected from three publicly available

---

TABLE 1 Number of healthy and DR images in the training dataset.

DR severity level	Number of healthy and DR images
0	920 (Healthy)
1	920 (DR)
2	
3	
4	
Total	1,840

TABLE 2 Number of healthy and DR images in the testing dataset.

DR severity level	Number of healthy and DR images
0	81 (healthy)
1	81 (DR)
2	
3	
4	
Total	162

datasets, namely High-Resolution Fundus (HRF) dataset (45), Indian Diabetic Retinopathy Image Dataset (IDRID) dataset (46), and MESSIDOR-2 datasets (47). The IDRID contained 81 color fundus images (4,288 x 2,848) with binary masks representing DR-affected eyes needed to extract and quantify the EXs, Mas, and HEMs lesion features. However, IDRID dataset does not contain any healthy eye images, so the healthy eye images in the testing dataset were randomly selected from HRF and MESSIDOR-2 datasets. Binary masks for healthy eye images are generated by creating all-black images with identical dimensions.

Similar to the training data, the images in the testing dataset have 0–4 levels of DR to indicate different DR severity levels. However, the dataset is imbalanced as it consists of 81 healthy images of DR-level 0 and 81 images of DR affected images with 2 images of DR-level 1, 34 images of DR-level 2, 22 images of DR level 3, and 23 images of DR level 4. Since there is not enough data from DR classes of DR-levels (1–4) that we can include in the testing dataset to balance it, it has been more convenient to use such data for binary classification of DR for detection of the disease. To achieve this, we adopt the testing data set for binary classification problem by merging all images of DR levels of 1–4 into a single positive class of 81 images labeled as DR and the remaining 81 images are labeled as healthy and assigned to the negative class as shown in Table 2.

Image feature extraction with transfer learning

In this approach, local representations of the retinal image's features are obtained from the TL model (either the

VGG16 or the Resent50 pretrained models) by extracting deep features from the final layers of the pre-trained models. When performing feature extraction with TL models, we treat the pre-trained network as an arbitrary feature extractor, allowing the input image to propagate forward, stopping at pre-specified layer, and taking the outputs of that layer as our features.

As for extracting deep features using VGG-16 pretrained model, the original VGG-16 model (15) is adopted first to address the automatic detection of DR (top subfigure of Figure 1A). For this task, the model expects input images of 224*224*3. Thus, images are reshaped to 224*224*3 before feeding them to this model. Next the soft-max layer and fully connected (FC) layers are removed from VGG-16 model (area after the solid vertical blue line in top subfigure of Figure 1A) and the model utilizes the VGG-16 network (15) for feature extraction *via* the final layer prior to the FC layers—that outputs volume of size 7 x 7 x 512 dim (area with dashed border in top subfigure of Figure 1A). This output will serve as VGG-16 extracted features which will be flattened later into a feature vector of 25,088-dim combined with the lesion features, as described later in section Combination of TL and lesion features.

As for extracting deep features using the ResNet-50 pretrained model, the original ResNet-50 model (17) is adopted first to address the detection of DR task (bottom subfigure of Figure 1A). For this task, the model expects input images of 224*224*3. Thus, images are reshaped to 224*224*3 before feeding them to this model. Next, the soft-max layer and fully connected (FC) layers are removed from ResNet-50 model (area after the solid vertical blue line in the bottom subfigure of Figure 1A) and the model utilizes the ResNet-50 network (18) for feature extraction *via* the final layer before the fully connected (FC) layers—that outputs volume of size 7 x 7 x 2,048 dim (area with dash border in bottom subfigure of Figure 1A). This output will serve as ResNet-50 extracted features which will be flattened later into a feature vector of 100, 352-dim combined with lesion features as described later in section Combination of TL and lesion features.

Lesion feature extraction with image segmentation

Retinal image lesions associated with DR

Retinal lesions that develop early over the course of DR, including MAs, EXs, and HEMs (Supplementary Figure 1), are clinically important markers used to distinguish between healthy and DR-affected eyes. Below, we elaborate about each of these three lesions:

Microaneurysms

MA are the earliest symptoms of DR. These lesions are widened protrusions extending from capillary walls and are associated with abnormal fluid leakage through breakdown of

the blood-retina barrier. MA can rupture to create hemorrhages, leading to greater leakage of capillary fluids and damage to surrounding retinal tissues. The number of microaneurysms can be used to gauge the progression of DR (48).

Exudates

EXs are lipids and proteins (fibrinogen, albumin) carried by filtrating fluids past the blood-retina barrier into the retinal tissue (49).

Hemorrhages

HEMs occur when MA burst, and leak blood and serum into the retina. Intraretinal bleeding is a sign of worsening DR. Blood can impair DR patient vision and increased intraretinal pressure can contribute to retinal damage (50).

Framework for UNET- model-based lesion detection and quantification

We have developed a framework to extract MAs, EXs, and HEMs lesions from retinal fundus images using U-Net semantic segmentation models (Figure 1B) and deployed it in DR-detector framework for extracting lesion features. The steps for the UNET-based retinal lesion detection and quantification workflow are described below:

Pre-processing

Binary thresholding is applied to set all pixels corresponding to the image background (the non-eye margins) equal to zero. Multiple studies demonstrate the green channel encodes the greatest contrast between retinal structures (32–34). Input RGB retinal fundus images are split by channel and the green channel is extracted. Contrast Limited Adaptive Histogram Equalization (CLAHE) (8x8 tile size) (51) is applied to the green channel to correct the contrast over-amplification. A gamma correction is utilized to adjust luminescence of the CLAHE output ($\gamma = 0.8$). The contrast enhancement stages are shown in Supplementary Figure 2. After contrast enhancement, images are divided into patches.

Each preprocessed retinal image and its corresponding ground truth mask is divided into overlapping square (n x n) patches. n is set to 128 pixels (px) for MAs (due to small lesion size) and is set to 256 px for EXs and HEMs. Created patches are randomly selected for augmentation operations.

Augmentation for image and corresponding binary mask patches involves creating new training instances from existing ones by applying a spatial or color operation to represent them in a new orientation or perspective. The random flip (horizontal, vertical) and random rotation (360°) techniques from Keras ImageDataGenerator (52) are used to augment training patches. Any augmentation technique applied to a fundus image patch is likewise applied to its ground truth patch.

Segmentation

The input retinal fundus images are preprocessed and divided into augmented patches as described above. K-fold cross-validation ($k = 5$) is applied to the patches. The batch size for each fold is set to 32 and the number of epochs is set to 3. Epoch training and validation steps are set as the number of training or validation patches per fold divided by batch size, respectively.

Patch probability maps output by the UNET DL model are merged to construct a probability map with equal dimensions to the input image. A threshold of 0.5 is applied to convert the reconstructed probability map into a binary image mask.

Feature counts

Canny edge detection (53) is applied to find the edges around mask foreground regions (white). Contour detection (54) is used to fill Canny edge gaps and fully close the foreground shapes. The number of lesions within the segmentation mask is defined as the number of distinct objects described by contours.

U-net model implementation

Keras (55), the free python deep learning API with TensorFlow (56) back end was used to construct a base UNET model for binary semantic segmentation. Input patches are supplied to the input layer as tensors with shape (32, n, n, 1). The UNET contracting path used for downsampling is defined by 5 convolution blocks, each with 2 (3 x 3) convolution layers (activation = "ReLU," padding = "same") and a (2 x 2) pooling layer. The UNET expansive path used for upsampling is defined by 5 convolution blocks, each with a (2 x 2) transpose convolution layer (activation = "ReLU," padding = "same"), concatenation layer, and 2 (3 x 3) convolution layers (activation = "ReLU," padding = "same"). A (1 x 1) output layer using a sigmoidal activation function returns the model output. Binary focal entropy is selected as the loss function due to large class imbalance between foreground and background pixels.

The model output is a pixelwise probability map for each input patch. The probability map values range from 0 to 1; values closer to 0 represent pixels more likely to belong to the negative class and pixels closer to 1 represent pixels more likely to belong to the positive class.

Metrics for performance evaluation of U-net model

We evaluate the performance of U-Net DL model in extracting each of the three lesions in terms of accuracy, recall, precision, F1-score, and IoU. The mathematical equations that describe each of these metrics are shown below:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Precision} = = \frac{TP}{TP + FP} \quad (3)$$

$$\text{F1} - \text{score} = 2*\frac{\text{Precision*Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{Intersection over union (IoU)} = \frac{A \cap A'}{A \cup A'} \quad (5)$$

Where:

TP = True Positives (the sum of positive (foreground) pixels classified by the model)

TN = True Negatives (sum of correctly classified negative pixels)

FP = False Positives (the sum of negative (background) pixels misclassified by the model)

FN = False Negatives (the sum of misclassified positive pixels)

A = Area of ground truth pixels and A′ = Area of predicted pixels.

High recall values indicate that most of pixels belonging to the positive class (lesions) are predicted correctly by U-Net segmentation models. High precision values across the three lesion types also demonstrate the U-Net models successfully differentiate between lesion foreground and non-lesion background regions. High accuracy and F1 scores values reflect the excellent model performance and robustness. IoU is a useful metric for image segmentation by measuring the overlap between predicted and ground truth segmentation masks. This can be done for a class by dividing the intersection (overlap) of ground truth and predicted pixels belonging to the class by the total number of pixels in both masks belonging to the class. IoU score ranges from [0–1], where scores closer to 1 indicate greater agreement between predicted and ground truth masks.

Combination of TL and lesion features

The DR-detector framework performs a fusion of the TL features with lesion features to get a better representation of the features used for detection of DR. This is achieved by concatenating the flat representation of the features obtained by TL model (either VGG-16 or ResNet-50) with three lesion features that captures EXs, MAs and HEMs *via* image segmentation technique (Figure 1C).

By combining the TL features and the lesion features, the resulting feature vector for each image in the training dataset that will be used for training the XGboost DR-detector model would be of size = 25,088 + 3 = 25,091 dim for VGG-16 and 100,352 + 3 = 100,355 dim for ResNet-50.

Metrics for performance evaluation of DR detection

The accuracy, precision, recall, F1-score and the area under the ROC curve (AUC) (57) have been used to evaluate the Xgboost model's performance deployed in the DR detector framework. The mathematical equations that define each of the first four metrics were previously defined in Equations (1–4) for evaluating the performance of U-Net segmentation models for lesion types. However, TP,FP,FN and TN terms for DR-detection task have a different indication from their previous annotation for evaluting the image segmentation task using U-Net model and hence are described below:

TP refers to the number of correctly classified DR images.

FP refers to the number of healthy images misclassified as DR images

FN refers to the number of DR images misclassified as healthy images

TN refers to the number of correctly classified healthy images.

As for the AUC metric, it measures the entire two-dimensional area under the ROC curve (58) which measures how accurately the model can distinguish between DR and healthy images.

Experimental setup and model development

All experiments in this study were executed on Ubuntu Linux server with 128 GB of RAM, 16 Intel Xeon E5-2609 1.7GHZ CPU cores, and 8 GPU cards. The optimized distributed gradient boosting python library (59) has been used for implementing the XGBoost model. The scikit-learn toolkit[3], the free machine learning python library has been used for implementing other ML models that were developed as a proof of concept to show that XGBoost was chosen because it outperforms other ML models (see Tables 4, 5). The Keras free python library (55) with tensorflow back end (56) was used to implement the TL models as well as the deep neural network (DNN) models to compare the performances in DR detection to the XGboost as an evidence to show the outperformance of later model.

Results

U-net model performance evaluation

U-Net model performance evaluation results are shown in Table 3. The table shows an outperformance of the U-Net model for predicting MAs and EXs lesions over HEMs lesions.

---

3   https://scikit-learn.org/stable/

In other words, the general trend we observed was that the performance evaluation results for the HEMs segmentation model were consistently lower than those for the MAs and EXs models. We attribute this to the variation in the appearance of retinal hemorrhages, which can range from small, concentrated regions (dot hemorrhages) to larger and more irregular shapes. Examples of U-Net model predictions are shown in Supplementary Figure 3.

DR detection using VGG-16 and lesion features

As a proof of concept, we developed different ML models and trained them on the combination of VGG-16 and lesion features including the support vector machine (SVM), K-Nearest Neighbors (KNN), Xgboost, Logistic Regression (LR), Multi-Layered Perceptron (MLP), Decision Tree (DT), and Random Forest (RF) with default settings in addition to a DNN model with different structures including one input layer with 128 nodes, one input layer with 256 nodes and one hidden layer with 128 nodes, and one input layer with 512 nodes and 2 hidden layers with 256 and 128 nodes, respectively. As shown in Table 4, XGBoost outperformed all other ML and DNN models achieving 99.38% accuracy in detecting DR.

DR detection using ResNet-50 and lesion features

Table 5 shows the performance evaluation results of all DR detection models using the combination of ResNet-50 and lesion features. As can be observed from the table, XGBoost continued to outperform all other ML and DNN models achieving 100% accuracy for detection of DR.

Performance of DR-detector with a single type of feature to understand feature importance and contribution

We decided to deploy XGBoost model in DR-detector framework for automatic detection of DR since it outperforms all other ML and DNN models as highlighted in the previous sections and documented in Tables 4, 5. For deep analysis of the features that highly contribute to the performance of XGBoost model in the detection of DR, we have analyzed the importance of each feature by evaluating the XGBoost model performance with each type of extracted features including the TL and lesion features. This has been achieved by building three versions of the XGBoost model, where each version of the model is trained with only one type of feature. Figure 2 shows a bar chart that

TABLE 3 Performance evaluation results of UNET model on extracting each type of lesions feature with the best 5-fold cross validation accuracy from all trials.

Lesion type	Recall	Precision	F1-score	Accuracy	IoU
MAs	0.94	0.93	0.98	0.98	0.84
EXs	0.94	0.95	0.98	0.98	0.83
HEMs	0.84	0.86	0.85	0.95	0.81

TABLE 4 The performance of all Ml models and DNN for detection of DR using a combination of VGG-16 and retinal lesion features.

Classifier	Accuracy %	Precision	Recall	F1-score	AUC
Xgboost	**99.38**	**1**	**0.99**	**0.99**	**0.994**
KNN	47.53	0.45	0.25	0.32	0.475
LR	51.85	0.52	0.57	0.54	0.519
SVM	43.21	0.45	0.62	0.52	0.432
MLP	53.09	0.52	0.93	0.66	0.531
DT	88.27	1.00	0.77	0.87	0.883
RF	68.52	0.71	0.63	0.67	0.685
DNN (1 layer)	48	0.48	0.58	0.53	0.481
DNN (2 layers)	50	0.5	0.38	0.43	0.5
DNN (3 layers)	52	0.52	0.56	0.54	0.519

The bold values highlight the best accuracy, precision, recall, F1-score, and AUC of ML and DNN learning models for detection of DR using a combination of VGG-16 and retinal lesion features.

represents the performance of XGBoost with VGG-16, ResNet-50 and lesion features (Supplementary Table 1). Clearly, we see a significant outperformance of the lesion features over the TL features that have been either extracted by VGG-16 or Resnt-50 models. This highlights the importance of the clinically manifested symptoms reflected in the form of lesion features in the detection of DR and how they can complement the textural features extracted by the TL model to improve the XGBoost performance and its interpretability for clinical use in detecting the presence of DR disease.

Performance of DR-detector with all possible feature combinations

Figure 3 shows a bar chart that represents the performance of XGBoost (DR-detector model) on the testing dataset with all possible combinations of TL and lesion features (Supplementary Table 2). Clearly, the figure shows that (Resent-50 and lesion) feature combination leads to the best performance of XGBoost model among all combinations and is slightly better than the performance of the model using (VGG-16 and lesions), and (VGG-16, ResNet-50, and lesions) combinations that lead to equal model performances. The figure also shows poor performance of XGBoost with VGG-16 and ResNet-50 combination (i.e., when the lesion features are specifically

excluded) which also highlights the importance of lesion features in the detection of DR.

Lesion feature importance and their effect on XGBoost performance

For deep analysis of the contribution of each lesion feature to the performance of DR-detector model in DR detection, we have reported the importance of each lesion feature through training and testing the XGBoost model with each type of lesion feature individually. This has been achieved by building three versions of the XGBoost model, where each version of the model is trained with one lesion feature. Figure 4 shows the performance of XGBoost with EXs, MAs and HEMs lesion features as well as with their combinations (Supplementary Table 3). Clearly, the table shows a significant outperformance of XGBoost model using HEMs lesion feature over its performance using either the MAs or EXs lesion features which equally contribute to the performance of the model. More importantly, the model also achieves better performance using HEMs alone than its performance using the combination of the three lesion features altogether. Thus, our results highlight the importance of HEMs as a feature in the detection of DR and how it can be used to allow the interpretability of the model for clinical use in DR detection.

TABLE 5 The performance of all Ml models and DNN for detection of DR using a combination of ResNet-50 and retinal lesion features.

Classifier	Accuracy %	Precision	Recall	F1-score	AUC
Xgboost	**100**	**1.00**	**1.00**	**1.00**	**1.0**
KNN	42.59	0.37	0.21	0.27	0.426
LR	53.09	0.52	0.68	0.59	0.531
SVM	45.68	0.48	0.84	0.61	0.457
MLP	47.53	0.49	0.95	0.64	0.475
DT	96.30	1	0.93	0.96	0.963
RF	56.79	0.55	0.70	0.62	0.568
1 layer-DNN	43.8	0.42	0.33	0.37	0.44
2 layer-DNN	48	0.49	0.75	0.59	0.481
3 layer-DNN	49	0.49	0.46	0.47	0.488

The bold values highlight the best accuracy, precision, recall, F1-score and AUC of ML and DNN learning models for detection of DR using a combination of Resnet-50 and retinal lesion features.

FIGURE 2
The performance of XGBoost with VGG-16, ResNet-50 and lesions features.
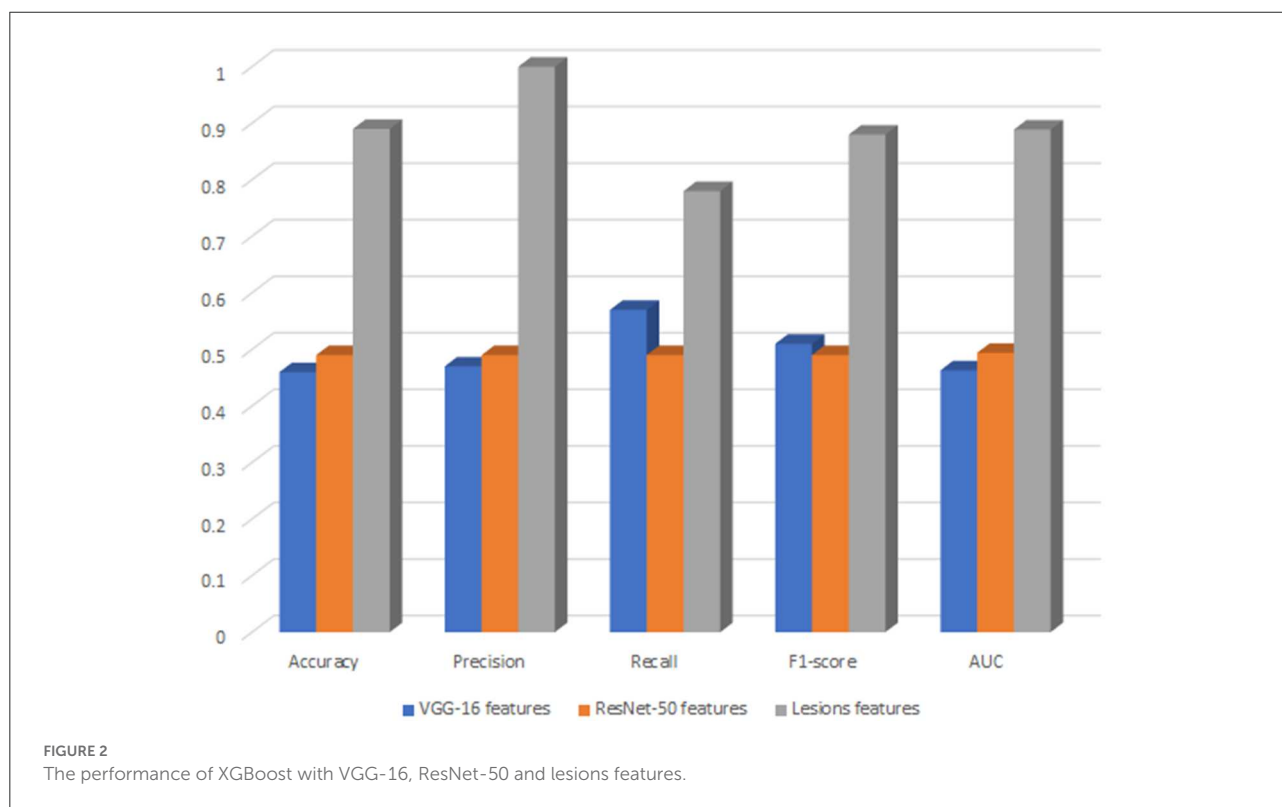
Discussion

There are several observations that can be summarized from our experimental results. First, using our proposed approach we found that the integration of lesion features with the TL features significantly improves the performance of the DR-detector model and adds a clear importance to its clinical interpretability. However, currently it is not possible to only work with the lesion features for training the DR-detector model as there are few retinal imaging datasets in

the literature that provide the image masks corresponding to the retinal images in the dataset that are needed to extract and quantify the lesion features using the U-Net segmentation models and obtaining those masks need the involvement of trained ophthalmologists which is costly and time-consuming.

It was also observed that deploying the XGBoost model as an ensemble of ML classifiers in the DR-detector framework led to a better binary classification of DR and error detection than deploying dense classifiers of DL models as introduced in
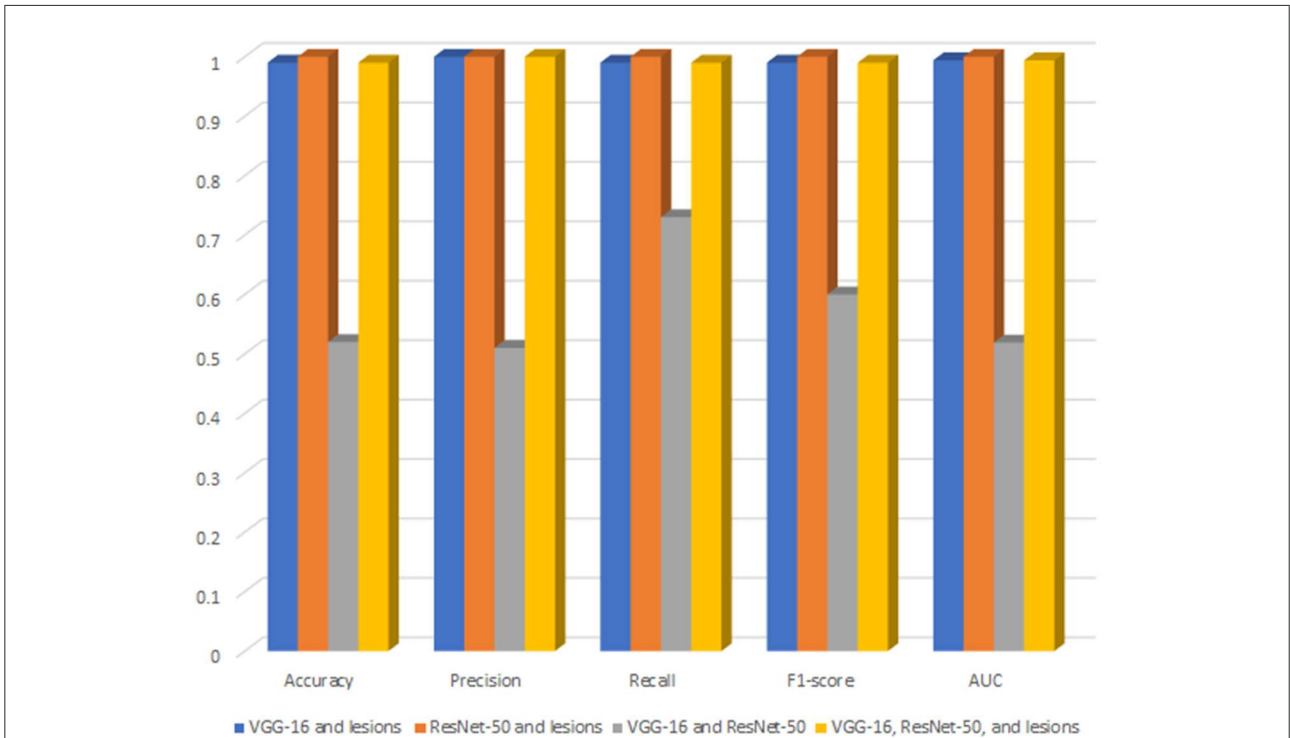
FIGURE 3
The performance of XGBoost with all possible combinations of TL and lesion features.
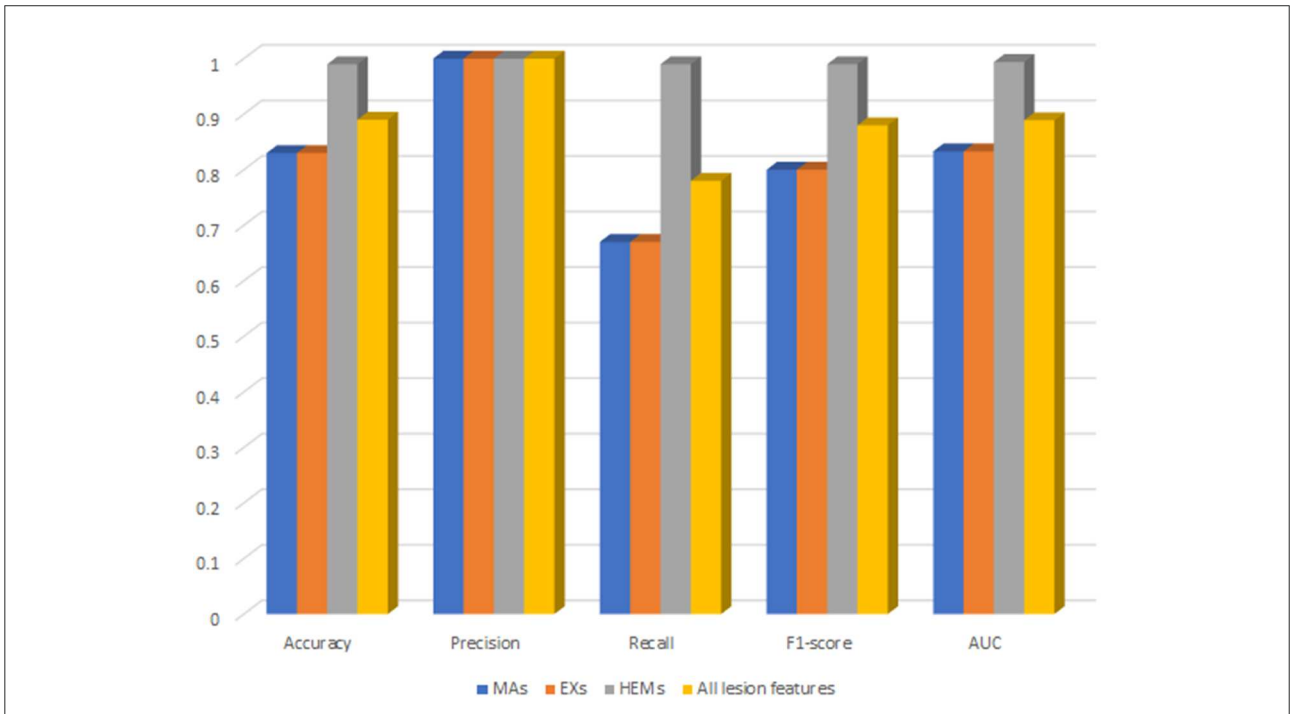
FIGURE 4
The performance of XGBoost with each type of lesion features as well as their combinations.

few previous studies (20, 37, 38). This motivated us to use the flat combination of TL and lesion features (Figure 1C) which is simpler, straight forward and more convenient to be applied to the XGBoost ML model than combining both types of features using different pooling approaches (20), or by extending the last fully connected (FC) layer of TL model with additional number of neurons equal to the number of lesion features and using softmax function to obtain the predictive DR class probabilities (37).

Regarding the contribution of lesion features to the DR-detector (XGBoost) model performance, our approach shows that lesion features have different contributions to the model performance. Particularly, it shows a significant contribution of the hemorrhages over the other two lesion features for binary identification of DR (Supplementary Table 3). This is likely because hemorrhages may arise early during the progression of DR and are associated with the worsening of vision and the development of other vision-threatening lesions. Our observations highlight the importance of such a feature to train the binary classifier of DR-detector framework and adds value for its clinical usage in DR diagnosis. However, it remains to be seen how our presented results will hold when testing our lesion feature-based models on different fundus retinal testing datasets with more complex demographics and varying quality.

It is also noteworthy to mention that although we have applied the DR-detector framework to a relatively small training and testing datasets in comparison with the large datasets that have been used in the literature for DR detection [e.g.,(17)] our datasets are from heterogenous sources i.e., have different variety of retinal images that were imported from multiple publicly available datasets with different settings of capturing the fundus retinal images. This, of course, highlights the efficiency of our proposed framework in detecting the initial signs of DR, even when applied to a set of retinal fundus images that were not imported from the same resource. Thus, we expect a better performance of DR-detector framework with larger training and testing datasets in the future.

Conclusions and future work

In this study, we have proposed a framework called DR-detector that combines the features extracted from fundus retinal image by transfer learning model and the lesion features extracted using semantic image segmentation *via* U-Net DL model for accurate detection of DR using the XGboost ML model deployed in this framework. The model was trained using the combination of both features on a training dataset collected from various publicly available datasets and was tested on an external dataset that consists of DR images from IDRID dataset and healthy images from HRF and Messior-2 dataset. Our experimental results show that our proposed framework for DR detection achieves a testing accuracy of 100% in detecting DR

using the combination of Resnt-50 and lesion features and 99% testing accuracy using the combination of VGG-16 and lesion features. Based on these results, we arrived at the conclusion that the extracted clinically relevant lesion features have a significant impact on the performance of the DR-detector model and would be an excellent complement to the textural features extracted by the TL model to improve the performance of DR-detector model and its interpretability for clinical use for detecting the presence of DR disease.

We anticipate a natural extension of our current work is to first extend to other TL models that are commonly used in the literature to study how they perform for DR classification task and then expand our general framework to be applied for the prediction of different severity levels of DR. Finally, we are looking forward to applying our approach to combine TL features with other types of DR lesion features such as cotton wool spots, foveal avascular zone, optic disc, and retinal blood vessels since these features were known in the literature to be associated with greater severity of DR (60). This might lead to better performance of the DR-detector model and its interpretability for clinical use on much larger and clinically diverse datasets, especially if it will be extended to predict the various grades of DR.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

DH, HG, MH, AB, AH, and SJ conceived and designed the study. DH did the TL feature extraction using VGG-16 and ResNet-50 models and combined TL and lesion features for training and evaluating the XGboost model of DR-detector framework. HG did the lesion feature extraction using image segmentation *via* U-Net model. DH and HG wrote the manuscript and contributed to the DR-detector GitHub repository. AB, AH, and SJ revised the manuscript and provided valuable comments. All authors contributed to the article and approved the submitted version.

Funding

IUPUI Institute of Integrative Artificial Intelligence (iAI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

AB is an *ad hoc* Staff Pharmacist at CVS Health/Aetna.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The content of this study does not reflect those of CVS Health/Aetna.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.1050436/full#supplementary-material

References

1. Hartnett ME, Baehr W, Le YZ. Diabetic retinopathy, an overview. *Vision Res.* (2017) 139:1–6. doi: 10.1016/j.visres.2017.07.006

2. Nentwich MM, Ulbig MW. Diabetic retinopathy - ocular complications of diabetes mellitus. *World J Diabetes.* (2015) 6:489–99. doi: 10.4239/wjd.v6.i3.489

3. Matuszewski W, Baranowska-Jurkun A, Stefanowicz-Rutkowska MM, Modzelewski R, Pieczyński J, Bandurska-Stankiewicz E. Prevalence of diabetic retinopathy in type 1 and type 2 diabetes mellitus patients in North-East Poland. *Medicina.* (2020) 56:164. doi: 10.3390/medicina56040164

4. Hietala K, Harjutsalo V, Forsblom C, Summanen P, Groop PH, FinnDiane Study Group. Age at onset and the risk of proliferative retinopathy in type 1 diabetes. *Diabetes Care.* (2010) 33:1315–9. doi: 10.2337/dc09-2278

5. Grossniklaus HE, Geisert EE, Nickerson JM. Introduction to the retina. *Prog Mol Biol Transl Sci.* (2015) 134:383–96. doi: 10.1016/bs.pmbts.2015.06.001

6. Kovoor E, Chauhan SK, Hajrasouliha A. Role of inflammatory cells in pathophysiology and management of diabetic retinopathy. *Surv Ophthalmol.* (2022) 67:1563–73. doi: 10.1016/j.survophthal.2022.07.008

7. Eisma JH, Dulle JE, Fort PE. Current knowledge on diabetic retinopathy from human donor tissues. *World J Diabetes.* (2015) 6:312–20. doi: 10.4239/wjd.v6.i2.312

8. Das A. Diabetic retinopathy: battling the global epidemic. *Invest Ophthalmol Vis Sci.* (2016) 57:6669–82. doi: 10.1167/iovs.16-21031

9. Alban M, Gilligan T. *Automated Detection of Diabetic Retinopathy Using Fluorescein Angiography Photographs.* Standford Education (2016).

10. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* (2018) 172:1122–31. doi: 10.1016/j.cell.2018.02.010

11. Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using deep learning. *AMIA Jt Summits Transl Sci Proc.* (2018) 2018:147–55. Available online at: https://pubmed.ncbi.nlm.nih.gov/29888061/

12. Aswathi T, Swapna TR, Padmavathi S. Transfer learning approach for grading of Diabetic Retinopathy. *J Phys.* (2021) 1767:012033. doi: 10.1088/1742-6596/1767/1/012033

13. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM.* 60, 84–90. doi: 10.1145/3065386

14. Mansour RF. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomed Eng Lett.* (2018) 8:41–57. doi: 10.1007/s13534-017-0047-y

15. Simonyan K, Zisserman A. Very deep convolutional networks for large scale image recognition. In: *Proceedings of ICLR.* San Diego, CA (2015).

16. Mateen M, Wen J, Nasrullah, Song S, Huang Z. Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry.* (2019) 11:1. doi: 10.3390/sym11010001

17. Yaqoob MK, Ali SF, Bilal M, Hanif MS, Al-Saggaf UM. ResNet based deep features and random forest classifier for diabetic retinopathy detection. *Sensors.* (2021) 21:3883. doi: 10.3390/s21113883

18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Las Vegas, NV (2016).

19. Bodapati JD, Shaik NS, Naralasetti V. Deep convolution feature aggregation: an application to diabetic retinopathy severity level prediction. *Signal Image Video Process.* (2021) 15:923–30. doi: 10.1007/s11760-020-01816-y

20. Bodapati JD, Veeranjaneyulu N, Shareef SN, Hakak S, Bilal M, Maddikunta PKR, et al. Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics.* (2020) 9:914. doi: 10.3390/electronics9060914

21. Liskowski P. Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging.* (2016) 35:2369–80. doi: 10.1109/TMI.2016.2546227

22. Wei Q, Li X, Yu W, Zhang X, Zhang Y, Hu B, et al. Learn to segment retinal lesions and beyond. In: *Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR).* Milan, Italy (2021). p. 7403–10.

23. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. *J Med Phys.* (2010) 35:3–14. doi: 10.4103/0971-6203.58777

24. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustainability.* (2021) 13:1224. doi: 10.3390/su13031224

25. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Cham: Springer (2015).

26. Tang MCS, Teoh SS, Ibrahim H, Embong Z. Neovascularization detection and localization in fundus images using deep learning. *Sensors.* (2021) 21:5327. doi: 10.3390/s21165327

27. Burewar S, Gonde AB, Vipparthi SK. Diabetic retinopathy detection by retinal segmentation with region merging using CNN. In: *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS).* Rupnagar (2018). p. 136–42.

28. Kou C, Li W, Liang W, Yu Z, Hao J. Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network. *J Med Imaging.* (2019) 6:025008. doi: 10.1117/1.JMI.6.2.025008

29. Albahli S, Ahmad Hassan Yar GN. Automated detection of diabetic retinopathy using custom convolutional neural network. *J Xray Sci Technol.* (2022) 30:275–91. doi: 10.3233/XST-211073

30. Li Q, Fan S, Chen C. An intelligent segmentation and diagnosis method for diabetic retinopathy based on improved U-NET network. *J Med Syst.* (2019) 43:304. doi: 10.1007/s10916-019-1432-0

31. Jiang Y, Wang F, Gao J, Cao S. Multi-path recurrent U-Net segmentation of retinal fundus image. *Appl Sci.* (2020) 10:3777. doi: 10.3390/app10113777

32. Kou C, Li W, Yu Z, Yuan L. An enhanced residual U-Net for microaneurysms and exudates segmentation in fundus images. *IEEE Access.* (2020) 8:185514–25. doi: 10.1109/ACCESS.2020.3029117

33. Rahebi J, Hardalaç F. Retinal blood vessel segmentation with neural network by using gray-level co-occurrence matrix-based features. *J Med Syst.* (2014) 38:85. doi: 10.1007/s10916-014-0085-2

34. Xu Y, Zhou Z, Li X, Zhang N, Zhang M, Wei P. FFU-Net: Feature Fusion U-Net for lesion segmentation of diabetic retinopathy. *BioMed Res Int.* (2021) 2021:6644071. doi: 10.1155/2021/6644071

35. Zong Y. U-net based method for automatic hard exudates segmentation in fundus images using inception module and residual connection. *IEEE Access.* (2020) 8:167225–35. doi: 10.1109/ACCESS.2020.3023273

36. Erwin S, Desiani A, Suprihatin BF. The augmentation data of retina image for blood vessel segmentation using U-Net convolutional neural network method. *Int J Comput Intell Appl.* (2022) 21:2250004. doi: 10.1142/S1469026822500043

37. Harangi B, Toth J, Baran A, Hajdu A. Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.* Berlin (2019). p. 2699–702.

38. Bogacsovics G, Toth J, Hajdu A, Harangi B. Enhancing CNNs through the use of hand-crafted features in automated fundus image classification. *Biomed Signal Process Control.* (2022) 76:103685. doi: 10.1016/j.bspc.2022.103685

39. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16).* San Francisco, CA (2016).

40. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. (2009). ImageNet: A large-scale hierarchical image database, in *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE),* 248–255. doi: 10.1109/CVPR.2009.5206848

41. Khojasteh P, Aliahmad B, Kumar DK. Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmol.* (2018) 18: 288. doi: 10.1186/s12886-018-0954-4

42. Wei Q, Li X, Yu W, Zhang X, Zhang Y, Hu B, et al. Learn to segment retinal lesions and beyond. In: *Proceedings of International Conference on Pattern Recognition (ICPR).* (2013) 34:196–203.

43. *Retinal Lesions.* Available online at: https://github.com/WeiQijie/retinal-lesions (accessed October 30, 2022).

44. *Aptos 2019- Blindness Detection.* Available online at: https://www.kaggle.com/c/aptos2019-blindness-detection/data (accessed October 30, 2022).

45. *High Resolution Fundus Image Database.* Available online at: https://www5.cs.fau.de/research/data/fundus-images/ (accessed October 30, 2022).

46. *Diabetic Retinopathy – Segmentation and Grading Challenge.* Available online at: https://idrid.grand-challenge.org (accessed October 30, 2022).

47. *Messidor – 2.* Available online at: https://www.adcis.net/en/third-party/messidor2/ (accessed October 30, 2022).

48. Sjølie AK, Klein R, Porta M, Orchard T, Fuller J, Parving HH, et al. Retinal microaneurysm count predicts progression and regression of diabetic retinopathy. Post-hoc results from the DIRECT Programme. *Diabet Med.* (2011) 28:345–51. doi: 10.1111/j.1464-5491.2010.03210.x

49. Idicula J, Nithyanandam S, Joseph M, Mohan VA, Vasu U, Sadiq M. Serum lipids and diabetic retinopathy: a cross-sectional study. *Indian J Endocrinol Metab.* (2012) 16:S492–4. doi: 10.4103/2230-8210.104142

50. Murugesan N, Üstunkaya T, Feener EP. Thrombosis and hemorrhage in diabetic retinopathy: a perspective from an inflammatory standpoint. *Semin Thromb Hemost.* (2015) 41:659–64. doi: 10.1055/s-0035-1556731

51. Zuiderveld K. Contrast limited adaptive histogram equalization. In: Heckbert PS, editor. *Graphics Gems IV.* Academic Press Professional, Inc (1994). p. 474–85.

52. *tf.keras.preprocessing.image.ImageDateGenerator.* Available online at: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (accessed October 30, 2022).

53. *Canny Edge Detection- OpenCV.* Available online at: https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html (accessed October 30, 2022).

54. *Contour Detection Using OpenCV (Python/C++).* Available online at: https://learnopencv.com/contour-detection-using-opencv-python-c/ (accessed October 30, 2022).

55. *Keras: The Python Deep Learning API.* Available online at: https://keras.io (accessed October 30, 2022).

56. *Tensorflow.* Available online at: https://www.tensorflow.org/ (accessed October 30, 2022).

57. Andrew E. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn.* (1997) 30:1145–59. doi: 10.1016/S0031-3203(96)00142-2

58. *Receiver Operating Characteristic.* Available online at: https://en.wikipedia.org/wiki/Receiver_operating_characteristic (accessed October 30, 2022).

59. Jain A. *Complete Guide to Parameter Tuning in XGBoost With Codes in Python.* (2016). Available online at: https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/

60. Das D, Biswas SK, Bandyopadhyay S. A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning. *Multimed Tools Appl.* (2022) 23:1–43. doi: 10.1007/s11042-022-12642-4