A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification From Fundus Images

GHULAM ALI[1], AQSA DASTGIR[1], MUHAMMAD WASEEM IQBAL[2], MUHAMMAD ANWAR[3], AND MUHAMMAD FAHEEM[4]

[1]Department of Computer Science, University of Okara, Okara 56310, Pakistan
[2]Department of Software Engineering, Superior University, Lahore 54000, Pakistan
[3]Department of Information Sciences, Division of Science and Technology, University of Education, Lahore 54000, Pakistan
[4]School of Technology and Innovations, University of Vaasa, 65200 Vaasa, Finland

CORRESPONDING AUTHOR: M. FAHEEM (muhammad.faheem@uwasa.fi)

ABSTRACT *Objective:* Diabetic Retinopathy (DR) is a retinal disease that can cause damage to blood vessels in the eye, that is the major cause of impaired vision or blindness, if not treated early. Manual detection of diabetic retinopathy is time-consuming and prone to human error due to the complex structure of the eye. *Methods & Results:* various automatic techniques have been proposed to detect diabetic retinopathy from fundus images. However, these techniques are limited in their ability to capture the complex features underlying diabetic retinopathy, particularly in the early stages. In this study, we propose a novel approach to detect diabetic retinopathy using a convolutional neural network (CNN) model. The proposed model extracts features using two different deep learning (DL) models, Resnet50 and Inceptionv3, and concatenates them before feeding them into the CNN for classification. The proposed model is evaluated on a publicly available dataset of fundus images. The experimental results demonstrate that the proposed CNN model achieves higher accuracy, sensitivity, specificity, precision, and f1 score compared to state-of-the-art methods, with respective scores of 96.85%, 99.28%, 98.92%, 96.46%, and 98.65%.

INDEX TERMS Diabetic retinopathy, fundus images, machine learning, computer vision.

Clinical and Translational Impact Statement— Diabetic Retinopathy is becoming a more common cause of visual impairment in working-age individuals. Optimal results for preventing diabetic vision loss require patients to undergo extensive systemic care. Early detection and treatment are the key to preventing diabetic vision loss, which results from long-term diabetes and causes blood vessel fluid leakage of the retina. Common indicators of DR include blood vessels, exudate, hemorrhages, microaneurysms, and texture. To address this issue, this study proposes a novel CNN model for diabetic retinopathy detection. The proposed approach is an end-to-end mechanism that utilizes Inceptionv3 and Resnet50 for feature extraction of diabetic fundus images. The features extracted from both models are concatenated and input into the proposed InceptionV3 Resnet50 convolutional neural network (IR-CNN) model for retinopathy classification. To improve the performance of the proposed model, several experiments, including image enhancement and data augmentation methods, are conducted. The use of DL models, such as Resnet50 and Inceptionv3, for feature extraction enables the model to capture the complex features underlying diabetic retinopathy, leading to more accurate and reliable classification results. The proposed approach outperformed the existing models for DR detection.

I. INTRODUCTION

Diabetic retinopathy is a retinal vascular disorder appears in the diabetic patients. The count of DR patients is expected to be doubled among Americans from 7.7 million to 14.6 million between 2010 and 2050. The Hispanic Americans are expected to be affected severely and rapidly from 1.2 million to 5.3 million. The duration of diabetes is a key factor in the arrival of retinopathy, with the increase

in diabetes duration will increase the risk of the DR development. It is also noticed that patients with diabetes usually unaware of the possibility of DR, which leads towards the delayed diagnosis and treatment [1]. Manual detection of DR is time-consuming and requires trained clinical experts to analyze digital color fundus images. However, the delayed outcomes can result in a lack of follow-up and misinformation for patients [2].

Diabetic retinopathy has been manually tested by ophthalmologists until now. Manual diagnosis of DR is time-consuming, and therefore, computer-aided diagnosis is gaining attention. Non-proliferative diabetic retinopathy (NPDR) causes retinal swelling and leakage of tiny blood vessels, leading to macular edema and vision loss. Other types of NPDR include blood vessel closure and macular ischemia, as well as the formation of exudates that can affect human vision [3]. Proliferative diabetic retinopathy (PDR) is the most severe stage of the disease, in which new blood vessels start developing in the retina through neovascularization. These new vessels can bleed in the vitreous, causing dark floaters, and if bleeding is extensive, it can result in blurred vision. Scar tissue formation is common in PDR and can cause macular problems or contribute to independent retinal tissue. PDR is a severe condition that can affect both central and peripheral vision.

The existing models are unable to detect the disease at early stages and complicated due to high computational cost with low performance. To address these issues, various techniques have been proposed for automatic detection of DR from fundus images, including DL-based approaches. In this study, we propose a novel DL model for DR detection, utilizing InceptionV3 and Resnet50 for feature extraction of fundus images. The extracted features are then concatenated and fed to the IR-CNN for classification of DR. Additionally, we conduct experiments with image enhancement and data augmentation methods to improve the performance of the proposed model. The proposed DL model efficiently diagnoses the diabetic retinopathy at early stage and perform significantly better than existing techniques. The main objectives of this approach are as follow:

- To develop a novel DL model for early classification of DR using color fundus images.
- To focus on the most critical aspects of the disease to exclude the irrelevant factors to ensure the high recognition accuracy.

II. RELATED WORK

In recent years, there has been a significant growth in the field of computer-aided diagnosis (CAD) in the medical industry. This emerging technology utilizes computer algorithms to assist medical professionals in the diagnosis of medical images. The CAD architecture is designed to address classification difficulties and has become increasingly necessary in the medical field [6]. Detection of DR is one of the primary goals of CAD by differentiating between infected and normal images, and analyzing various parameters such

as microaneurysms (MAs), veins, texture, hemorrhages, node points, and exudate areas [4]. Machine learning (ML) based classification techniques are commonly used to classify the presence or absence of DR [5]. DR is normally categorized into two stages based on the number and severity of symptoms present [7].

Several ML techniques has been developed to detect the DR. Keerthi et al. [8] proposed a novel technique for early detection of diabetic retinopathy symptoms using clutter rejection. In the feature extraction stage, the authors used an anisotropic Gaussian filter in conjunction with scaled difference-of-Gaussian, and inverse Gaussian filter. The final decision about the presence of DR was decided using support vector machine (SVM) classifier. The proposed approach reported 90% specificity and 79% sensitivity. Istvan and Hajdu [9] utilized the Radon transform combined with principal component analysis for prominent features extraction, with SVM classifier to recognize the DR. The proposed model reported the area under curve (AUC) of 0.96. Balint and Hajdu [10] developed an ensemble approach for the detection of microaneurysms (MAs). In feature extraction phase, the following feature extraction techniques: 2D Gaussian transform, grayscale diameter closing, circular Hough transformation, and top-hat transformations were used. An automated approach for DR detection based on the classification and recognition of variation in time series data is given in [11]. The research presented in [12] involves a series of processes that include equipment and patient medical examination, dust particle visualization, training samples, segmentation error, and alignment error of the retinal optic disk, fovea, and vasculature. In this approach, pathological findings are automatically obtained, and ML technique ensures the robustness of proposed approach. The technique presented in [13], demonstrated for branching and geometric models without the user participation. It is worth noting that such automated approaches have the potential to provide accurate and reliable diagnosis of DR, while also saving time and reducing the burden on healthcare professionals.

In the realm of DR diagnosis, DL models have shown promising results. One such model is the CNN model proposed by [14], which utilized Kaggle for training and DiaretDB1 for analysis. The binary data is rated as normal or infected. The proposed DL model is evaluated using binary classification, yielding a sensitivity of 93.6% and a specificity of 97.6% for DiaretDB1. Another research work adopted VGG architecture and the residual neural networks architecture, to identify the DR from color images. In prevalence and systemic risk factor associations of reproducible diabetic retinopathy diagnosis, the ML model and human classifications obtained similar results. The model identified longer durations of diabetes and higher levels of glycated hemoglobin, as well as elevated systemic blood pressure, as risk factors for reproducible diabetic retinopathy. The hyperparameter tuning Inception-v4 (HPTI-v4) model proposed in [15] is another effective diagnostic model of DR. HPTI-v4 provides a segmentation protocol based on
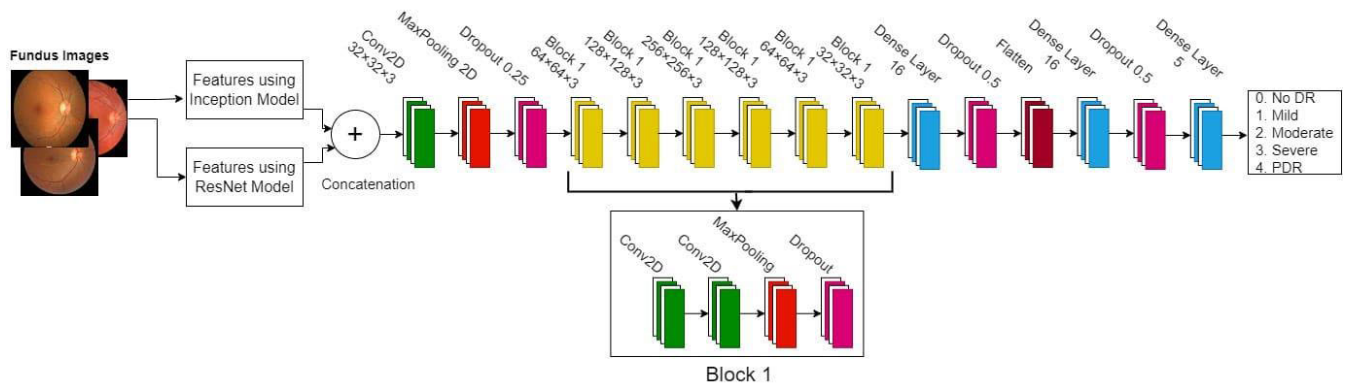
FIGURE 1. The proposed IR-CNN model for diabetic retinopathy detection.
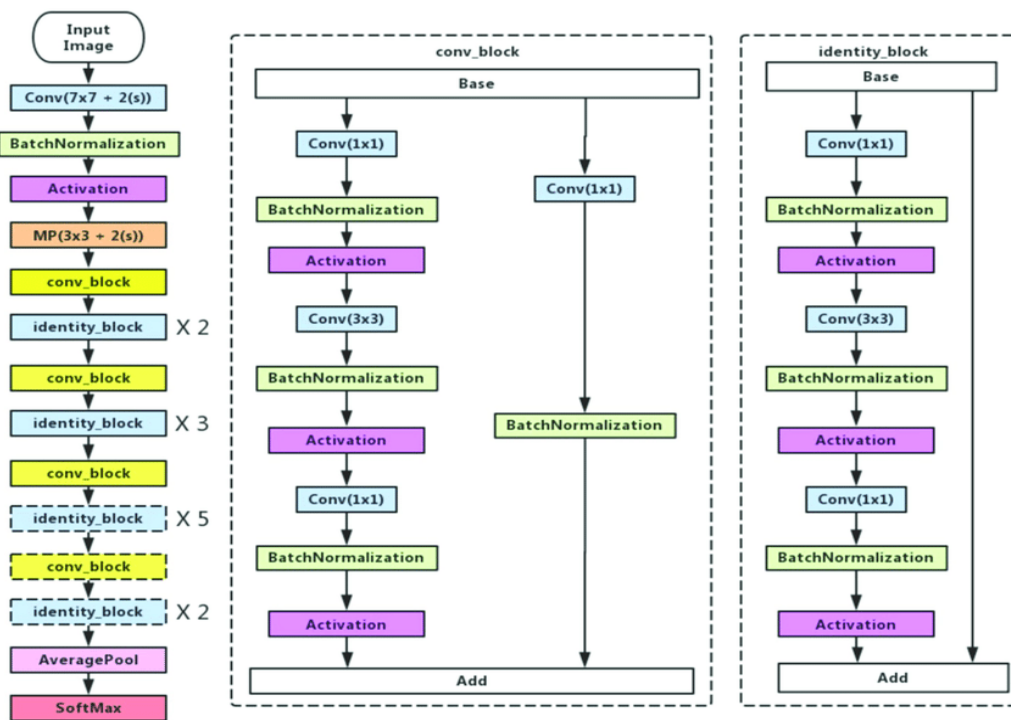

FIGURE 2. ResNet50 model used in this study.

histograms and InceptionV4 function extraction processes. The Bayesian optimization technique is used to change hyper-parameters at the initial stage of Inceptionv4. Finally, the multi-layer perceptron is used for classification processes. Test results show that the presented model provides excellent outcomes with 99.49%, 98.83%, 99.68%, and 100% accuracy. However, it is worth noting that this model can only classify two basic DR classes, namely normal and NPDR. Furthermore, almost 70% of the investigations classified fundus images using binary classifiers such as DR or non-DR, whereas only 27% classified the input to one or more classes.

In the field of DR, the use of CNN models has been explored to achieve more automated detection. In [16], a CNN model was introduced, which took data from two networks as input and sent lesions to a worldwide grading network. The evaluation of the model included class weight and kappa values, but the level of PDR was not considered. Another study [17] used a Kaggle dataset and trained three neural network models, including feedforward neural network, deep neural network, and CNN. The deep neural network achieved the highest training precision of 89.6%. PDR is a serious vision-threatening disease, and automated approaches for detecting new vasculature in retinal images are of great interest. In [18], a DL technique was developed for vessel segmentation that uses a conventional line operator and a unique modified line operator to segment vessels in two ways. The latter is designed to reduce erroneous reactions to edges that are not vessels. Two binary vessel maps are created, and a dual categorization method is used to independently analyze each map's critical information. Two feature sets

are generated by measuring local morphological characteristics from each binary vessel map. Results from a dataset of 60 images show a per-patch sensitivity and specificity of 0.862 and 0.944, respectively, and 1.00 and 0.90, for the first binary vessel map, and second binary vessel map respectively.

This study proposed a novel approach for the detection of DR through the use of CNNs on fundus images. Specifically, the proposed method involved conducting experiments with two distinct CNN architectures, namely Inceptionv3 and Resnet50, followed by feature extraction using these networks. The resulting features were concatenated and used as input to the proposed CNN model for classification. To further enhance the quality of the fundus images, pre-processing techniques such as Histogram Equalization and Intensity normalization were applied. The experimental results demonstrate that the proposed approach outperformed existing methods for DR detection.

III. METHODOLOGY

Deep learning is a powerful ML technique that has been utilized to solve various medical imaging tasks such as object detection, segmentation, and classification. Unlike traditional CNNs that rely on feature extraction methods, DL techniques can directly learn from the input. As a result, DL techniques have been used in various fields including bioinformatics [25], finance [26], drug discovery [27], medical imaging [28], and education systems [29]. Among the different DL techniques, CNN is the most commonly used approach to solve image-based medical problems.

The proposed method is an end-to-end mechanism for DR classification using a hybrid approach with ResNet50 and Inceptionv3. The Inceptionv3 and Resnet50 models are used to extract features from the input images. These features are passed to the Convolutional Neural Network for classification of DR.

A. ResNet50 MODEL FOR FEATURE EXTRACTION

The ResNet50 [30] model was utilized for feature extraction from the DR images. The ResNet50 model introduced a novel structure named residual block, which is a feedforward model with a connection that allows for the addition of new inputs and the production of new outputs. This approach increases the performance of the model without significantly increasing its complexity. Resnet50 yielded the highest accuracy among the DL models considered, and thus, it was selected for DR detection.

B. Inceptionv3 MODEL FOR FEATURE EXTRACTION

In the domain of medical imaging, the InceptionV3 model, which is the most prevalent adaptation of the GoogLeNet architecture [31]. It is extensively used for classification purposes. The architecture of Inceptionv3 model is illustrated in Figure 3. The InceptionV3 model is famous for fusing filters of distinct sizes to form a novel filter, resulting in a reduction of trainable parameters and a corresponding decline in computational cost.

C. EVALUATION METRICS

The objective of evaluation metrics is to assess the efficacy of ML models. Following is the brief description of performance evaluation metrics used in this research.

1) ACCURACY

The overall effectiveness of proposed model in classifying the various DR classes is evaluated using the accuracy metrics. The proportion of correctly classified and miss classified samples, divided by cumulative sum of samples, estimates the correctness of the IR-CNN model. Mathematically it is presented as:

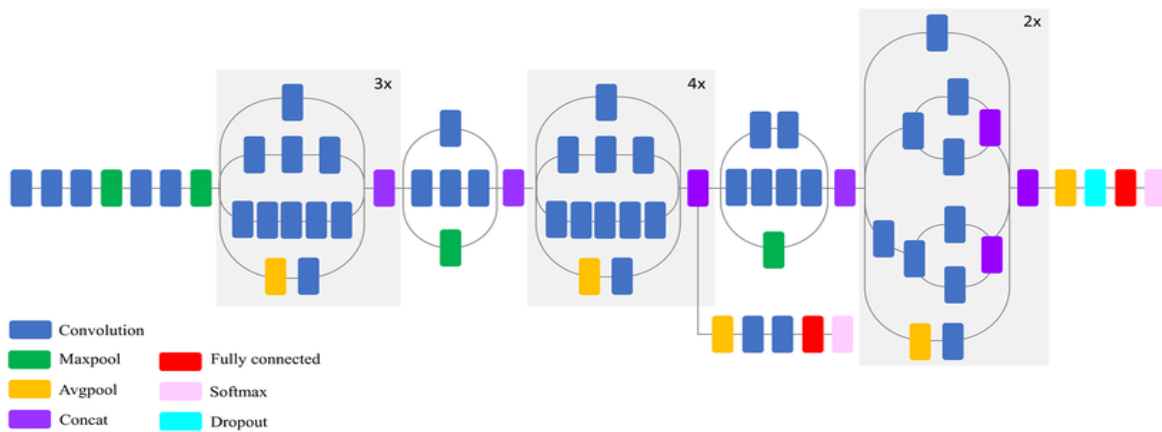$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

FIGURE 3. The Inceptionv3 model used in this study for diabetic retinopathy detection.

2) PRECISION

Precision is used to assess the ability of a ML model to precisely predict positive cases. It represents the relationship of true positive estimates to the sum of true positive false positive estimates. Mathematically, precision can be expressed as follows:

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

3) SENSITIVITY

This is the sum of factual positive instances that are classified accurately. The arithmetic formula of sensitivity is given as:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{3}$$

The ability of a model to appropriately recognize persons with a disease is measured by its sensitivity in medical diagnosis.

4) SPECIFICITY

Specificity refers to accurately classifying the true negative cases, which is computed as the fraction of true negatives to the sum of false positive and true negatives samples. It is also referred to as selectivity or true negative rate. Mathematically, specificity can be expressed as:

$$Specificity = \frac{TN}{(TN + FP)} \tag{4}$$

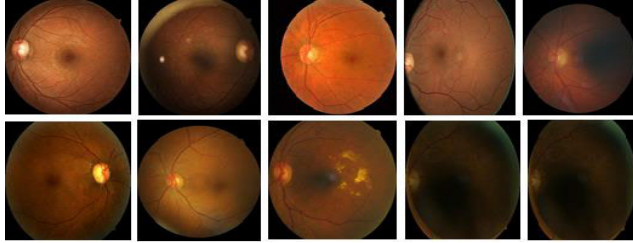In the world of medicine, specificity denotes to a model's ability to accurately classify samples with negative DR.

FIGURE 4. Fundus images dataset of diabetic retinopathy.

IV. EXPERIMENTS AND RESULTS
A. DATASET DEFINITION

In ML based medical diagnostic techniques, the quality of data is crucial for establishing the validity and generalization capabilities of models. The ML paradigm centers on constructing models that can learn from data. To this end, numerous publicly available datasets are accessible for the detection of DR and retinal vessels. These datasets serve as standard resources for training, validating, and testing ML systems, as well as for comparing the performance of different systems. Retinal imaging modalities include fundus color images and optical coherence tomography (OCT). Fundus images are two-dimensional images acquired by reflecting light, while OCT images are three-dimensional images obtained with

low-coherence light that can provide structural and thickness information about the retina [32]. OCT retinal scans have been obtainable for a few years and are widely used for various purposes. In addition, a vast array of publicly available fundus images accessible for research and development in the field of ML.

Individuals diagnosed with severe NPDR face a 17% probability of developing high-risk PDR within a year, and a 40% chance within three years, marking the disease's most advanced stage. PDR is characterized by the emergence of new, fragile, and anomalous blood vessels on the retina or optic nerve. The rupture of these blood vessels can lead to impaired vision. The presence of pre-retinal or vitreous hemorrhages, as well as noticeable neovascularization, is commonly detected during clinical examinations.

This research utilizes an open-source dataset that is publicly available [33]. The dataset comprises high-resolution retinal images of both left and right eyes for each patient, amounting to a total of 44,119 images. The dataset encompasses five distinct classes of DR, as illustrated in Table 1.

TABLE 1. Distribution of images according to the presence of DR.

Category	No. of Images	Labels
No DR	25,806	0
Mild	2,440	1
Moderate	5,291	2
Severe	5,291	3
PDR	5,291	4

B. IMAGE ENHANCEMENT METHODS

The pre-processing step plays a pivotal role in enhancing the quality of retinal images by eliminating noise. The proposed approach employs a set of pre-processing techniques aimed at optimizing image quality, which are briefly outlined below.

1) HISTOGRAM EQUALIZATION

Histogram Equalization is a method utilized to redistribute the intensity values of an input image, resulting in a uniform distribution of intensity values in the output image. The formula for histogram equalization is presented below:

$$Specificity = \frac{Pixels\ with\ intensity\ N}{Total\ Pixels} \tag{5}$$

The histogram equalized image is defined by:

$$g_{i,j} = floor(L - 1) \sum_{n=0}^{f_{i,j}} P_n \tag{6}$$

where floor () rounds down to the closest number. Figure 5 shows the original images and images after applying the histogram equalization.

2) INTENSITY NORMALIZATION

Intensity normalization causes the image histogram in the region of interest to be extended across the whole accessible
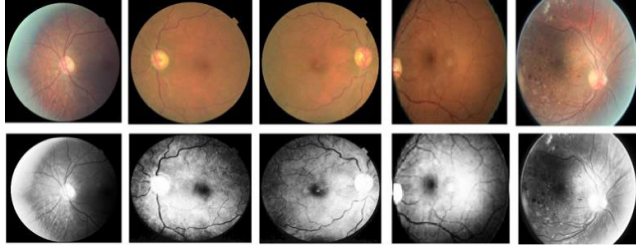
FIGURE 5. The fundus images dataset before and after applying histogram equalization.

intensity array. This approach [19] reported two different structural effects, bright and dark, that enhance the structural contrast, and the effect of the insular mean intensity is minimized. The quality of the features obtained is frequently improved by both effects. Both discriminative and generative [20] tasks require normalization in neural network training. It uses a shared mean and variance to make the input features an independent and identical distribution.

These characteristic speeds up neural network training convergence and makes deep network training efficient. Batch normalization [21], instance normalization [22], layer normalization [23], and group normalization [24] are all practical normalization layers in deep learning-based classifiers. After normalizing the given feature maps, they are often affine transformed, which is learned from other features or circumstances. These conditional normalization methods can help the generator generate more convincing label@lpzsprelevant content.

The intensities in each volume are often standardized by using the following equation.

$$X_{norm} = (I - Min) \frac{Max' - Min'}{Max - Min} + Min' \qquad (7)$$

$$X_{norm} = Max' - Min' \frac{1}{1 + exp\left(\frac{x - \beta}{\sigma}\right)} + Min' \qquad (8)$$

The image intensities are normalized for the dataset used in this study. Supports in decreasing the computational cost and training time of the DL models. The images after applying intensity normalization are given below in Figure 6.
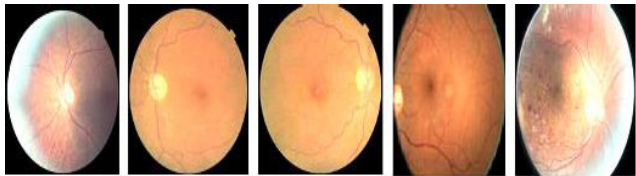

FIGURE 6. The fundus images dataset after applying intensity normalization.

C. MODEL TRAINING

The proposed model for classification of fundus images is explained in this section, which is based on convolutional

neural network architectures. The proposed framework combines two well-known network architectures to perform classification tasks efficiently. If only a single architecture had been used instead of combining multiple architectures, the probability of missing some useful features would have been high, which could have decreased the performance.

After performing normalization, cropping, and resampling of the retinal images, the subsequent stage entails training the model to autonomously extract the multiclass DR features. Due to the high dimensionality of the data, individual samples were processed rather than in batches. The training dataset was segregated into an 80-20 train-test split, and all three network models were trained over 100 epochs, employing a learning rate of 0.001. The training process utilized the Adam optimization algorithm [36], which is an adaptive first-order gradient-based optimization method. The mini-batch size utilized during the training phase was 32 image crops. The Kaggle notebook platform is used to train the proposed model with 16GB RAM and two GPUs. To prevent overfitting, the training process incorporated an early stopping strategy. Specifically, if there was no improvement in validation data after ten epochs, the training process was terminated. The learning rate was also reduced by a factor of 0.4 when the validation loss displayed no improvement for five epochs. Unless otherwise specified, the default loss function employed during the training phase was cross-entropy.

D. RESULTS BEFORE DATA AUGMENTATION

This section provides a detailed analysis and discussion of the performance of the proposed models. To quantify and assess the enhancements made to the final model, several experiments were conducted. The OCT fundus images dataset was used, and the results were obtained after selecting the best model configuration. The pre-processing step, which is an essential initial stage in every data-driven study, involved resizing the images to 256 × 256 pixels because of different image sizes of the dataset to feed into DL models.

1) PROPOSE IR-CNN MODEL
The proposed model is a hybrid model which uses features extracted from the above models described. The features from the Inceptionv3 and ResNet50 are concatenated and given to the CNN model for final prediction. This model is evaluated using OCT fundus images dataset. Table 2 demonstrates how the model's results were compared. In comparison to the model without augmentation, the findings of the augmentation model were promising.

TABLE 2. Results of the proposed IR-CNN architecture on the OCT fundus images dataset.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score
IR-CNN	92.66	96.15	97.13	97.63	94.67

The result of all the five classes using IR-CNN is discussed in the Table 3, which shows the highest accuracy is achieved in class 0.

TABLE 3. Result of IR-CNN model on all the classes of DR.

Model	Class	Label	Accuracy
IR-CNN	No DR	0	94.07
	Mild	1	92.90
	Moderate	2	92.36
	Severe	3	92.10
	PDR	4	91.90

The results presented in Table 3 demonstrates that the proposed model IR-CNN outperform the individual model Resnet50 and InceptionV3. The proposed model achieved the classification accuracy on all classes No-DR, Mild, Moderate, Severe and PDR as 94.07%, 92.90%, 92.36%, 92.10% and 91.90% respectively, which shows the highest accuracy is achieved in class 0.

2) ResNet50 MODEL

After that ResNet-50 is used in this study, which has already been trained on the regular ImageNet database [34]. Residual Network-50 is a deep convolutional neural network that achieves remarkable results in ImageNet database categorization [35]. ResNet-50 is made from a variety of convolutional filter sizes to reduce training time and address the degradation problem caused by deep structures. Table 4 illustrate the outcomes of this model. It is found that the Resnet50 achieves accuracy, sensitivity. Specificity precision and F1 score was 84.15%, 92.58%, 89.29%, 90.47% and 93.47% respectively.

TABLE 4. Results of the ResNet50 architecture on the OCT fundus images dataset.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score
ResNet50	84.15	92.58	89.29	90.47	93.67

The result of all the five classes using IR-CNN is discussed in the Table 5, which shows the highest accuracy is achieved in class 0.

TABLE 5. Result of ResNet50 model on all the classes of DR.

Model	Class	Label	Accuracy
ResNet50	No DR	0	87.06
	Mild	1	82.53
	Moderate	2	82
	Severe	3	82.25
	PDR	4	86.16

3) Inceptionv3 MODEL

The Inceptionv3 model is trained on the fundus images. The results of the model are presented in Table 6. The Inception-v3 architecture is intended for image classification and recognition. Inceptionv3 provides accuracy, sensitivity, specificity precision and F1 score as 82.97%, 94.71%, 96.12%, 94.26% and 93.79% respectively.

TABLE 6. Results of the InceptionV3 architecture on the OCT fundus images dataset.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score
InceptionV3	82.97	94.71	96.12	94.26	93.79

The result of all the five classes using InceptionV3 is discussed in the Table 7, which shows the highest accuracy is achieved in class 0.

TABLE 7. Result of InceptionV3 model on all the classes of DR.

Model	Class	Label	Accuracy
InceptionV3	No DR	0	85.31
	Mild	1	82.43
	Moderate	2	81.80
	Severe	3	82.90
	PDR	4	82.04

The results presented in Table 7 demonstrates that inceptionV3 model classifies all classes No-DR, Mild, Moderate, Severe and PDR as 85.31%, 82.43, 81.80, 82.90 and 82.04 respectively which shows the highest accuracy is achieved in class 0.

E. RESULTS WITH DATA AUGMENTATION

To improve the model accuracy data augmentation methods such as scaling and rotation, are applied to all three models, which gradually increases the accuracy for all the three models. Table 8 demonstrates the results of all models with augmentation. From the results obtained one can observe the proposed model gives the highest accuracy with accuracy, sensitivity, specificity, precision and F1 score as 96.85%, 99.28%, 98.92%, 96.46% and 98.65% respectively.

TABLE 8. Results of the InceptionV3, ResNet50, and the proposed IR-CNN model architecture on the OCT fundus images dataset.

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score
InceptionV3	87.18	95.43	93.71	94.12	90.39
ResNet50	90.65	97.75	96.48	97.25	94.28
Proposed IR-CNN Model	96.85	99.28	98.92	96.46	98.65

Based on the obtained results, it is evident that data augmentation has conferred benefits across all the evaluation

criteria and has substantially enhanced the classification abilities of the model. A comprehensive breakdown of the results of all the DR classes with augmentation, is presented in Table 9.

TABLE 9. Result of InceptionV3 model on all the classes of DR.

Model	Class	Label	Accuracy
InceptionV3	No DR	0	90.20
	Mild	1	88.53
	Moderate	2	85.56
	Severe	3	87.41
	PDR	4	83.82
ResNet50	No DR	0	92.32
	Mild	1	89.72
	Moderate	2	90.41
	Severe	3	89.83
	PDR	4	90.92
Proposed IR-CNN Model	No DR	0	97.33
	Mild	1	97.15
	Moderate	2	97.32
	Severe	3	96.07
	PDR	4	96.40

The analysis of Table 9 leads to the conclusion that data augmentation has resulted in a notable improvement in the performance of all three deep learning models under consideration. Nonetheless, the hybrid model proposed in this study has demonstrated particularly promising results when compared to the other deep learning models.

F. RESULTS WITH 5-FOLD CROSS VALIDATION

In order to demonstrate the generalization of model. K-fold cross-validation is employed as a means of selecting the most effective model DR detection. To execute the cross-validation process, the dataset is partitioned into 5 folds, with four of these folds serving as training sets and the remaining fold used for testing. This process is repeated five times while altering the test set each time. Table 10 displays the outcomes of the DR detection with cross-validation.

TABLE 10. Accuracy, sensitivity, specificity, precision, and F1-score obtained for each fold of the proposed model.

Fold	Accuracy	Sensitivity	Specificity	Precision	F1 Score
1	89.75	92.88	91.25	92.65	90.13
2	94.78	96.59	95.05	93.48	94.59
3	92.35	94.26	93.19	94.79	93.19
4	96.85	99.28	98.92	96.46	98.65
5	90.24	93.06	92.42	94.58	91.25

Table 10 demonstrates that the proposed model achieves 92.79%, 94.39%, 95.21%, 93.56%, 94.16%, average

accuracy, precision, sensitivity, f1 score, and specificity respectively. The most outstanding accuracy score of 96.85%, along with sensitivity, specificity, precision, and F1-score values of 99.28%, 98.92%, 96.46%, and 98.65%, respectively, was obtained from the 4th iteration of the cross-validation process. The least accurate outcome of the cross-validation procedure was obtained in the 5th iteration, with an accuracy score of 90.24% and corresponding sensitivity, specificity, precision, and F1-score values of 93.06%, 92.42%, 94.58%, and 91.25%, respectively.

G. COMPARISON OF THE PROPOSED METHOD WITH OTHER DL MODELS

In recent years, there have been significant improvements in image classification accuracy due to advancements in DL technologies and high-performance computers. Currently, efforts are being made to create AI based techniques capable of performing ocular tasks. The Inceptionv3 model has been evaluated on OCT dataset, obtained accuracy, sensitivity, specificity, precision, and f1 score values of 87.18%, 95.43%, 93.71%, 94.12%, and 90.39%, respectively. ResNet, is a classic neural network that serves as a backbone for many computers vision tasks. We evaluated ResNet50 model on OCT fundus images dataset, the ResNet50 model yielded accuracy, sensitivity, specificity, precision, and f1 score values of 90.65%, 97.75%, 96.48%, 97.25%, and 94.28%, respectively. By combining the features of Inceptionv3 and ResNet50, the proposed model learns deep features and achieves accuracy, sensitivity, specificity, precision, and f1 score values of 96.85%, 99.28%, 98.92%, 96.46%, and 98.65%, respectively.

TABLE 11. Comparison of the proposed models with other DL models.

	Method	Accuracy	Sensitivity	Specificity	Precision	F1 Score
[37]	SVM and NN	0.88	71.9	81.5	-	-
[38]	CNN	95	-	-	-	-
[39]	WP-CNN	94.23	95	90	-	0.90
[4]	SVM	91.2	90.5	91.6	-	-
[40]	REMURS-3D	80.7	-	-	-	-
[41]	Ensemble microaneurysm detector	0.87	51	96	-	-
[42]	k-NN	0.87	66	93	-	-
IR-CNN	InceptionV3, ResNet50 and CNN Model	96.85	99.28	98.92	96.46	98.65

1) DISCUSSION

In [37], fundus image analysis was conducted using SVM neural networks. It presents a comparative analysis of the recognition accuracy of fundus images from both Japanese and American populations, based on their severity index. The trained classification model was evaluated for grading accuracy using 200 Japanese fundus images obtained from Keio University Hospital. The sensitivity and specificity of the

model was 81.5% and 71.9%, respectively. On the American validation dataset, these values were 90.8% and 80.0%, respectively. Controlling blood glucose levels and timely treatment of DR can help prevent many complications associated with diabetes mellitus. However, manual diagnosis of DR is a time-consuming and challenging task due to the diversity and complexity of DR. The work presented in [38] is to develop an automated technique for classifying a batch of fundus images. The authors employed advanced CNNs such as AlexNet, VggNet, GoogleNet, and ResNet, in combination with transfer learning and hyper-parameter tweaking. The models were trained on the freely accessible Kaggle platform. The results demonstrate that CNNs and transfer learning outperform DR image classification by 95.68% in terms of classification accuracy. The low screening rates for diabetic retinopathy highlight the need for an automated image evaluation system that can benefit from the development of DL methods.

The work reported in [39] an innovative method based on ensemble learning is presented, which integrates multiple weighted pathways into a convolutional neural network (WP-CNN). Backpropagation is utilized to enhance various route weight coefficients, thereby reducing redundancy and improving convergence. The experiment results reveal that WP-CNN achieves an impressive accuracy of 94.23%, an AUC, sensitivity and F1-score of the model was reported as 0.9823, 0.908, 90.94% respectively. Similarly, the work presented in [4] focuses on individuals aged 50-85 years from South India, with fundus images captured using a high-resolution CARL ZEISS FF 450 plus Visual camera. The ground truth data confirmed the presence of pathological states such as microaneurysms, exudates, and hemorrhages. A CAD system based on an SVM kernel classifier was used to detect the presence or absence of DR. The proposed system achieved the highest classification accuracy using 5-fold cross-validation, with average sensitivity, specificity, and accuracy values of 91.6%, 90.5%, and 91.2%, respectively.

Considering three-dimensional regional statistics, [40] demonstrates a novel approach to extract the macular disease area in the human retinal layer from OCT images. Previous studies relied on the mean and standard deviation of the two-dimensional illness portion identified by clinical practitioners to determine the disease area, which is not capable to accurately capture the disease region in few cases. To test the proposed technique, OCT images of five individuals with retinal disorders were evaluated, and the anomalous region's volume was measured with an average accuracy of 80.7%. In the context of diabetic retinopathy, an experiment was conducted to test a microaneurysm detector, which was developed by a research group using an ensemble-based method that demonstrated promising results in previous trials [41]. The detector achieves the sensitivity, specificity and AUC, 96%, 51% and 87% respectively. The number of detected microaneurysms increases with greater confidence, making this strategy effective depending on the severity of the condition. Tang et al. [42] proposed a method

for detecting retinal hemorrhages based on splat characteristics. Utilized 357 diverse splat features, including color, difference of gaussian filter, Gaussian and Schmid filter, local texture filter, area, orientation, and solidity. Moreover, feature selection was carried out using a wrapper approach. The authors employed a k-nearest neighbor classifier, which achieved an AUC of 0.87 in receiver operating characteristics, resulting in 66 percent specificity and 93 percent sensitivity. Furthermore, the author presents a novel technique for segmenting blood vessels and the optic disc in fundus retinal images, which aid in detection of ocular diseases such as diabetic retinopathy, glaucoma, and hypertension. The retina vascular tree is first extracted using the graph cut approach. The segmentation of the optic disc is achieved through the Markov random field image reconstruction technique, which eliminates vessels, and the compensation factor approach that relies on previous local intensity knowledge of vessels in the optic disc region.

In the early stages of the disease, diabetic retinopathy may not be symptomatic, and various physical tests are necessary to diagnose DR effectively. To prevent visual loss, diabetic retinopathy must be detected early. For this purpose, Convolution Neural Networks is a better option to detect and grade non-proliferative DR from retinal fundus images. The proposed deep learning technique was evaluated on MESSIDOR and IDRiD datasets. The CNN layers were applied after images were pre-processed and resized to 256 × 256 pixels. The maximum accuracy achieved is 90.89% using MESSIDOR images. The proposed model for the classification of diabetic retinopathy is an end-to-end mechanism in which Inceptionv3 and Resnet50 are used for the feature extraction using diabetic fundus images. The system only requires the fundus image of the patient as input, and all the processing in next examinations can be carried out automatically by the system itself. The features of each model are concatenated and given to the proposed model for classification of diabetic retinopathy. Image enhancement methods and data augmentation are also applied to increase the classification accuracy.

V. CONCLUSION

In working-age people, DR is becoming a more prevalent cause of visual loss. In order to get optimal results, patients must undergo extensive systemic care, including glucose control and blood pressure control. The key to prevent diabetic vision loss is through early detection and treatment. Long-term diabetes causes the blood vessel fluid leakage of retina. Blood vessels, exude, hemorrhages, microaneurysms, and texture are commonly used to determine the stage of DR. In this study, a novel CNN network is proposed for the diabetic retinopathy detection. The proposed approach is an end-to-end mechanism in which Inceptionv3 and Resnet50 are used for the feature extraction of diabetic fundus images. The features extracted using both models are concatenated and given to the proposed model IR-CNN for classification of the retinopathy. Different experiments are performed including image enhancement and data augmentation

methods to increase the performance of the proposed model. The proposed model achieves promising results as compared to the existing model.

ACKNOWLEDGMENT

REFERENCES

[1] S. Akara and B. Uyyanonvara, "Automatic exudates detection from diabetic retinopathy retinal image using fuzzy C-means and morphological methods," in *Proc. 3rd IASTED Int. Conf. Adv. Comput. Sci. Technol.*, 2007, pp. 359–364.

[2] V. Deepika, R. Dorairaj, K. Namuduri, and H. Thompson, "Automated detection and classification of vascular abnormalities in diabetic retinopathy," in *Proc. 38th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, 2004, pp. 1625–1629.

[3] H. Li, W. Hsu, M. L. Lee, and T. Y. Wong, "Automatic grading of retinal vessel caliber," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 7, pp. 1352–1355, Jul. 2005.

[4] E. Trucco et al., "Validating retinal fundus image analysis algorithms: Issues and a proposal," *Investigative Opthalmol. Vis. Sci.*, vol. 54, no. 5, p. 3546, May 2013.

[5] G. A. G. Vimala and S. Kajamohideen, "Diagnosis of diabetic retinopathy by extracting blood vessels and exudates using retinal color fundus images," *WSEAS Trans. Biol. Biomed.*, vol. 11, pp. 20–28, Jun. 2014.

[6] M. Bilal, G. Ali, M. W. Iqbal, M. Anwar, M. S. A. Malik, and R. A. Kadir, "Auto-prep: Efficient and automated data preprocessing pipeline," *IEEE Access*, vol. 10, pp. 107764–107784, 2022.

[7] C. P. Wilkinson et al., "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003.

[8] K. Ram, G. D. Joshi, and J. Sivaswamy, "A successive clutter-rejection-based approach for early detection of diabetic retinopathy," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 664–673, Mar. 2011.

[9] I. Lazar and A. Hajdu, "Microaneurysm detection in retinal images using a rotating cross-section based model," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Mar. 2011, pp. 1405–1409.

[10] B. Antal and A. Hajdu, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 6, pp. 1720–1726, Jun. 2012.

[11] J. Amin, M. Sharif, and M. Yasmin, "A review on recent developments for detection of diabetic retinopathy," *Scientifica*, vol. 2016, pp. 1–20, May 2016.

[12] H. Narasimha-Iyer et al., "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1084–1098, Jun. 2006.

[13] Z. Munawar et al., "Predicting the prevalence of lung cancer using feature transformation techniques," *Egyptian Informat. J.*, vol. 23, no. 4, pp. 109–120, Dec. 2022.

[14] Y. Yehui et al., "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland, 2017, pp. 533–540.

[15] R. A. Welikala et al., "Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification," *Comput. Methods Programs Biomed.*, vol. 114, no. 3, pp. 247–261, May 2014.

[16] M. J. Iqbal, M. W. Iqbal, M. Anwar, M. M. Khan, A. J. Nazimi, and M. N. Ahmad, "Brain tumor segmentation in multimodal MRI using U-Net layered structure," *Comput., Mater. Continua*, vol. 74, no. 3, pp. 5267–5281, 2023.

[17] M. Anwar et al., "Green communication for wireless body area networks: Energy aware link efficient routing approach," *Sensors*, vol. 18, no. 10, p. 3237, Sep. 2018.

[18] I. Sergey and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[19] M. Anwar, F. Masud, R. A. Butt, S. M. Idrus, M. N. Ahmad, and M. Y. Bajuri, "Traffic priority-aware medical data dissemination scheme for IoT based WBASN healthcare applications," *Comput., Mater. Continua*, vol. 71, no. 3, pp. 4443–4456, 2022.

[20] S. Naseem, A. Alhudhaif, M. Anwar, K. N. Qureshi, and G. Jeon, "Artificial general intelligence-based rational behavior detection using cognitive correlates for tracking online harms," *Pers. Ubiquitous Comput.*, vol. 27, no. 1, pp. 119–137, Feb. 2023.

[21] W. Yuxin and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[22] Y. Li, C. Huang, L. Ding, Z. Li, Y. Pan, and X. Gao, "Deep learning in bioinformatics: Introduction, application, and perspective in the big data era," *Methods*, vol. 166, pp. 4–21, Aug. 2019.

[23] Y. A. Ahmed et al., "Sentiment analysis using deep learning techniques: A review," *J. Jilin Univ., Eng. Technol. Ed.*, vol. 42, no. 3, pp. 789–808, 2023.

[24] M. Anwar, A. H. Abdullah, K. N. Qureshi, and A. H. Majid, "Wireless body area networks for healthcare applications: An overview," *Telkomnika*, vol. 15, no. 3, pp. 1088–1095, 2017.

[25] L. A. Selvikvåg and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019.

[26] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: A comparison of deep learning frameworks," *Educ. Inf. Technol.*, vol. 25, no. 3, pp. 1951–1963, May 2020.

[27] R. Olaf, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Cham, Switzerland, 2015, pp. 234–241.

[28] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," 2016, *arXiv:1603.08029*.

[29] E. Rezende, G. Ruppert, T. Carvalho, A. Theophilo, F. Ramos, and P. de Geus, "Malicious software classification using VGG16 deep neural network's bottleneck features," in *Information Technology—New Generations*, Cham, Switzerland: Springer, 2018, pp. 51–59.

[30] X. Tio and A. Emmanuel, "Face shape classification using inception v3," 2019, *arXiv:1911.07916*.

[31] *Diabetic Retinopathy Detection*. Accessed: Mar. 1, 2015. [Online]. Available: https://www.kaggle.com/c/diabetic-retinopathy-detection

[32] Y. Yuhai et al., "Modality classification for medical images using multiple deep convolutional neural networks," *J. Comput. Inf. Syst.*, vol. 11, no. 15, pp. 5403–5413, 2015.

[33] Y. Katada, N. Ozawa, K. Masayoshi, Y. Ofuji, K. Tsubota, and T. Kurihara, "Automatic screening for diabetic retinopathy in interracial fundus images using artificial intelligence," *Intell.-Based Med.*, vols. 3–4, Dec. 2020, Art. no. 100024.

[34] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *Comput. Electr. Eng.*, vol. 72, pp. 274–282, Nov. 2018.

[35] Y.-P. Liu, Z. Li, C. Xu, J. Li, and R. Liang, "Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network," *Artif. Intell. Med.*, vol. 99, Aug. 2019, Art. no. 101694.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[37] I. Nakahara et al., "Extraction of disease area from retinal optical coherence tomography images using three dimensional regional statistics," *Proc. Comput. Sci.*, vol. 22, pp. 893–901, Jan. 2013.

[38] B. Antal, I. Lázár, A. Hajdu, Z. Török, A. Csutak, and T. Peto, "Evaluation of the grading performance of an ensemble-based microaneurysm detector," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 5943–5946.

[39] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff, "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 364–375, Feb. 2013.

[40] Q. Chen, X. Xu, L. Zhang, and D. W. K. Wong, "REMURS: Remote medical ultrasound simulation on the cloud," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 2, pp. 173–181, Sep. 2017.

[41] X. Chen et al., "Deep learning for diagnosis and evaluation of diabetic retinopathy: A review," *J. Healthcare Eng.*, vol. 6, 2019, Art. no. 5481723.

[42] H. Tang, Z. Wu, Y. Li, and J. Li, "Automated detection of retinal hemorrhages in fundus images using splat feature classification," *J. Med. Syst.*, vol. 40, no. 4, p. 95, 2016.

● ● ●