



Multi-Label Diabetic Retinopathy Detection Using Transfer Learning Based Convolutional Neural Network

Raghad. H. Abood¹, Ali. H. Hamad²

¹Informatics Institute for Postgraduate Studies, Iraqi Commission for Computers and Informatics, Baghdad, Iraq

²Department of Information and Communication Engineering, University of Baghdad, Baghdad, Iraq

Emails: ms202220726@iips.edu.iq; ahamad@kecbu.uobaghdad.edu.iq

Abstract

Retinopathy is a progressive and common retinal disease that most progressive diabetics suffer from and causes blood vessels in the retina to swell and leak blood and fluid. This condition requires timely diagnosis via medical experts to prevent causing visual loss among patients. To enhance the feasibility of checking many persons, diverse deep-learning schemes have recently been developed for diabetic retinopathy detection. In this paper, retinopathy image detection system based on diverse deep learning schemes (VGG-19, DenseNet-121, and EfficientNet-B6) has been presented. The implemented deep learning schemes with multi-label classification are trained and tested using the Asia Pacific Tele Ophthalmology Society (APTOS-2019) dataset, and the two combined datasets Indian Diabetic Retinopathy Image Dataset (IDRiD) and Messidor-2. The system outcomes of classification are exhibited as sensitivity, precision, F1Score, and accuracy measurements, and the system performance is compared with recently existing related systems. The attained outcomes indicate that the implemented EfficientNetB6 network outperforms peers' schemes and related systems via realizing supreme accuracy using balanced multi-class retinopathy datasets.

Keywords: Retinopathy, Classification System; Deep Learning Schemes; Multi-label classification; VGG-19, DenseNet-121; EfficientNet-B6

1. Introduction

Retinopathy is a disorder of microvascular which has a long-term influence on the eyes of diabetics. The prospect of progressing retinopathy is commensurate with the diabetics' age and duration of infection, in addition to variation in the levels of blood pressure and a rise in the level of blood glucose [1]. Anatomical alterations in the pattern of vascular and changes in the ratio of blood vessels in the vein and artery are the commonly noticed diabetic retinopathy symptoms [2]. Diabetic retinopathy (DR) advances from an absence of abnormalities (No DR) to the presence of mild micro aneurysms, followed by moderate alterations. Subsequently, Severe DR is characterized by the occurrence of multiple hemorrhages, venous beading, or intraregional microvascular abnormalities (IRMA), albeit lacking any signs of proliferative changes. Finally Proliferative DR involves neovascularization or vitreous hemorrhage [3].

The global prevalence of diabetes has been steadily increasing, with the number of individuals affected rising from 285 million in 2010 to a projected 439 million by 2030 [4]. Among those with diabetic retinopathy, approximately one-quarter experience vision-threatening complications, potentially leading to irreversible blindness. Early detection of retinopathy is therefore crucial to preserving patients' vision. Addressing this challenge requires a focus on the fundamental issues associated with retinopathy.

Manual detection are the effort and time consumption that are needed by an ophthalmologist for examining the retinal fundus images of the eyes of diabetics. Figure 1 depicts some samples of fundus retinal images with diverse retinopathy severity levels [4]. Consequently, there is a requirement to avoid these issues and find earlier and more effective detection approaches to decrease blindness risk. Accordingly, diverse diabetic retinopathy detection and

classification systems have made substantial proceed using machine and deep learning approaches that can extract distinct (low and high levels) features for blood vessels in the fundus images [1] [5].

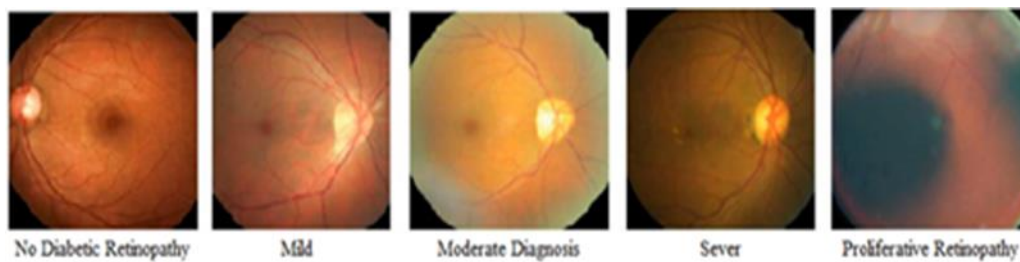


Figure 1. Levels of diabetic retinopathy, graded as severity

When dealing with massive data sizes, machine learning schemes couldn't supply satisfactory responses. Therefore, deep learning schemes are utilized to conquer this issues [6]. Deep learning schemes especially, Convolution Neural Networks (CNNs) have revolutionized lots of application fields dealing with real-world issues [7]. These schemes encompass input, output, and many hidden layers. The omnipresence and wide scope of employment fields of CNNs can be ascribed to various benefits: intertwining extraction of essential features and classification processes into a single body of learning, learning to optimize the extracted features throughout the training phase from the input raw images, and processing more inputs (of various sizes) that have impervious to minor changes with an outstanding computational capability. Over the last few years, deep learning schemes hold their benefits and obstacles, however, various researchers have utilized these schemes to construct automatic retinopathy classification systems [8] The attributes of diabetic retinopathy (DR) fundus images typically encompass a variety of lesion types that furnish compelling evidence for ophthalmologists in the diagnostic process. It is of paramount importance to identify an effective approach that not only precisely classify DR fundus images but also detects the diverse lesions present within them. In order to alleviate the burdensome task of annotation and enhance the efficiency of labeling, this paper has innovatively proposed treating various lesion types as distinct labels for a fundus image, thereby transforming the challenge of lesion detection into that of image classification [9]. Also the process of multi-label classification in diabetic retinopathy enables the concurrent recognition of multiple retinal irregularities or level of the disease, providing a more comprehensive evaluative diagnosis [10]. The principal contributions of this proposed system can be given as follows:

- Exploiting diverse image pre-processing approaches and three deep-learning schemes are utilized to classify diabetic retinopathy images into five diverse classes, then expert supervision isn't needed.
- Exhibiting an approach of SMOTE capable of mitigating issues resulting from dataset imbalance, in addition to an augmentation approach that can discern relevant features from a diverse array of data variations.
- The proposed system accurately classifies retinopathy images into five fundamental classes by using multi label classification approach.
- The schemes' performance has been assessed by utilizing sensitivity, precision, F1Score, and accuracy measurements.

The residual organization of this paper is delineated as follows. The subsequent section depicts the recently presented relevant works, the third section explains the proposed system methodology at length, and the fourth section provides the attained results of the proposed retinopathy classification system with a comparison with the attained results of other closely relevant works. The conclusion and the future works are depicted in the final section.

2. Related Work

In Deep-learning schemes hold the capability to automate and improve screening aspects for diabetic retinopathy, and various studies have investigated their clinical relevance [11]. Based on well-developed deep-learning schemes, various researchers presented distinct multi-classification systems for retinopathy-aided diagnosis. Concerning retinopathy multi-classification systems, the motivation of researchers is to provide the detailed severity (from 0 to 4 grades), by which it is potential to observe the progression of diabetic retinopathy. To diminish the effect of imbalanced datasets across distinct retinopathy grades, some researchers applied diverse augmentation and oversampling techniques for training and testing the deep learning classification schemes [12].

[13] proposed a model for multi-label classification based on deep learning techniques, employing Gradient-weighted Class Activation Mapping, this study provide lesion types as separate labels for a fundus image, thereby transforming the task of lesion detection into that of image classification. The architectural framework of

the deep learning model was constructed by the researcher team, drawing upon the ResNet architecture as a foundational basis. They used Messidor dataset and achieved 93.9% sensitivity and 94.4% specificity of on DR classification. [14] Presented Convolutional Neural Networks, specifically DenseNet, ResNet50, and VGG16, were subjected to training to facilitate the process of multi-label classification, signifying that the algorithms possessed the capability to concurrently discern four distinct retinal lesions. The dataset that used in this work FFA images from Eye Centre at the Second Affiliated Hospital of Zhejiang University School medicine were classified with four distinct categories of lesions associated with diabetic retinopathy, DenseNet achieve the best result with AUC of 0.8703, 0.9435, 0.9647, and 0.9653 for detecting NP, micro aneurysms, leakages, and laser scars. [15] Utilized two deep learning schemes (Xception and InceptionV3) for classifying retinopathy fundus images into five grads. These schemes were trained and tested using the APTOS-2019 dataset. The utilized dataset was firstly pre-processed using Gaussian Blur as well as applying augmentation approaches such as uniform brightening, rotation, and vertical and horizontal flips. The attained results depicted that the Xception scheme outperformed the other implemented scheme, and the achieved accuracy was 93.10%. [16] Presented a deep learning-based system for classifying 5000 retinal fundus images into various severity levels (five classes) for diabetic retinopathy disease. First, these raw images were pre-processed by resizing them to 224×224×3 pixels, then over-sampling them to equalize the count of classes. After that, these pre-processed images were fed to the deep-learning classification schemes (simple CNN, VGG-16, and VGG-19). The performance of VGG-16 outperformed the other classification schemes. The achieved accuracy of VGG-19 was 83%, and the achieved accuracy of VGG-16 was 90%. However, the accuracy of this presented system could be enhanced by utilizing more pre-processing methods and other deep learning schemes.[17] presented a retinopathy classification system using various deep learning schemes. In this system, several image pre-processing techniques were utilized to emphasize the essential features, and various auto-tuned deep-learning schemes (VGG-16, Mobilenet-v2, adapted Resnet50, Inception-v3, and Inception Resnet-v2) were implemented to provide superior results of retinopathy classification. These schemes were trained and tested using the Asia Pacific Tele Ophthalmology Society (APTOS) dataset. Resnet schemes attained superior performances compared with the other implemented schemes. The highest classification accuracy was 94.7% achieved using adapted Resnet50. [18] Leveraged retinal images to train and test two deep-learning schemes (ResNet-34 and DenseNet-121) using augmented APTOS- 2019 dataset. The utilized dataset was firstly pre-processed by adjusting the contrast of retinal images with Gaussian Blur and contrast limited adaptive histogram equalization. In this presented system, the DenseNet-121 classification scheme outperformed the other scheme with an accuracy of 89.57% and some comparative schemes.[19] examined ResNet's potential application in retinal image classification side by side with other deep learning schemes using various datasets (APTOS, Messidor, EyePACS, and IDRiD). In the presented system, the exudates, blood vessels, and spots of cotton wool were underlined using various image pre-processing methods. Compared with the implemented VGGNet-16 and MobileNet-V2 schemes, the ResNet-18 scheme with the Swish function reached higher results, and the attained accuracy, sensitivity, precision, and F1-score using APTOS dataset were 93.51%, 93.42%, 93.77%, and 93.59%, respectively.

The prevailing state-of-the-art methodologies presents several salient limitations that warrant emphasis, as delineated in the related work section The utilization and execution of deep learning methodologies for the categorization of multi-label datasets are still in its early stages. Furthermore, the lack of high-volume datasets significantly undermines the performance efficacy of deep learning models.

Also noticed that utilizing the deep learning schemes without retinal image pre-processing methods could not supply adequate performance in benchmark retinopathy datasets. Some of these relevant systems have stressed pre-processing methods and others stressed utilizing recent effective deep-learning schemes to increase their performance. In the proposed system, various pre-processing methods effective deep learning schemes and multi-label classification were utilized to attained high results. Comparisons between various relevant retinopathy classification systems are depicted in Table 1.

Table 1: Comparison of various retinopathy classification system

Ref	Datasets	Deep learning Schemes	No. of Classes	Merits	Demerits
[13]	Messidor	ResNet	5 Classes	improving the efficiency of labeling work.	They should try another pre-training model to prove the effectiveness of the proposed work.

[14]	FFA images	DenseNet, ResNet50, VGG16	4 Classes	Achieved high accuracy for each level.	Didnt provide any detail about preprocessing the images.
[15]	APTOS-2019 dataset	Xception InceptionV3	5 Classes	Improving the essential highlighted features by utilizing Gaussian Blur and applying data augmentation	Other deep learning schemes are required to enhance the system's performance
[16]	25000 retinal fundus images	VGG-16 VGG-19	5 Classes	The oversampling technique was performed on imbalanced data classes	The highest obtained accuracy was 90% which could be enhanced by utilizing more pre-processing methods and other deep-learning schemes
[17]	APTOS dataset	Inception Resnet-v2 Adapted Resnet50	3 Classes	Gaussian blur and circular crop methods were utilized to extract the essential features	The three-class classification should be expanded to a five-class issue with more sophisticated features
[18]	APTOS-2019 dataset	ResNet-34 DenseNet-121	5 Classes	Adjusting the contrast of images with Gaussian Blur and contrast limited adaptive histogram equalization	Low accuracies were achieved, other deep-learning schemes should be utilized
[19]	APTOS-2019 dataset	ResNet-18	5 Classes	Achieved high accuracy (93.51%)	Still need more accuracy improvement

3. Proposed System

This section presents the utilized datasets, the raw retinal images pre-processing pipeline, and the implementation of effective deep-learning classification schemes. The proposed system, as illustrated in Figure 2, involves several stages; Inputting and pre-processing the retinopathy dataset, splitting pre-processed retinopathy images into test/train groups, implementing diverse deep learning schemes with multi label classification, and finally assessing the system performance using diverse assessment measurements.

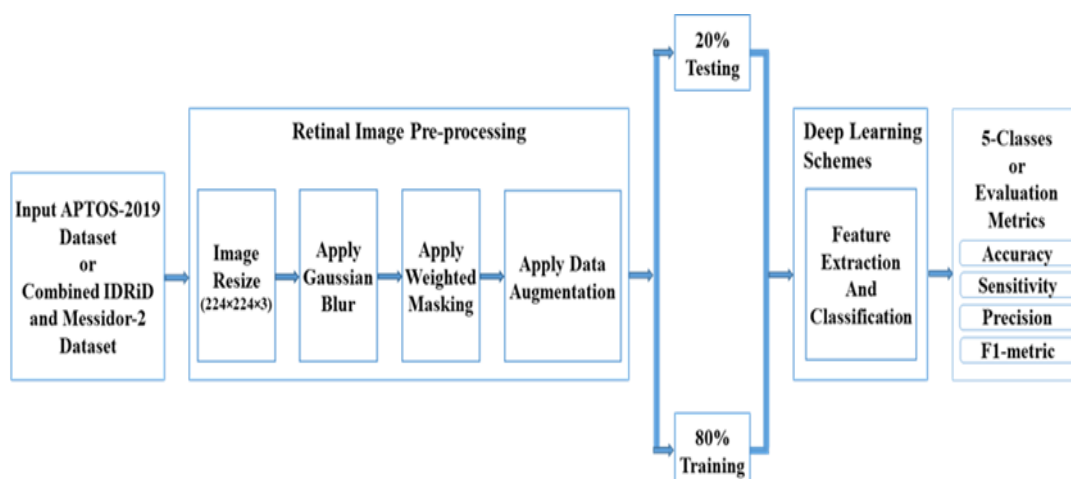


Figure 2. Architecture of retinopathy classification system

3.1 Retinopathy datasets

In the proposed system, two datasets are utilized; the first dataset is the APTOS-2019 multi-class retinopathy dataset, and the second one represents a combination of two multi-class retinopathy datasets IDRiD and Messidor-2 that are combined for accurately classifying more retinal images.

Figure 3 indicates the whole distribution of retinal images in diverse classes for the utilized datasets. These datasets are explained below depending on their sample sizes (large, medium, and small datasets).

- APTOS-2019 dataset [20] represents a largest dataset utilized for diagnosing diabetic retinopathy. It is imbalanced and involves 3662 images collected in a real-world environment. This dataset exhibited variations owing to the settings of cameras and noise in the data and labels.
- IDRiD [21] represents one of the commonly utilized benchmark datasets involving normal and retinopathy images (of size 4288×2848) for the Indian population. This dataset encompasses 516 retinal images of which 413 images are intended for training, and 103 images are considered for testing.
- Messidor dataset [22] involved 1200 retinal fundus images of varying sizes (collected between 2005 and 2006), 800 images were acquired using pupil dilation and 400 were acquired without using pupil dilation. Messidor-2 dataset [23] involved 1748 good-quality images of which 1058 images were taken from the Messidor dataset, in addition to 690 images collected between 2009 and 2010.

Table 2: Number of retinal images in each class

APTOS 2019 dataset				
class	Num Of image	Percentage of image	Num after balance	Percentage after balance
Normal	1805	49.3%	1805	20%
Mild	999	10.1%	1805	20%
Moderate	370	27.3%	1805	20%
Severe	295	5.3%	1805	20%
Proliferative	193	8.1%	1805	20%
Messidor-2 & IDRiD				
Normal	1146	52.1%	1046	20%
Mild	503	13.3%	1046	20%
Moderate	292	22.9%	1046	20%
Severe	159	7.2%	1046	20%
Proliferative	99	4.5%	1046	20%

The number of retinal images in APTOS-2019 are considered sufficient, but the number of retinal images for Messidor-2 and IDRiD individually considered not effective, so in this work we combine the two datasets Messidor-2 and IDRiD. By doing so, dataset has been increased which could enhance the learning performance. Figure 3 shows the retinopathy classes distribution after the dataset augmentation.

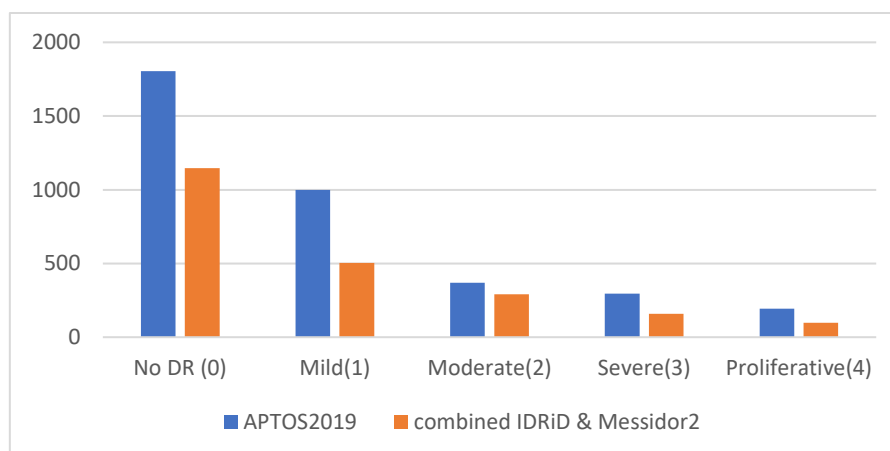


Figure 3. Distribution of the retinopathy classes in each utilized dataset

3.2. Pre-processing Retina images

Owing to the acquisition of a camera, noises caused during retina photographing, and variation of images in size and brightness, the essential attributes for classification like exudates, blood vessels, and the spots of cotton wool could not be distinguished. To handle these issues, several pre-processes can be applied to obtain better-quality images.

Resizing retinal fundus images: Since the raw images were acquired from diverse sources at various periods, these collected images had varying storage memory and sizes. Image resizing pre-process works on adjusting the image size while preserving the quality to the fullest extent [24]. Thus, to obtain superior computation ability, the entire images within the utilized datasets are rescaled to a uniform size (224×224×3 pixels).

Applying Gaussian Blur: To decrease the noise amount in the retinal images, a Gaussian Blur processing technique is applied. In this pre-process, each retinal fundus image of a 2D matrix of color pixels is passed through (convolved with) a 2D-Gaussian filter. Generally, the Gaussian filter can be depicted as follows:

$$2DG(l, h, \delta) = \frac{1}{2\pi\delta^2} e^{-\frac{(l^2+h^2)}{2\delta^2}} \quad (1)$$

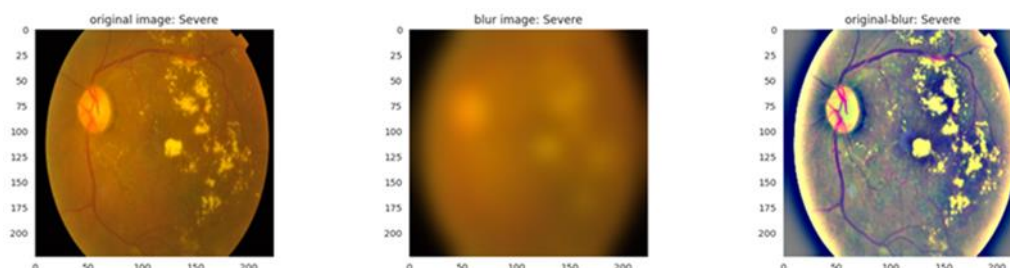
Where l and h denote the position indices, and δ denotes the obscuring influence degree around a pixel (standard deviation). After implementing this pre-process, the areas that included essential information were highlighted [25].

Applying weighted masking: After applying the previous pre-process, weighted masking is applied to sharpen the image edges, where pixel values in regions of interest will be emphasized and regions of no interest will be suppressed [26].

Applying SMOTE: In order to alleviate class imbalance and enhance model performance and generalization, SMOTE is applied during the preprocessing stage of retinal fundus image datasets. A disproportionate amount of photos from various retinopathy classifications can be found in retinal fundus imaging collections. There could be a lot of images showing healthy retinas, but not many showing serious retinopathy. By boosting the proportion of minority class samples, SMOTE aids in dataset balance. SMOTE can be directly applied to image datasets, thus it encompasses creating synthetic images via interpolating between image pairs from the underrepresented category [27]. A model can be trained on this well-balanced dataset to improve its diagnostic performance by learning to recognize and classify each stage of retinopathy more correctly.

Utilize various approaches of augmentation: Image augmentation represents implementing diverse approaches to increase the amount and variety of images accessible for training machine and deep datasets learning models. In other words, image augmentation assists in generating expanded substantial from smaller ones[28], data augmentation tackles the issue of lower accuracy in multi-class classification by generating additional training data through techniques like resizing, rotation, and zooming and adjustments in brightness. These approaches generate further retinal fundus images and the model gains the ability to recognize important features across various data variations, thereby improving its adaptability and enhancing its overall accuracy in classifying retinopathy and hence improve the system's capability for generalizing efficiently and making accurate classifications. After SMOTE and augmentation, the size of the first utilized dataset (APTOS-2019) becomes 10108, and the size of the combined IDRiD and Messidor-2 dataset becomes 7217.

Compared with the original raw images, the pre-processed retinal images that included the essential highlighted features of red lesions, blood vessels, and exudates were improved, as depicted in Figure 4.



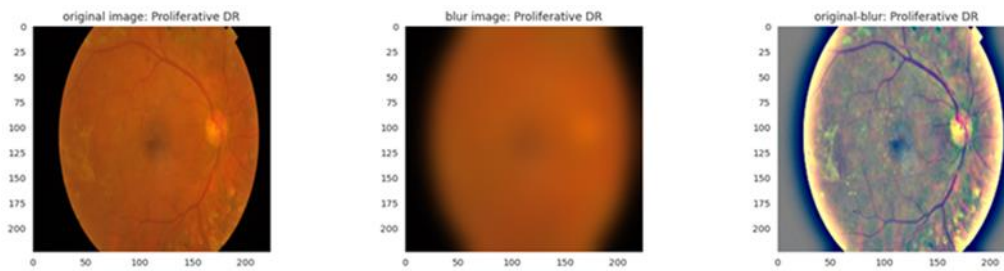


Figure 4. An example of the retinal image pre-processing stage

3.3 Implementing Deep learning schemes

Before the classification stage, the pre-processed image datasets are separated into train and test groups. The proposed system implements Convolution Neural Network and three pre-trained CNNs schemes (VGG-19, DenseNet-121, and EfficientNet-B6), in our work involves first extracting features from the input images. The classification process is then carried out using fully connected layers. The output layer employs sigmoid activation, making these models suitable for multi-label classification and specifically for categorizing retinopathy into five classes. The training process for these schemes is the same. The required configurations are; a learning rate of 0.00005, the count of Epoch and Batch Size are 35 and 12, respectively, and the utilized optimizer is Adam.

In our work we use Convolution Neural Network that provides high performance in terms of image segmentation, improvement and classification, especially in the medical field. The structure of CNN that we proposed consists of several convolutional layers with ReLU activation, followed by batch normalization and max-pooling layers.

Initially, it commences with a pair of Conv2D layers containing 32 filters, advances to another pair with 64 filters, and subsequently to a further pair with 128 filters. The ultimate convolutional layers entail 512 filters. Following this, the model utilizes global average pooling, succeeded by three dense layers with 1024, 512, and 256 units respectively, and culminating in a final dense layer for the purpose of classifying inputs into 5 classes.

Many models deal with complex problems and require large datasets. In order to train the CNN model, you need a large dataset, but collecting a lot of classified data requires time and effort by the experts who collect it, and it is considered an arduous and difficult challenge, especially to obtain a large database.

Therefore, benefiting from pre-trained models regardless of the field they are in is called transfer learning.

- **VGG-19:** VGG [29] represents a deep learning scheme that evolved in 2014. This scheme encompasses diverse versions (VGG11-VGG19) that vary in the count of contained layers. In the proposed retinopathy classification system, VGG-19 of nineteen layers is utilized.[30] VGG19 excels at extracting deep features by using small filters that were effective at extracting fine details, making VGG19 simpler and more effective.
- **DenseNet-121:** DenseNets [31] are named densely connected convolution networks since every layer inside these schemes is connected to the entire other layers. The input concerning every layer represents feature maps of the entire former layers, and concerning every succeeding layer, its feature maps are utilized as input. DenseNets schemes mitigate the issue of disappearing gradients, fostering feature reutilization, strengthening feature propagation, and considerably reducing the count of parameters. In the proposed retinopathy classification system, DenseNet-121 is utilized.
- **EfficientNet-B6:** EfficientNet [32] represents a deep learning scheme evolved by a team of Google researchers in 2019. This scheme encompasses diverse versions (B0-B7) that were evolved depending on the assumption of expanding the resolution, width, and depth of the scheme. MBConvs represent the fundamental building blocks for the EfficientNet that generate shortcuts between the starting and ending of a convolution block. EfficientNet assists in decreasing the scheme size and the total count of processes needed. In the proposed retinopathy classification system, EfficientNet-B6 is utilized. [33]EfficientNet-B6 represents one of the larger versions of EfficientNets, with a larger scale factor compared with former versions (B0-B5). In other words, it has greater resolution, width, and depth, which results in higher computational cost, while the accuracy is higher.

3.4 Multi-label classification

Since our objective involves multi-label classification, we are unable to train the model using the conventional loss function. We are looking for a classification function that can minimize the loss function because every image is linked to the actual label. We utilize training data sets that have been tagged and employ the one-hot approach to encode each y with five labels. We must use the optimization function to optimize the learning parameters after

we have the loss function. We paid special attention to Adam's influence on model performance because different optimizers can have varied effects on parameter training.

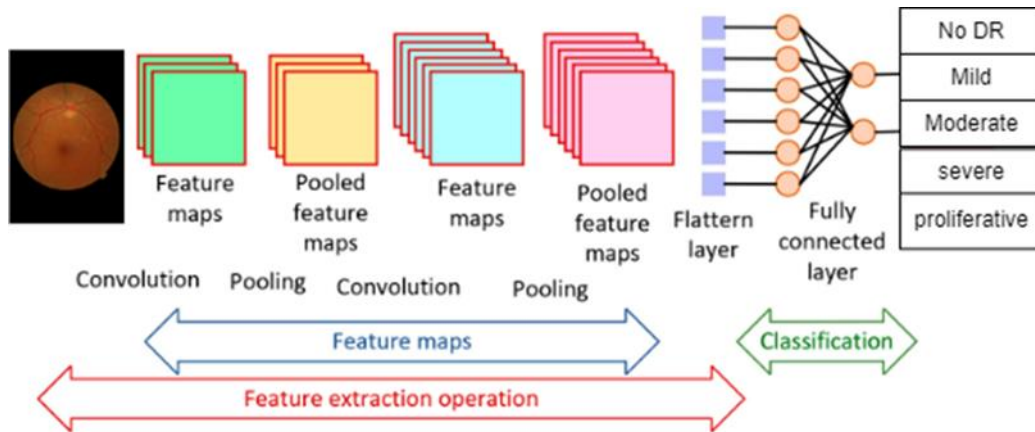


Figure 5. Convolution Neural Network Architecture of retinopathy classification

In Figure 5 the feature maps obtained from CNN or pre-training model are applied to global average pooling layer (GAP). The final feature vector is optimized using the combination of sigmoid activation function and binary cross-entropy loss function for multi-label classification of fundus images. The sigmoid activation function (s) (Eq. (2)) [34] limits the predicted value of every feature s between $[0, 1]$.

$$f(s) = \frac{1}{1 + e^{-s}} \quad (2)$$

Binary cross-entropy loss calculates the difference between the predicted probability and the actual label (0 or 1). The model is subsequently instructed to reduce the binary cross-entropy loss through training. In the case of a specific retinal image x accompanied by true labels y (where y represents a binary vector denoting the existence of different lesion types) and predicted probabilities \hat{y} , the loss pertaining to an individual sample is determined. The formula (3) for binary cross-entropy loss is:

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p(\hat{y}_i)) + (1 - y_i) \log(1 - (p(\hat{y}_i))) \quad (3)$$

The parameter N represents the quantity of samples, where y_i denotes the classification of sample i . The class denoted as positive is assigned the value of 1, while the class identified as negative is assigned 0. And $p(y_i)$ represents the probability that sample i will be predicted as positive. The aggregate loss is calculated across all instances in the data collection to derive the total loss, which the model strives to reduce throughout the training process [35].

4. Experimental Result

In this section, the proposed retinopathy classification system is assessed using the testing sets, and the deep learning schemes' performance was examined and compared with the other relevant works regarding performance assessment measurements like scheme Recall (Rec) precision (Pre), F1-Score ($F1_{score}$), and accuracy (Acc). These measurements are calculated dependent on the confusion matrix by obtaining True negative (Tn) False negative (Fn) True positive (Tp) and False positive (Fp). The formulas of performance assessment measurements are as follows:

$$Acc = \frac{\sum_{i=1}^N TP(C_i)}{\sum_{i=1}^N \sum_{j=1}^N C_{i,j}} \quad (4)$$

$$Pre = \frac{TP(C_i)}{TP(C_i) + FP(C_i)} \quad (5)$$

$$Rec = \frac{TP(C_i)}{FN(C_i) + TP(C_i)} \quad (6)$$

$$F1(Ci) = \frac{2 * Pre(Ci) * Sens(Ci)}{Pre(Ci) + Sens(Ci)} \quad (7)$$

To present an evident understanding of the proposed system performance, Figures 6, 7, and 8 illustrate the ROC curve for balanced APTOS-2019 and combined IDRiD and Messidor-2 datasets. It displays the relationship between the True Positive Rate and the False Positive Rate across a variety of possible decision thresholds. That shows Efficientnet B6 network works better as compared to CNN, VGG 19, and DenseNet 121 architectures. The Area under the Curve (AUC) achieved better with Efficientnet for two input approaches. This indicates that the proposed system is robust and effective.

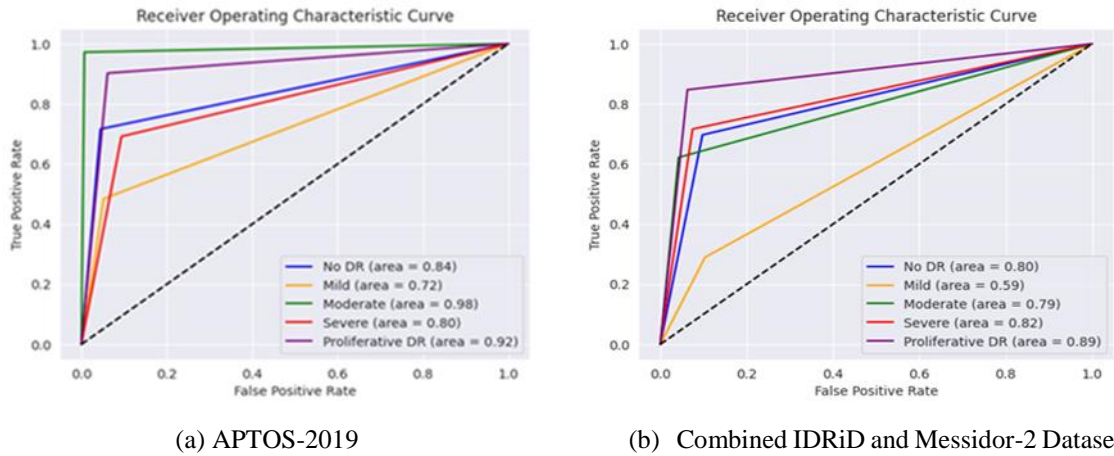


Figure 6. ROC curve for CNN Schemes

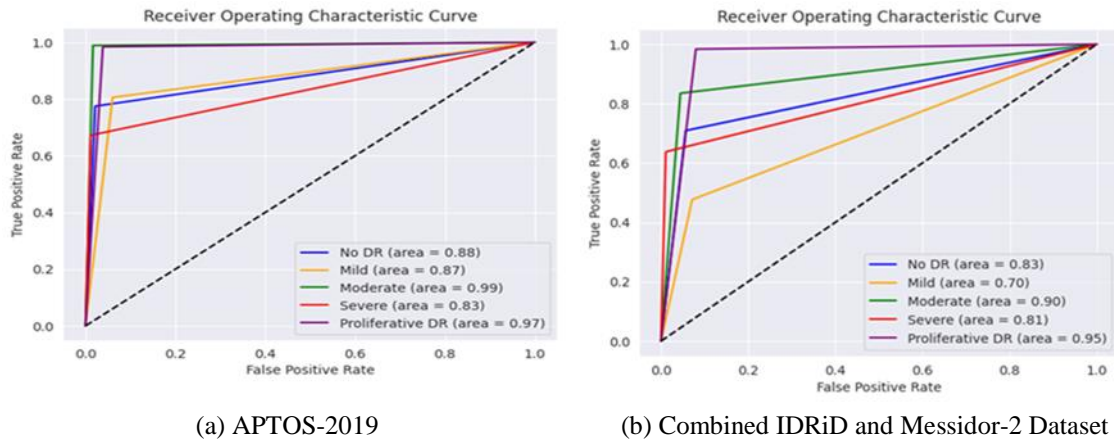


Figure 7. ROC curve for VGG 19 Schemes

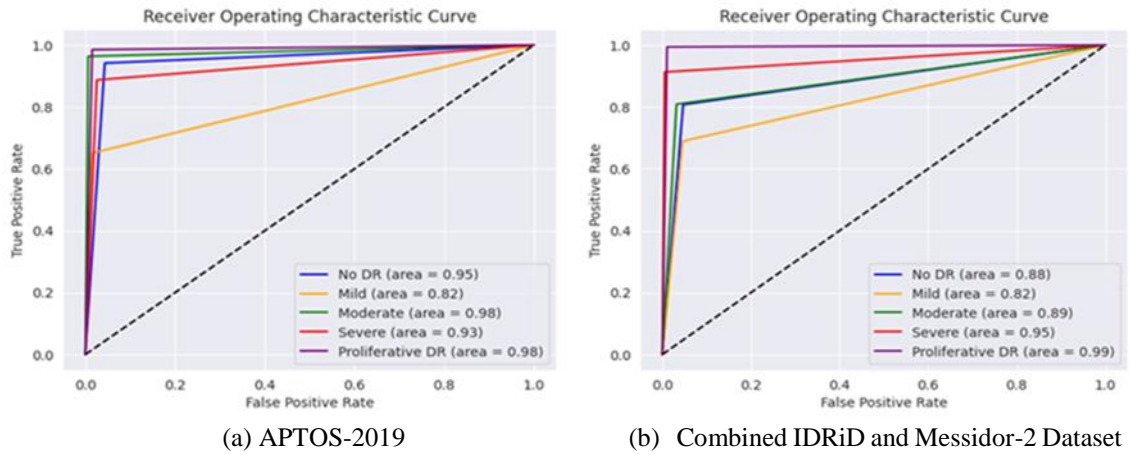


Figure 8. ROC curve for DenseNet 121 Schemes

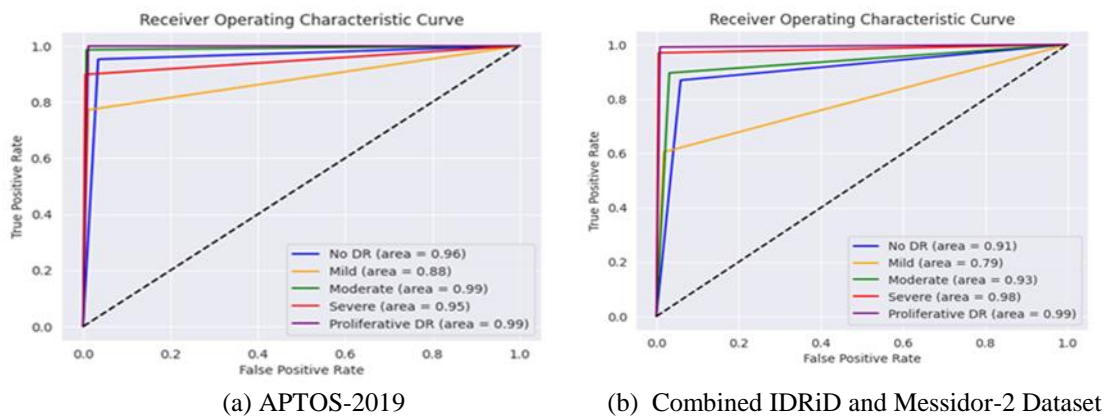


Figure 9. ROC curve for EfficientNet B6 Schemes

Tables 3, 4 and Table 5 depict the classification results of the implemented deep learning schemes assessed on the balanced APTOS-2019 and combined IDRiD and Messidor-2 datasets. The EfficientNet-B6 scheme reports the highest accuracy among the others, and the achieved results of sensitivity, precision, F1Score, and accuracy of this superior scheme were all 94% for the APTOS-2019 dataset, and 90% for the combined IDRiD and Messidor-2 dataset, respectively.

Table 3: Achieved results using balanced APTOS-2019 dataset & combined IDRiD and Messidor-2 dataset

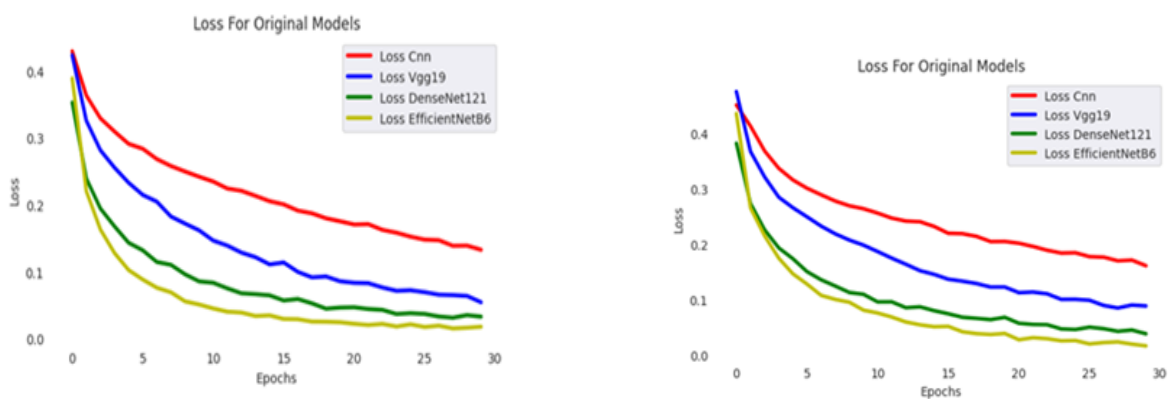
Classes	APTOS 2019				IDRiD and Messidor-2			
	Pre	Recall	F1Sc		Pre	Recall	F1Sco	
0	0.74	0.72	0.73	377	0.56	0.70	0.62	244
1	0.60	0.48	0.54	355	0.32	0.29	0.31	235
2	0.96	0.97	0.97	358	0.71	0.62	0.66	230
3	0.55	0.69	0.61	363	0.61	0.72	0.66	229
4	0.92	0.90	0.91	1074	0.91	0.85	0.88	667
Acc	0.80			2527	0.69			1605
MacroAvg	0.75	0.75	0.75	2527	0.62	0.63	0.63	1605
WeightedAvg	0.80	0.80	0.80	2527	0.70	0.69	0.69	1605

Table 4: Achieved results using balanced APTOS-2019 dataset

Classes	VGG-19			DenseNet-121			EfficientNet-B6			Support
	Pre	Recall	F1 _{Sc}	Pre	Recall	F1 _{Sc}	Pre	Recall	F1 _{Sc}	
0	0.86	0.77	0.82	0.79	0.94	0.86	0.83	0.95	0.89	377
1	0.69	0.81	0.74	0.88	0.65	0.75	0.92	0.77	0.84	355
2	0.91	0.99	0.95	0.96	0.96	0.96	0.96	0.99	0.97	358
3	0.91	0.67	0.77	0.85	0.89	0.87	0.97	0.90	0.93	363
4	0.95	0.98	0.97	0.98	0.99	0.98	0.98	1.00	0.99	1074
Acc	0.88			0.91			0.94			2527
Macro _{Avg}	0.86	0.85	0.85	0.89	0.89	0.88	0.93	0.92	0.92	2527
Weighted _{Avg}	0.89	0.88	0.88	0.92	0.91	0.91	0.95	0.94	0.94	2527

Table 5: Achieved results using balanced combined IDRiD and Messidor-2 dataset

Classes	VGG-19			DenseNet-121			EfficientNet-B6			Support
	Pre	Recall	F1 _{Score}	Pre	Recall	F1 _{Score}	Pre	Recall	F1 _{Score}	
0	0.69	0.71	0.70	0.75	0.81	0.78	0.73	0.87	0.79	244
1	0.53	0.48	0.50	0.72	0.69	0.70	0.86	0.60	0.71	235
2	0.76	0.83	0.80	0.81	0.81	0.81	0.83	0.90	0.86	230
3	0.91	0.64	0.75	0.97	0.91	0.94	0.97	0.97	0.97	229
4	0.90	0.98	0.94	0.99	0.99	0.99	0.99	0.99	0.99	667
Acc	0.80			0.88			0.90			1605
Macro _{Avg}	0.76	0.73	0.74	0.85	0.84	0.84	0.87	0.87	0.86	1605
Weighted _{Avg}	0.79	0.80	0.79	0.88	0.88	0.88	0.90	0.90	0.90	1605



(a) APTOS-2019 Dataset

(b) Combined IDRiD and Messidor-2 Dataset

Figure 10. Loss for the implemented schemes

At the end, the proposed deep learning-based detection system can accurately classify retinopathy images, permitting healthcare professionals to concentrate their expertise and time on the care of patients instead of manual image interpretation.

Improving the accuracy of retinopathy grade classification through preprocessing and deep learning schemes with multi-label classification can have several practical implications for healthcare professionals in detecting signs of diabetic retinopathy early. Early detection permits timely intervention and treatment, which can inhibit or retard the disease progression, hence, minimize the vision loss risk and other complications related to diabetic retinopathy. Additionally, healthcare professionals can evolve more treatment maps for patient's dependent on their specific retinopathy grade, hence streamlining workflow efficiency.

The results attained from the proposed retinopathy classification system have been compared with some of the recently presented relevant works to exhibit the effectiveness of the implemented deep learning classification schemes. This comparison is illustrated in Table 6.

Table 6: Comparison with closely relevant retinopathy classification systems in various terms of assessment measurements

Ref.	Datasets	Deep learning Schemes	No. of Classes	Acc
[13]	Messidor	ResNet	Five Classes	92.78%
[15]	APTOS-2019 dataset	Xception	Five Classes	93.10%
		InceptionV3		91.90%
[18]	APTOS-2019 dataset	ResNet-34	Five Classes	88.41%
		DenseNet-121		89.57%
[19]	APTOS dataset (80% training and 20% testing)	ResNet-18 (Swish)	Five Classes	93.51%
Proposed Retinopathy classification System	APTOS-2019 dataset (80% training and 20% testing)	VGG-19	Five Classes	88%
		DenseNet-121		91%
		EfficientNet-B6		94%
	combined IDRiD and Messidor-2 dataset(80% training and 20% testing)	VGG-19	Five Classes	80%
		DenseNet-121		88%
		EfficientNet-B6		90%

5. Conclusion

This paper aims to implement diverse effective deep-learning schemes to classify the severity levels induced by diabetic retinopathy disease. To strengthen the schemes' representative capacity, Gaussian Blur and weighted masking have been applied in the stage of pre-processing. Additionally, to adopting multi-label in the classification process. Three benchmark datasets (APOS-2019 and combined IDRiD and Messidor-2) were utilized for testing how well the implemented deep learning schemes work. It is obvious from analyzing and comparing various deep learning schemes that EfficientNetB6, has considerable possibility for classifying retinopathy. The proposed system outperforms other closely relevant retinopathy classification systems (In terms of utilized dataset), especially, the related works, and hence it will assist in the early diagnosis of diabetic retinopathy levels, i.e. treatments are possibly started at the right time. The application of the suggested system in real world presents considerable advantages, notably in enhancing the precision and rapidity of disease diagnosis, which can contribute to the mitigation of complications. Nevertheless, in spite of these merits, deep learning architectures are typically voluminous, necessitating significant computational resources and storage capabilities. To address this, future work should focus on the exploration of lightweight methodologies, thereby rendering these models more efficient and feasible for extensive implementation.

References

- [1] Krishnan Sangeetha, K. Valarmathi, T. Kalaichelvi, S. Subburaj (2023). A broad study of machine learning and deep learning techniques for diabetic retinopathy based on feature extraction, detection and classification, *Measurement: Sensors*, Vol. 30, 100951, <https://doi.org/10.1016/j.measen.2023.100951>.
- [2] V. Sathananthavathi, G. Indumathi (2023). 9 - Deep learning approaches for the retinal vasculature segmentation in fundus images, *Computational Methods and Deep Learning for Ophthalmology*, Academic Press, pp. 139-155. <https://doi.org/10.1016/B978-0-323-95415-0.00010-3>.
- [3] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D. I., & Marias, K. (2021b). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, 135, 104599. <https://doi.org/10.1016/j.compbiomed.2021.104599>
- [4] Mohammad Z. Atwany, Abdulwahab H. Sahyoun, Mohammad Yaqub (2022), Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey, in *IEEE Access*, vol. 10, pp. 28642-28655, doi: 10.1109/ACCESS.2022.3157632.
- [5] Nikos Tsiknakis, Dimitris Theodoropoulos, Georgios Manikis, Emmanouil Ktistakis, Ourania Boutsora, Alexa Berto, Fabio Scarpa, Alberto Scarpa, Dimitrios I. Fotiadis, Kostas Marias (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review, *Computers in Biology and Medicine*, Vol. 135, 104599. <https://doi.org/10.1016/j.compbiomed.2021.104599>.
- [6] Saad Albawi, M. H. Arif, Jumana Waleed (2022). Skin cancer classification dermatologist-level based on deep learning model, *Acta Scientiarum. Technology*, vol. 45, no. 1, e61531. <https://doi.org/10.4025/actascitechnol.v45i1.61531>.
- [7] Muhammad H. Obaid, Ali H. Hamad (2023). Deep learning approach for oil pipeline leakage detection using image-based edge detection techniques, *Journal Européen des Systèmes Automatisés*, Vol. 56, No. 4, pp. 663-673. <https://doi.org/10.18280/jesa.560416>.
- [8] Rubina Sarki, Khandakar Ahmed, Hua Wang, Yanchun Zhang (2020). Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey, in *IEEE Access*, vol. 8, pp. 151133-151149. doi: 10.1109/ACCESS.2020.3015258.
- [9]] Jiang, H., Xu, J., Shi, R., Yang, K., Zhang, D., Gao, M., Ma, H., & Qian, W. (2020). A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification. *IEEE*. <https://doi.org/10.1109/embc44109.2020.9175884>
- [10] Gour, N., & Khanna, P. (2020). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomedical Signal Processing and Control*, 66, 102329. <https://doi.org/10.1016/j.bspc.2020.102329>
- [11] Paisan Ruamviboonsuk, Richa Tiwari, Rory Sayres, Variya Nganthavee, Kornwipa Hemarat, Apinpat Kongprayoon, Rajiv Raman, Brian Levinstein, Yun Liu, Mike Schaekermann, Roy Lee, Sunny Virmani, Kasumi Widner, John Chambers, Fred Hersch, Lily Peng, Dale R Webster (2022). Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit Health*. 4(4):e235-e244. doi: 10.1016/S2589-7500(22)00017-6.
- [12] Hongyang Jiang, Yilin Hou, Hanpei Miao, Haili Ye, Mengdi Gao, Xiaoling Li, Richu Jin, Jiang Liu (2023). Eye tracking based deep learning analysis for the early detection of diabetic retinopathy: A pilot study, *Biomedical Signal Processing and Control*, Volume 84, 104830. <https://doi.org/10.1016/j.bspc.2023.104830>.
- [13]] Jiang, H., Xu, J., Shi, R., Yang, K., Zhang, D., Gao, M., Ma, H., & Qian, W. (2020b). A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification. *IEEE*. <https://doi.org/10.1109/embc44109.2020.9175884>
- [14] Pan, X., Jin, K., Cao, J., Liu, Z., Wu, J., You, K., Lu, Y., Xu, Y., Su, Z., Jiang, J., Yao, K., & Ye, J. (2020). Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus

- fluorescein angiography based on deep learning. Graefe S Archive for Clinical and Experimental Ophthalmology, 258(4), 779–785. <https://doi.org/10.1007/s00417-019-04575-w>
- [15] Vinuja S., Krishna Sameera A., Kaushek Kumar T. R., Uma Meenakshi R., Karthika R. (2021). Performance Analysis of Diabetic Retinopathy Classification using CNN, 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 823-828. doi: 10.1109/ICIRCA51532.2021.9544730.
- [16] Sai Kiran Reddy Meruva, Venkata Guru Sukesh Tulasi, Nithin Vinnakota, V Bhavana (2022). Risk Level Prediction of Diabetic Retinopathy based on Retinal Images using Deep Learning Algorithm, Procedia Computer Science, Volume 215, pp. 722-730. <https://doi.org/10.1016/j.procs.2022.12.074>.
- [17] T.R. Athira, Jyothisha J Nair (2023). Diabetic Retinopathy Grading From Color Fundus Images: An Autotuned Deep Learning Approach, Procedia Computer Science, Vol 218, pp. 1055-1066. <https://doi.org/10.1016/j.procs.2023.01.085>.
- [18] Bam Bahadur Sinha, R. Dhanalakshmi, K. Balakrishnan (2023). 2 - Early diagnosis of diabetic retinopathy using deep learning techniques, Computational Methods and Deep Learning for Ophthalmology, Academic Press, pp. 17-33. <https://doi.org/10.1016/B978-0-323-95415-0.00006-1>.
- [19] Serena Sunkari, Ashish Sangam, Venkata Sreeram P., Suchetha M., Rajiv Raman, Ramachandran Rajalakshmi, Tamilselvi S. (2024). A refined ResNet18 architecture with Swish activation function for Diabetic Retinopathy classification, Biomedical Signal Processing and Control, Volume 88, Part A, 105630, <https://doi.org/10.1016/j.bspc.2023.105630>.
- [20] Aravind Eye Hospital, APTOS 2019 blindness detection. <https://www.kaggle.com/datasets/mariaherrero/ptos2019>. Accessed on 20/09/2023.
- [21] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabudhe, Fabrice Meriaudeau (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research, Data 3 (3), <https://doi.org/10.3390/data3030025>.
- [22] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, Béatrice Charton, Jean-Claude Klein (2014). Feedback on a publicly distributed database: the Messidor database, Image Anal. Stereol. 33 (3). <https://doi.org/10.5566/ias.1155>.
- [23] Messidor-2. <https://www.kaggle.com/datasets/xyaustin/messidor2>.
- [24] H. Zafari et al., "Using Deep Learning with Canadian Primary Care Data for Disease Diagnosis", Deep Learning for Biomedical Data Analysis, Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-71676-9_12
- [25] Devi, T. G., Patil, N., Rai, S., & Philipose, C. S. (2023). Gaussian Blurring Technique for Detecting and Classifying Acute Lymphoblastic Leukemia Cancer Cells from Microscopic Biopsy Images. Life, 13(2), 348. <https://doi.org/10.3390/life13020348>
- [26] W.-H. Wang, D. C. Reutens, Z. Yang, G. Nguyen, V. Vegh, "Modified human contrast sensitivity function based phase mask for susceptibility-weighted imaging", NeuroImage: Clinical, vol. 4, pp. 765-778, 2014. <https://doi.org/10.1016/j.nicl.2014.04.012>.
- [27] L. C. M. Liaw, S. C. Tan, P. Y. Goh, C. P. Lim, "A histogram SMOTE-based sampling algorithm with incremental learning for imbalanced data classification", Information Sciences, 2024. <https://doi.org/10.1016/j.ins.2024.121193>
- [28] D. Wan, R. Lu, T. Xu, S. Shen, X. Lang, Z. Ren, "Random Interpolation Resize: A free image data augmentation method for object detection in industry", Expert Systems with Applications, vol. 228, 2023. <https://doi.org/10.1016/j.eswa.2023.120355>
- [29] Karen Simonyan, Andrew Zisserman (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556v6. <https://doi.org/10.48550/arXiv.1409.1556>.

- [30] Z. Cao, J. Huang, X. He, Z. Zong, "BND-VGG-19: A deep learning algorithm for COVID-19 identification utilizing X-ray images", *Knowledge-Based Systems*, vol. 258, 2022. <https://doi.org/10.1016/j.knosys.2022.110040>
- [31] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger (2016), "Densely connected convolutional networks". arXiv:1608.06993v5. <http://arxiv.org/abs/1608.06993>.
- [32] Mingxing Tan, Quoc V. Le (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv:1905.11946v5. <http://arxiv.org/abs/1905.11946>.
- [33] M. Tan, Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", 2019. arXiv:1905.11946v5. <http://arxiv.org/abs/1905.11946>
- [34] Chougrad, H., Zouaki, H., & Alheyane, O. (2020). Multi-label transfer learning for the early diagnosis of breast cancer. *Neurocomputing*, 392, 168–180. <https://doi.org/10.1016/j.neucom.2019.01.112>
- [35] Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. (2020). Multi-Label classification of Fundus images with EfficientNet. *IEEE Access*, 8, 212499–212508. <https://doi.org/10.1109/access.2020.3040275>