





Article

Balancing Data through Data Augmentation Improves the Generality of Transfer Learning for Diabetic Retinopathy Classification

Zahra Mungloo-Dilmohamud ^{1,*} , Maleika Heenaye-Mamode Khan ¹ , Khadiime Jhumka ¹ ,
Balkrish N. Beedassy ², Noorshad Z. Mungloo ² and Carlos Peña-Reyes ^{3,4} 

- ¹ Faculty of Information, Communication and Digital Technologies, University of Mauritius, Réduit 80837, Mauritius; m.mamodekhan@uom.ac.mu (M.H.-M.K.); khadiime@gmail.com (K.J.)
 - ² Ministry of Health and Wellness, Quatre Bornes 72259, Mauritius; drkrisb12@gmail.com (B.N.B.); noorshad@hotmail.com (N.Z.M.)
 - ³ School of Management and Engineering Vaud (HES-SO), University of Applied Sciences and Arts Western Switzerland Vaud, 1400 Yverdon-les-Bains, Switzerland; carlos.pena@heig-vd.ch
 - ⁴ CI4CB—Computational Intelligence for Computational Biology, SIB—Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
- * Correspondence: z.mungloo@uom.ac.mu



Citation: Mungloo-Dilmohamud, Z.; Heenaye-Mamode Khan, M.; Jhumka, K.; Beedassy, B.N.; Mungloo, N.Z.; Peña-Reyes, C. Balancing Data through Data Augmentation Improves the Generality of Transfer Learning for Diabetic Retinopathy Classification. *Appl. Sci.* **2022**, *12*, 5363. <https://doi.org/10.3390/app12115363>

Academic Editors: Lucian Mihai Itu, Constantin Suciuc and Anamaria Vizitiu

Received: 30 March 2022

Accepted: 10 May 2022

Published: 25 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The incidence of diabetes in Mauritius is amongst the highest in the world. Diabetic retinopathy (DR), a complication resulting from the disease, can lead to blindness if not detected early. The aim of this work was to investigate the use of transfer learning and data augmentation for the classification of fundus images into five different stages of diabetic retinopathy. The five stages are No DR, Mild nonproliferative DR, Moderate nonproliferative DR, Severe nonproliferative DR and Proliferative. To this end, deep transfer learning and three pre-trained models, VGG16, ResNet50 and DenseNet169, were used to classify the APTOS dataset. The preliminary experiments resulted in low training and validation accuracies, and hence, the APTOS dataset was augmented while ensuring a balance between the five classes. This dataset was then used to train the three models, and the best three models were used to classify a blind Mauritian test datum. We found that the ResNet50 model produced the best results out of the three models and also achieved very good accuracies for the five classes. The classification of class-4 Mauritian fundus images, severe cases, produced some unexpected results, with some images being classified as mild, and therefore needs to be further investigated.

Keywords: deep learning; diabetic retinopathy; retinal fundus images; transfer learning; data augmentation

1. Introduction

Diabetes is one of the most challenging health problems in the world, impacting roughly 537 million individuals according to the IDF Diabetes Atlas Tenth edition 2021 (Diabetes Atlas, 2021). According to the same atlas, countries have spent over USD 966 billion on diabetes patients worldwide, a 316 percent increase over the previous 15 years, and yet diabetes will be responsible for 6.7 million deaths in 2021, or 1 death every 5 s. Diabetes poses a danger to the health-care systems of low- and middle-income nations, which account for 75 percent of the world's diabetic population, resulting in many cases going undetected. The most common complication in advanced or uncontrolled diabetic patients is diabetic retinopathy, one of the leading cause of vision loss worldwide, accounting for 21.8 percent of patients across the globe [1]. With Mauritius currently ranking fifth in the global standardized diabetes prevalence among ages 20–79 in 2019 and predicted to reach the second position in 2030 [2], diabetic retinopathy is a serious threat to Mauritians. This is especially true for people in their working years, since this group is more susceptible as

per the article “Global estimates of the prevalence of diabetes for 2010 and 2030 in Diabetes Atlas”. Patients who have had vision loss as a result of this condition typically have a late diagnosis of diabetes or are unaware that they have diabetes and eye difficulties. A recent study [3] found that diagnosing retinopathy early can prevent or delay a substantial amount of vision loss. This can also help to speed up the healing process or halt disease development. However, establishing a precise diagnosis and the stage of the disease is difficult. Ophthalmologists conduct screenings by visually inspecting the fundus and evaluating colour images. They rely on detecting the presence of microaneurysms, small saccular outpouching of capillaries, retinal haemorrhages and ruptured blood vessels, among many indicators, in the fundoscopic images. This manual method, however, results in inconsistency among readers [4] and is costly and time-consuming. To address the growing number of undiagnosed retinal patients, early disease identification and treatment are critical.

Advancements in convolutional neural networks (CNNs), a type of deep learning, has motivated researchers to use them in medical image analysis for different tasks, amongst which is image classification of diabetic retinopathy. CNNs exhibit a better performance, but they also need a lot of computing resources and large datasets to train. Transfer learning (TL) strategies have been proposed to solve this problem [5–7]. It involves using a previously learned model, on different images, to train a new model. The traits learned by pre-training on the large dataset can be transferred to the new network, where only the classification component is trained on the new smaller dataset, to fine-tune the new data [7]. TL reduces the amount of time spent constructing and training a deep CNN model as well as the computing resources needed. The visual geometry group (VGG) [8], inception modules (GoogleNet) [9], residual neural network (ResNet) [10] and neural architecture search network (NasNetLarge) [10] are examples of the many high-performing pre-trained models found in the literature. In 2017, Masood et al. [11] applied a pre-trained Inception V3 model on the Eye-PACS fundus dataset and achieved an accuracy of 48.2%. Meanwhile, Li et al. [12] investigated the use of transfer learning for identifying DR by comparing several network topologies, such as AlexNet, VGG-S, VGG16 and VGG19, to two datasets: the Messidor and DR1 datasets. With an area under the curve (AUC) of 98.34%, the VGG-S architecture scored the best AUC for the Messidor dataset while an AUC score of 97.86% was obtained for the DR1 dataset. Similarly, in 2019, using the EYE-PACS dataset, Challa et al. [13] proposed a deep All-CNN architecture for DR classification. The model obtained an accuracy of 86.64%, a loss of 0.46 and an average F1 score of 0.6318. Meanwhile, using the Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection (APTOS 2019 BD) dataset [14], Kassani et al. [15] described a classification method using a modified Xception architecture model, which is an extension of the Inception architecture, on the dataset and obtained an accuracy of 83.09%, a sensitivity of 88.24% and a specificity of 87.00%. Khalifa et al. [16] implemented transfer learning using four pre-trained models, namely AlexNet, Res-Net18, SqueezeNet and GoogleNet. AlexNet obtained the highest accuracy of 97.9%. In Hagos et al. [17], a pre-trained Inception V3 model was applied to a subset of the APTOS dataset for DR classification, and the accuracy was 90.9% and the loss was 3.94%. Sikder et al. [18] presented a method incorporating the ExtraTree classifier, which is a popular ensemble learning algorithm based on decision trees and bagging learning techniques, and achieved a classification accuracy of 91%. In 2020, Shaban et al. [19] proposed a modified version of the VGG-19 that achieved an accuracy of 88%–89% when both 5-fold, and 10-fold cross validation methods were used, respectively. Using the same APTOS 2019 BD dataset, Mushtaq et al. [20] achieved a classification accuracy of 90% using a pre-trained Dense169 model. Before they trained the images, the latter were pre-processed by removing the black border and applying Gaussian blur filter. Moreover, Thota et al. [21] fine-tuned a pre-trained VGG16 model for classifying the severity of DR. An average class accuracy of 74%, sensitivity of 80%, specificity of 65% and AUC of 0.80 were achieved. Gangwar et al. [22] developed a novel deep learning hybrid model with pre-trained Inception-ResNet-v2 as a base model and it obtained a test accuracy of 72.33%

on Messidor-1 and 82.18% on the APTOS dataset. On the other hand, Dai et al. [23] used a deep learning model based on the ResNet architecture to classify fundus images into five different classes. Images were obtained from the Shanghai Integrated Diabetes Prevention and Care System study. Firstly, the different features, such as microaneurysm, hard exudate and haemorrhage were detected, and then they concatenated the model used and the base model for DR classification. The model achieved AUCs of 0.943, 0.955, 0.960 and 0.972, for mild, moderate, severe and proliferative cases. Benson et al. [24] discussed the usage of transfer learning by using a pre-trained Inception V3 on the DR dataset obtained from the VisionQuest Biomedical database. The model classified fundus images into six classes including identifying scars, and it achieved a sensitivity and specificity of 90%, with an AUC of 95%.

The reviews described above highlight the fact that all work carried out to date was for images from a specific country, and hence they were not targeted at a local multiracial population such as Mauritius [25,26]. Therefore, this research work makes the following contributions:

- (1) Application of three pre-trained models, VGG16, DenseNet169 and ResNet50, on a publicly available diabetic retinopathy dataset and the data-augmented version of the dataset to solve the class imbalance problem;
- (2) Enhance the pre-trained models to improve the performance obtained in (1);
- (3) Apply the enhanced models on a blind Mauritian local cohort to predict the different stages of diabetic retinopathy;
- (4) Compare the predicted results obtained for the Mauritian dataset using the enhanced models to an actual ophthalmologist's diagnosis.

The paper is structured as follows. Section 2 presents the proposed solution and describes the different components. Section 3 discusses the experimental results. Finally, Section 4 concludes the paper.

2. Materials and Methods

This section highlights the methodology used in implementing deep transfer learning for classification.

2.1. Proposed Workflow and Components

Figure 1 shows the proposed workflow for the system, which can accept different datasets. For this work, two datasets, the APTOS original dataset and a constructed Mauritian dataset, were used. The data were first pre-processed, and data augmentation was applied to the APTOS dataset only. Next, three pre-trained models were applied to the original and augmented APTOS dataset. The results were analyzed, and the models were tuned to reach their ideal minima. The enhanced models were then applied to the blind testing data from the APTOS dataset and the labelled Mauritian dataset, which was not used for the training phase.

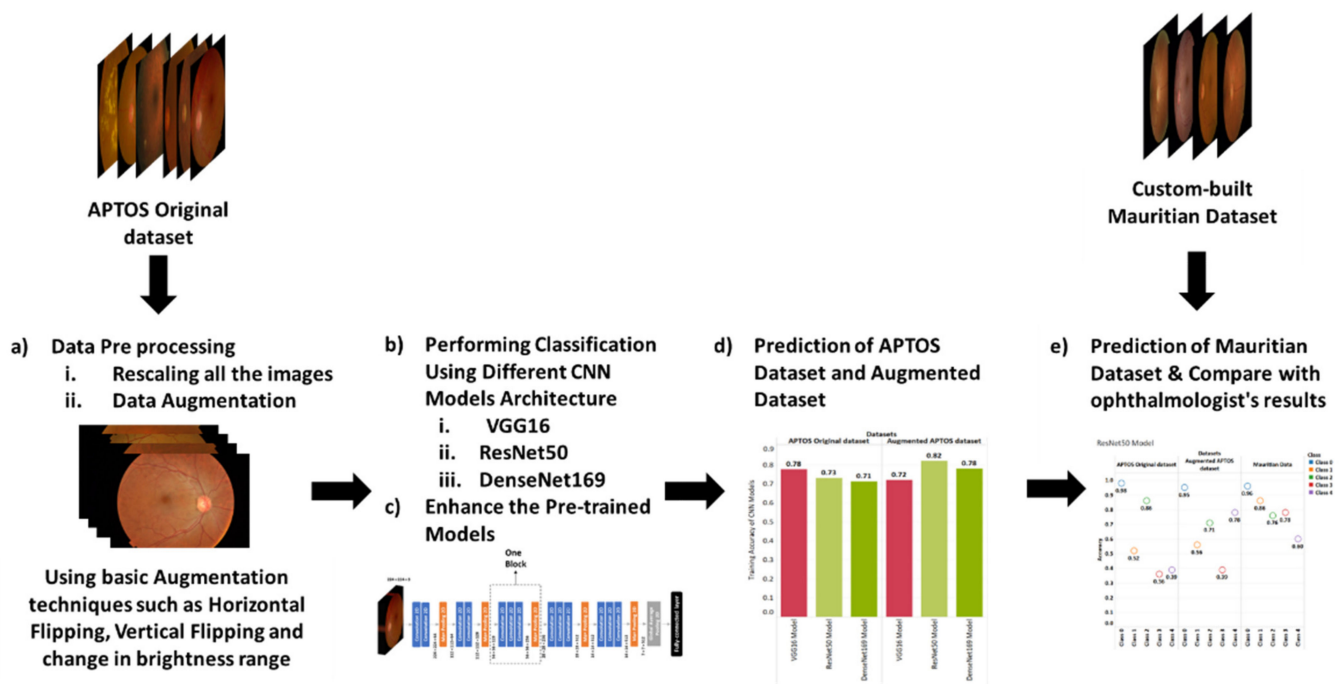


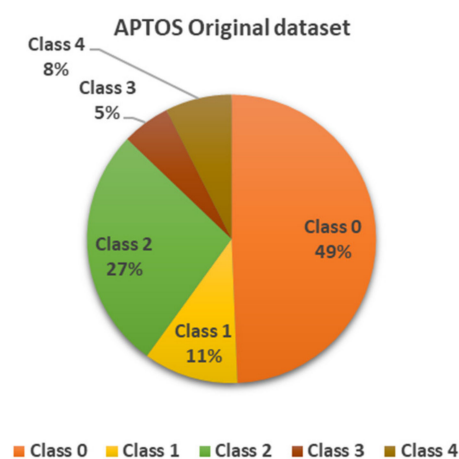
Figure 1. Workflow of our proposed system.

The workflow shown in Figure 1 is as follows: (1) data pre-processing and augmentation (for the APTOS dataset only); (2) training and enhancing the CNN models using the original and augmented APTOS dataset; (3) analyzing results; and (4) classification of the images for the 3 datasets and comparison to actual data.

2.2. Datasets

In this research work, two fundus image datasets were used. The first dataset was the APTOS 2019 diabetic retinopathy dataset, which is publicly available online on Kaggle (<https://www.kaggle.com/c/aptos2019-blindness-detection/data>, accessed on 17 February 2022). This dataset was selected among the other publicly available datasets since it is from India, which is close to the Mauritian population in terms of ethnicity. The second dataset was created locally from the images obtained from the hospitals in Mauritius. Each image in the APTOS 2019 dataset was assigned a class label of 0–4 according to the severity of the disease, as shown in Figure 2. Each image from the local cohort was also assigned a class label of 0–4 by a local doctor. The original dataset obtained from Kaggle is termed as the original APTOS dataset. The class distribution of the original APTOS dataset is illustrated in Figure 2.

Figure 2 reveals that, despite the data belonging to five different classes, the number of samples in each class varied substantially, resulting in an unbalanced dataset. As discussed in [27–29], an unbalanced dataset leads to a high misclassification rate and sub-optimal performance. To mitigate this challenge, we applied data augmentation, which is one possible solution to this problem. Traditional data augmentation techniques, namely horizontal and vertical flipping and changes in the brightness range [30], were applied to the original APTOS datasets to produce the augmented APTOS dataset.



Class Label	Severity of Diabetic Retinopathy (DR)
0	No DR
1	Mild nonproliferative DR (Mild DR)
2	Moderate nonproliferative DR (Moderate DR)
3	Severe nonproliferative DR (Severe DR)
4	Proliferative

Figure 2. Original APTOS dataset.

Table 1 shows the total number of images for each class in the original APTOS dataset, the augmented APTOS dataset and the local Mauritian dataset. We divided both the APTOS dataset and the augmented APTOS dataset into a training set and testing set. There were 3662 images in the original APTOS dataset, whereby 70% (2563 images) were considered for training and 30% (1099 images) were taken for the testing phase. For the augmented APTOS dataset, data augmentation was performed on the training set only as performed by Gangwar et al. [22]. Only the data from classes 1, 3 and 4 were augmented since the model could not correctly classify these 3 classes in the original APTOS dataset. All the images in these 3 classes were augmented. In this paper, we used two sets for testing data, one which is made up of fundus images from the APTOS 2019 dataset (the remaining 30% of which were not used as training data) and the second being the Mauritian dataset composed of fundus images obtained from a local hospital in Mauritius. Table 1 presents the image count for each class in the training and testing data for the original and augmented APTOS datasets as well as the Mauritian dataset.

Table 1. Number of images class-wise in the 3 datasets.

Training Data	Number of Images in Training/Validation Dataset					Number of Images in Testing Dataset				
	Class 0	Class 1	Class 2	Class 3	Class 4	Class 0	Class 1	Class 2	Class 3	Class 4
Original APTOS dataset	1265	272	697	138	191	540	98	302	55	104
	Total images—2563					Total images—1099				
Augmented APTOS dataset	1265	1306	697	935	1264	540	98	302	55	104
	Total images—5467					Total images—1099				
Mauritian dataset	No training performed using Mauritian data					54	62	45	12	33
						Total images—208				

Figure 3 presents the number of images in each of the 5 classes after the application of data augmentation on the original APTOS dataset. It can be observed that the augmented dataset was more balanced.

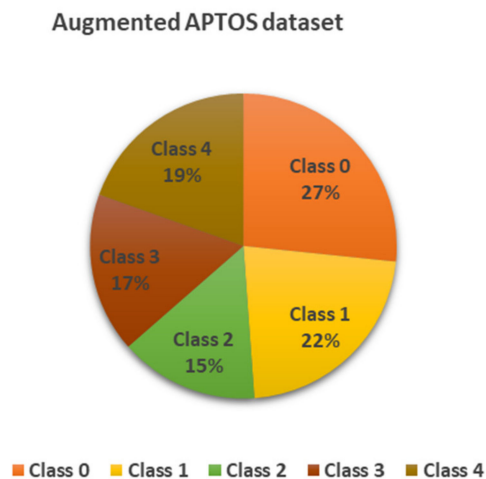


Figure 3. Augmented APTOS dataset.

2.3. Data Pre-Processing

The images were subjected to a pre-processing phase to improve their quality. They were resized as each model accepted images of different resolutions. For the ResNet50 Model, the images were resized to 512×512 pixels, whereas they were resized to 224×224 pixels for the VGG16 and DenseNet169 models. Another reason for performing pre-processing was the varying size and resolution of photos collected from the Kaggle website. These pictures ranged from 474×358 pixels to 3388×2588 pixels in width and height. After pre-processing the images, the different CNN models were applied to the training data of the two APTOS datasets to perform classification.

2.4. Transfer Learning Using ResNet50, VGG16 and DenseNet169

In this paper, transfer learning (TL) using the architectures of the three CNNs models, ResNet50, VGG16 and DenseNet169, was applied to the diabetic retinopathy images. In TL, learned features from one task are applied to a different task without having to learn from scratch. This is commonly used when building CNN models since the process of training from scratch requires a lot of computational resources, large datasets and a lot of time [31]. CNN models consist of multiple layers, namely: the convolution layer, pooling layer and fully connected layer. CNN models employ multiple perceptrons to evaluate picture inputs and eventually extract different patterns from the images to output to the fully connected layer. Our CNN models extracted representative patterns to form the feature maps. A 3×3 kernel was passed over the input matrix of the diabetic retinopathy image, as illustrated in Figure 4.

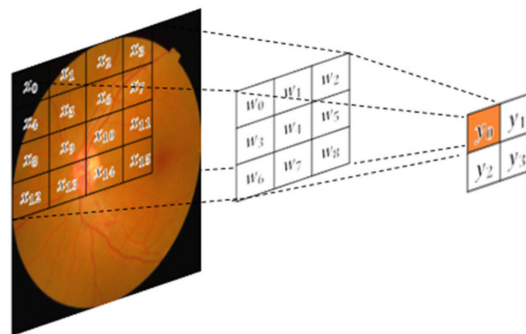


Figure 4. Convolution layer.

The classification function, which is the output of the fully connected layer, plays an important role in the process, whereby the different patterns of the five stages of diabetic retinopathy, learnt by the feature extraction layers, are used to perform the multiclass classification.

The VGG16 model, a CNN architecture pre-trained on the ImageNet dataset, was adopted for the development of our diabetic retinopathy application as it has been fully tested in a similar domain, achieving good performance [32,33]. VGG16 consists of 13 convolutional layers and 3 fully connected layers. There are 5 blocks each containing 2 or 3 convolution layers and ending with a max-pooling layer, as illustrated by Figure 5. A fixed-size image of dimensions (224, 224, 3) is the input to the VGG16 model.

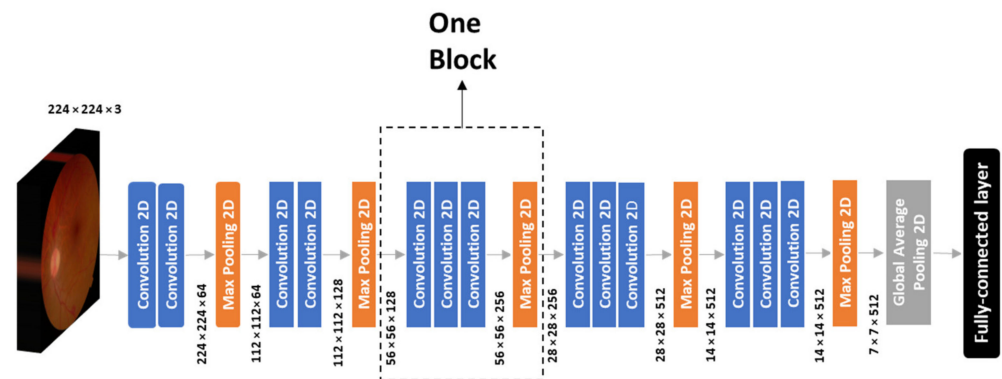


Figure 5. Architecture of the VGG16 model.

ResNet50, another popular CNN architecture, consists of 50 layers organized in so-called residual blocks [9]. It is known for its skip connection approach, which eventually solves the vanishing gradient problem. ResNet50 contains 48 convolution layers along with 1 MaxPool and 1 AveragePool layer. This was desired in our diabetic retinopathy application as it allows the later layers to learn lesser semantic information that was captured in the early layers. A 3×3 filter was used to perform the spatial convolution, which was eventually reduced using the max-pooling method. Figure 6 illustrates the ResNet 50 model with the 48 convolution layers and the 16 skip connections.

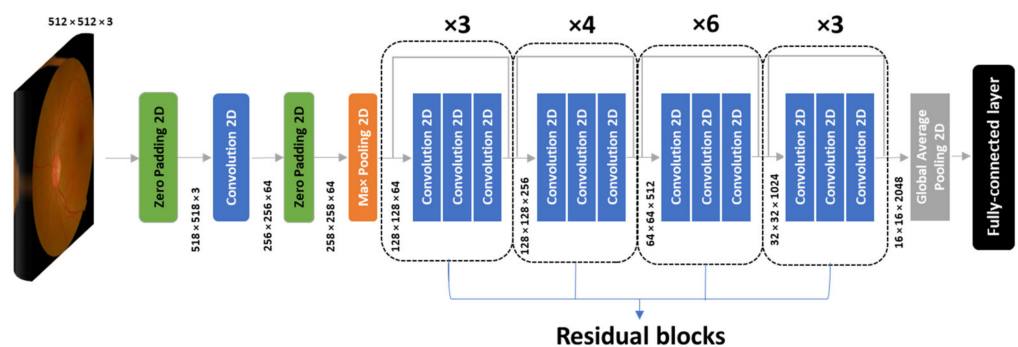


Figure 6. Architecture of ResNet50 model.

The third model that was considered was the DenseNet169 model [34]. Compared to the ResNet50 model, it has more layers. However, it contains a similar block to skip connections called the dense block. With the increase in the number of layers, it gives the model the opportunity to learn more distinctive features. In fact, the architecture consists of four dense blocks with varying numbers of layers as illustrated in Figure 7. Our design for this model consisted of the 2D average pooling, which is in the original architecture, where a dropout layer set to 0.5 was added.

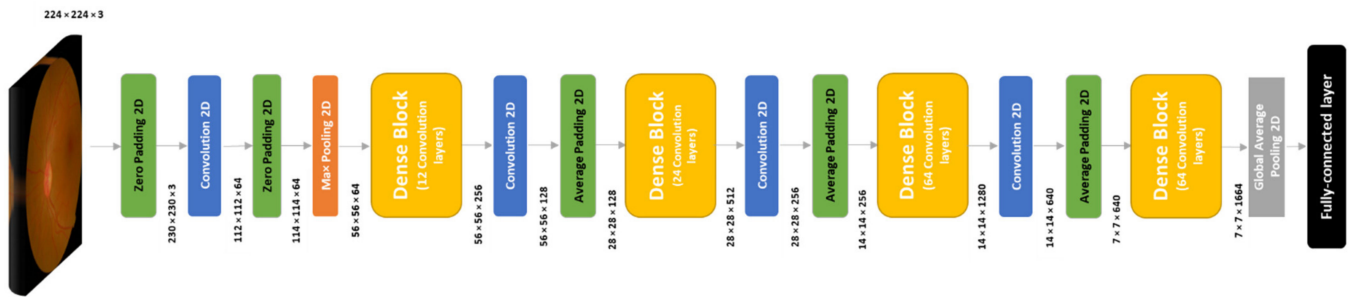


Figure 7. Architecture of DenseNet169 model.

2.5. Enhanced CNN Models

Initially, the architectures of VGG16, ResNet50 and DenseNet169 were applied to the APTOS dataset. To be able to use these architectures in transfer learning and for classifying the diabetic retinopathy images into five classes, fully connected layers were added. The 3-dimensional feature map obtained from the last convolutional layer was converted to one dimension by using global average pooling 2D and passed to a series of a dropout layer, a dense layer and a dropout layer and finally to a dense layer with five nodes, representing the normal and the DR grades. The fully connected layers were selected as in the ResNet model in Taormina et al. [35], and Zhang et al. [36] shows that adding fully connected layers yields better results. The activation function used in the last dense layer was Softmax, as used in ElBedwehy et al. [37] for face detection classification. The Adam optimizer was applied to the 3 models with a learning rate of 1×10^{-3} , and the loss entropy used was the categorical cross entropy. In this work, data balancing was performed using basic image manipulation techniques [38]. In the deep neural network, the Adam optimizer was used instead of the stochastic gradient descent (SGD) since the former is computationally more efficient. The Adam optimizer has been found to be faster in converging the algorithm to the minima, hence reducing the training time [39]. The use of the SGD and other approaches will be explored in future works. Here, only the last 5 layers, namely the global average pooling 2D, dropout, dense, dropout and dense layers, were trained. The other layers were frozen as we were only extracting the features from the base model. These steps resulted in the models producing the relevant learnable parameters during the training process. For example, for the ResNet model, out of the 27,794,309 parameters, 4,206,597 were trainable. In this work, the sequential modelling approach was adopted for adding and customizing the convolution, dropout, dense and optimizer layers. The sequential model is appropriate for a plain stack of layers whereby each layer has exactly one input tensor and one output tensor, which was the case in this application.

To improve the performance of the models and cater for underfitting/overfitting, the 3 models were fine-tuned. The Adam optimizer was again used but this time with a learning rate of 10^{-4} . The learning rate was decremented by 10 as this has been shown to both reduce the risk of overfitting [40] and to improve classification [41]. When the validation loss metric stopped improving, the learning rate was halved as in [42]. Several parameters were changed and added to the models for fine-tuning. Firstly, the loss function was changed to binary cross entropy. Using the latter along with a SoftMax classifier helped the model in reducing the cross entropy loss of each iteration in multiclass classification [43]. Afterwards, an early stopping feature was added to end training when the network began to overfit the data according to the validation loss [44]. Eventually, all the convolutional layers were unfrozen, and the models were set to be trained.

The enhanced transfer learning model that was trained on the augmented APTOS dataset was tested on APTOS test data and on a blind Mauritian test datum annotated by a medical practitioner.

3. Results and Discussions

To evaluate the trained models both before and after fine-tuning, the accuracy regarding training, validation and test sets was calculated. Classification accuracy is the fraction of predictions that a given model predicted correctly. Firstly, a custom-built CNN model similar to that developed by Jayalakshmi et al. [45] was used. The same fully connected layers as in the case of our pre-trained models were joined, and the hyperparameters were tuned to obtain the optimal accuracy. A classification accuracy of 0.73 was obtained here. The model was only able to correctly predict classes 0 and 2. Although the accuracy is quite satisfactory for a binary classification of DR and NoDR, this custom-built model is very limited in the case of a multiclass DR classification. Next, pre-trained networks were implemented. The training and validation accuracy obtained before fine-tuning of the pre-trained networks are illustrated in Figure 8. From the results, it was found that the accuracies were quite low for the models ResNet50 and DenseNet169. Hence, it was deduced that these models were underfitting.

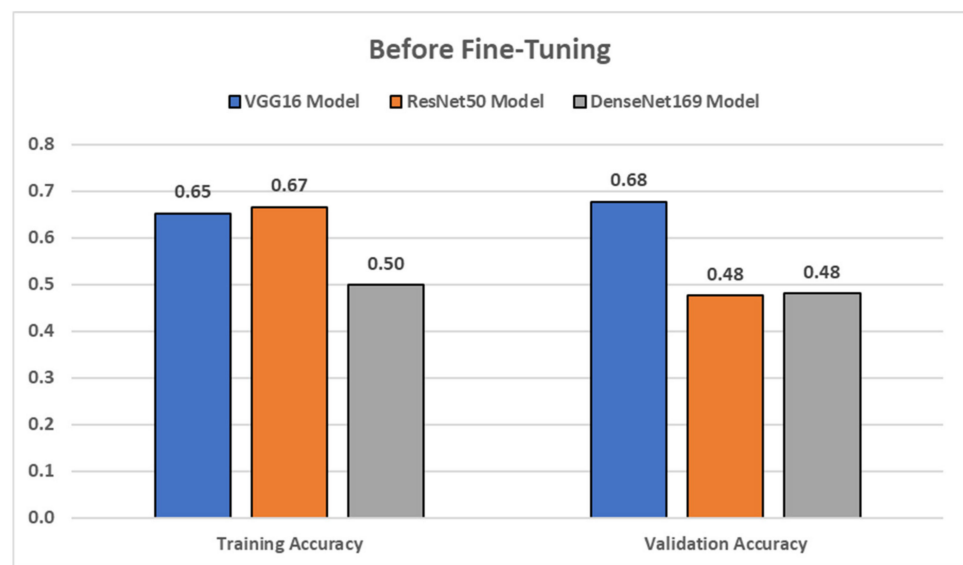


Figure 8. Overall training and validation accuracy before fine-tuning for the original APTOS dataset (after 2 epochs).

Consequently, the models were enhanced, and the weights were adjusted. Different learning rates were applied and evaluated to reach the minima. In addition, the number of epochs were adjusted while analyzing the different accuracies, thus fine-tuning the models. Each model was trained on the same training set used in the previous process. Figure 9 shows the results obtained for training and validation accuracy for each of the three models after fine-tuning.

From Figures 8 and 9, it can be clearly seen that fine-tuning the models improved both the training and the validation classification accuracy of the three models for the original APTOS dataset. We also noticed that using the augmented data improved the generality of transfer learning for the models for both the training and validation data. This can be deduced from the accuracy for the augmented dataset being maintained or increasing across all models compared to the original dataset. Furthermore, ResNet, with the highest accuracy in all cases, showed a better generalization. In parallel, it was also observed that the time taken to train the model decreased considerably (by at least 3 h).

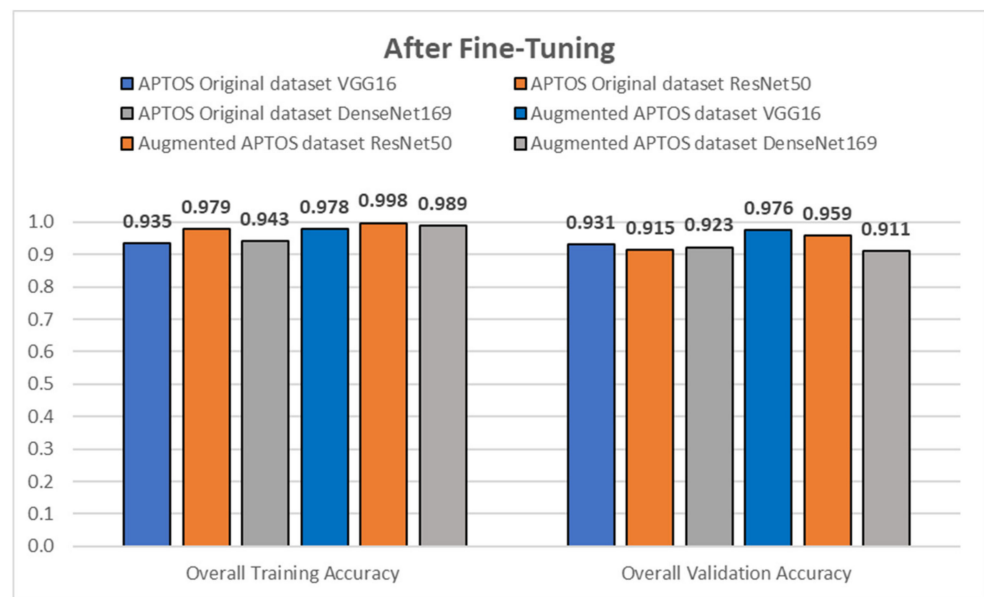


Figure 9. Overall training and validation accuracy of the CNN models after fine-tuning.

Both the overall training accuracy and the validation accuracy were above 90, which is a good indication that the six trained models were able to guess the label for nearly all of the training and validation sets of images. In three out of the six different CNN model training, with ResNet50 using both the original APTOS dataset and the augmented APTOS dataset as the training data, and the DenseNet169 model using the original APTOS dataset as the training data, early stopping occurred to prevent the models from overfitting.

Next, the six models were used to predict the class of the images in the testing data of both the APTOS and the Mauritian datasets. Figure 10 shows the overall testing accuracy obtained with the three CNN models for the original and augmented APTOS datasets. For the ResNet50 model and DenseNet169 model, increases of 9% and 7% were observed, respectively, when dealing with the augmented and balanced dataset. As for the VGG16 model, a decrease of 6.9% was noted for the augmented APTOS dataset.

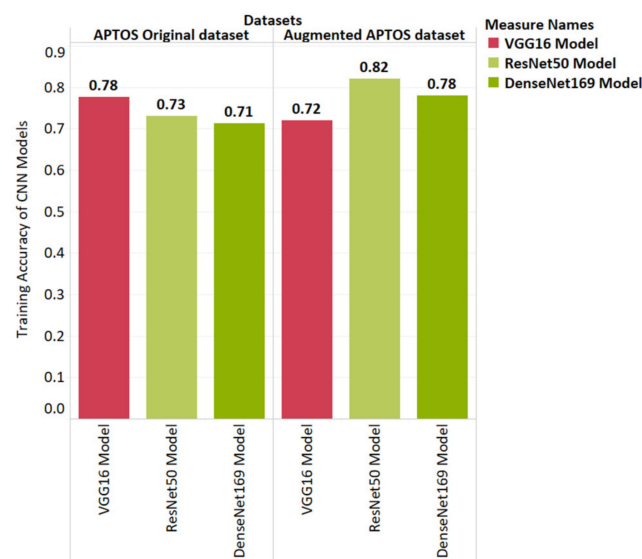


Figure 10. Testing accuracy of the CNN models for the APTOS dataset.

However, this overall testing accuracy for the data is not a good indicator of performance as the proportion of classes in the datasets was different. For example, in the

original APTOS dataset, the number of images belonging to class 0 makes up nearly half of the original data, whereas the images in the augmented APTOS dataset are more or less equally distributed among the different classes. Hence, the models will exhibit bias towards class 0 when they are applied to the original APTOS dataset, whereas for the augmented APTOS dataset, the proportion is nearly the same, so comparing the overall testing accuracy between the two datasets is not recommended. To address this issue, the class-wise accuracy was calculated for the three datasets and plotted as shown in Figure 11.



Figure 11. Detailed testing accuracy for each class for the 3 datasets and the 3 models.

A closer study of the plots in Figure 11 shows that the three models were able to predict class 0, “No DR” cases, quite easily for both the original and augmented APTOS datasets; however, only the ResNet50 model was able to classify “No DR” cases for the Mauritius dataset. This is to be expected since class 0 is quite distinct from the other classes given the absence of DR features such as microaneurysms.

For the VGG16 model, class 3 was the one that achieved the lowest accuracy out of all three datasets with none of the 55 cases being correctly classified for the original APTOS dataset. We also noted that none of the cases of class 1 for the Mauritian dataset were correctly identified. This shows that the model was unable to learn to distinguish the features of these two classes. A closer look at the results obtained shows that most of the

cases for class 3 were misclassified as class 2 and a few cases as classes 1 and 4. Class 3 represents the moderate cases, which fall between the mild and proliferative cases and therefore may be difficult to identify. There may be intraretinal haemorrhage, which also complicates the task.

For the ResNet50 model, classes 1, 3 and 4 were the most difficult to classify for the original dataset, classes 1 and 3 were the most difficult to classify for the augmented dataset, and class 4 was the most difficult class to classify for the Mauritian dataset. The difficulty in the classification of class 4 for the Mauritian cohort may be due to choroidal fronts and troughs being more pronounced in the local dataset due to presence of pigments. This is due to the local population having different skin colours.

For the Densenet169 model, the results obtained for the three datasets are variable with classes 1 and 4 being the less distinctive for the original dataset, classes 1 and 3 being less distinctive for the augmented dataset and classes 1, 2 and 3 being less distinctive for the Mauritian dataset. Here, none of the 202 cases for classes 1 and 4 in the original APTOS dataset were correctly identified. A closer look at the class-wise results shows that most of the images from class 1 were wrongly classified as class 2, and a few were classified as classes 0 and 3. Similarly, for class 4, we found that most of the images from class 4 were wrongly classified as class 2 and the rest as class 3. Based on these results, we concluded that for the APTOS dataset, classes 1 and 3 were the most difficult to learn.

Although none of the models had been trained with the data from Mauritius, the ResNet50 model achieved quite good results on this blind test dataset, achieving accuracies of 60% and above. It also obtained the best results compared to the other two models. This can be explained by the fact that the Densenet169 has more layers and may be overlearning and therefore generalizing less. Resnet50 has residual connections between layers, meaning that the output of a layer is a convolution of its input plus its input. It is also deeper than VGG16 with fewer parameters and is better able to identify the features to distinguish between the different classes of diabetic retinopathy. Moreover, although ResNet is much deeper than VGG16, the model size is substantially smaller due to the use of global average pooling rather than fully connected layers. Based on the results of the ResNet50 model, the results were further investigated, and a confusion matrix of the predicted vs. actual results was plotted, as shown in Figure 12.

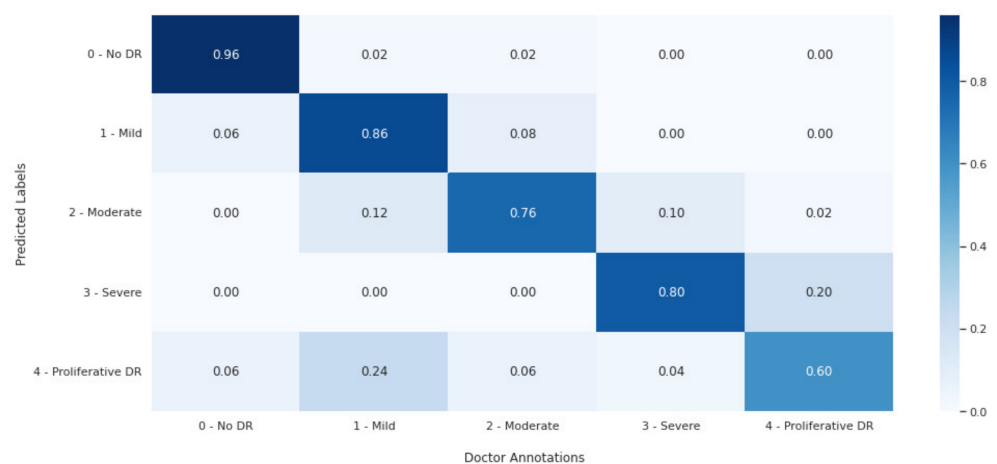


Figure 12. Confusion matrix for Mauritian data classified by ResNet50.

The precision, recall and F1 score were computed for each individual class and are displayed in Table 2. Additionally, the weighted average was also calculated.

Table 2. Performance metrics for Mauritian data classified by ResNet50.

	Precision	Recall	F1 Score
Class 0	0.9600	0.8889	0.9231
Class 1	0.8600	0.6935	0.7679
Class 2	0.7551	0.8222	0.7872
Class 3	0.7778	0.5000	0.6087
Class 4	0.6000	0.9091	0.7229
Weighted Average	0.8165	0.7933	0.7945

From the confusion matrix, we found that very good accuracies were achieved for all the classes, and the cases that were wrongly classified were close to the diagonal, being either from the class just before or just after. Thus, for class 0, the cases that were wrongly classified were actually from classes 1 and 2, for class 1, they were from classes 0 and 2, for class 2, they were from classes 1 and 3 with few cases from class 4, and for class 3, they were from class 4. This behaviour is not followed by class 4, where the wrongly classified classes were from all classes with the majority from class 2, which is quite far from class 4. Class 4 is of interest and requires further investigation. A comparison of our proposed model with the other available works in DR classification is given in Table 3.

Table 3. Comparison table of similar work.

Authors	Techniques Used	Discussions
Dai et al. [23]	Model: deep model based on ResNet Dataset: Shanghai Integrated Diabetes Prevention and Care System (Shanghai Integration Model, SIM) between 2014 and 2017 Number of images: 666,383 images	Pre-trained models (ResNet and R-CNN) were used. ROC was used to evaluate performance. Performance: AUC scores of 0.943, 0.955, 0.960 and 0.972 for mild, moderate, severe and proliferative cases were achieved, showing good performance using transfer learning
Masood et al. [11]	Model: pre-trained Inception V3 model Dataset: Eye-PACS dataset Number of images: 3908 images (800 from each class except 708 from class 4)	Performance: accuracy—48.2%, limitations: low accuracy
Li et al. [12]	Model: different pre-trained networks such as AlexNet, VGG-S, VGG16 and VGG19 Dataset: the Messidor and DR1 datasets Number of images: 1014 images (DR1), 1200 images (Messidor)	Performance: best area under the curve (AUC) (VGG-S)—98.34% (Messidor dataset), 97.86% (DR1 dataset) Limitations: number of classes is limited to DR and No DR only
Challa et al. [13]	Model: developed a deep All-CNN architecture Dataset: Eye-PACS dataset Number of images: 35,126 images	Performance: accuracy—86.64%, loss—0.46, average F1 score—0.6318 Limitation: no detailed information on overfitting
Khalifa et al. [16]	Model: AlexNet, Res-Net18, SqueezeNet and GoogleNet Dataset: APTOS dataset Number of images: 3662 images	Performance: best accuracy (AlexNet)—97.9% Limitation: high computational power needed (Intel Xeon E5-2620 processor (2 GHz), 96 GB of RAM) since the model needed to train on 14,648 images. Additionally, no detailed information was given for model overfitting during the training phase. The only method used to counter overfitting was data augmentation, which takes place before the model training phase.

Table 3. Cont.

Authors	Techniques Used	Discussions
Hagos et al. [17]	Model: pre-trained Inception V3 model Dataset: APTOS dataset Number of images: 2500 images (1250 for NoDR and 1250 for DR)	Performance: accuracy—90.9%, loss—3.94% Limitation: number of classes is limited to DR and No DR only
Gangwar et al. [22]	Model: deep learning hybrid model with pre-trained Inception-ResNet-v2 as a base model Dataset: Messidor-1 and APTOS dataset Number of images: 1200 images (Messidor-1), 3662 images (APTOS)	Performance: accuracy—72.33% (Messidor-1), 82.18% (APTOS dataset) Limitation: did not check whether model was overfitting
Benson et al. [24]	Model: pre-trained Inception V3 model Dataset: DR dataset obtained from VisionQuest Biomedical database Number of images: 6805 images	Performance: sensitivity—90%, specificity—90%, AUC—95% Limitation: results for No DR, MildDR, Moderate DR were 47%, 50% and 35%
Thota et al. [21]	Model: Fine-tuned and pre-trained VGG16 model Dataset: Eye-PACS dataset Number of images: 34,126 images	Performance: accuracy—74%, sensitivity—80%, a specificity—65%, AUC—80% Limitation: low accuracy compared to similar experimentations
Our proposed Model	Model: Fine-tuned and pre-trained ResNet50, VGG16, DenseNet169 models Dataset: APTOS dataset, Mauritian dataset Number of images: 3662 images (APTOS), 208 images (Mauritius)	Performance: accuracy (ResNet50)—82% (APTOS dataset), 79% (Mauritian dataset) Novelty: performed multiclass classification (5 different classes) for Mauritian dataset

Compared to similar work carried out in the field of DR classification, our proposed enhanced model was able to classify the different stages of diabetic retinopathy for a Mauritian dataset. The enhanced model was trained using the APTOS augmented dataset, and this model was used to classify the Mauritian dataset images with an overall accuracy of 79%. Furthermore, it can be said that our proposed model can be used for early detection of DR compared to Benson et al. [24], where the proposed model had a low accuracy for the early stages of DR. Meanwhile, Li et al. [12] and Hagos et al. [17] applied transfer learning for a binary classification, namely images having DR or No DR, whereas our model was used to classify all 5 stages of DR both for the APTOS and Mauritian dataset. In this paper, we have reported the use of several parameters to address overfitting of the models compared to the work of Gangwar et al. [22] and Challa et al. [13]. Finally, our model outperforms Thota et al. [21] and Masood et al. [11] in terms of accuracy.

4. Conclusions

In this work, transfer learning was applied at multiple levels with the aim of training multiple models to classify diabetic retinopathy for a completely blind dataset, the Mauritian cohort. At the initial stage, transfer learning was performed with three general pre-trained models, VGG16, ResNet50 and DenseNet169, using the APTOS dataset for diabetic retinopathy. Even after fine-tuning the three models, some classes were not being classified, and accuracies were not very high. This could be due to the dataset being highly imbalanced with almost 50% of the dataset belonging to “No DR” cases and the remaining 50% being distributed amongst the four DR classes. Hence, the dataset was augmented to achieve a comparable number of cases in each of the classes. Transfer learning was performed on the augmented APTOS dataset, and a better performance was achieved in the various experiments. It was found that the ResNet50 model produced equivalent or better results for all the classes compared to the VGG16 and DenseNet169 models. These trained enhanced models were then applied to the blind Mauritian dataset, and the results obtained are compared to the annotated local images. Again, the ResNet50, given its architecture, achieved the best results amongst the three models, and the accuracies obtained were very good. Class 0 achieved accuracies of 98%, 95% and 96% for the original APTOS dataset,

the augmented APTOS dataset and the Mauritian dataset, respectively, clearly indicating that the model is able to easily distinguish this class from the other classes, thus confirming the potential of training a precursor model for class 0 versus others. It was observed that some classes performed much better than others, and this needs to be further investigated. Classes 1, 2 and 3 achieved acceptable performances while class 4 was the most difficult to classify. The diabetic retinopathy expert observed that class 3 was graded more precisely. Moreover, retinal images with pronounced choroidal fronds seemed to be identified as class 4 by the software, which clinically rates as normal variants. This is an unexpected behaviour of class 4, representing a major difference between the training APTOS data and the Mauritius data. This can be solved by further transfer learning (or fine-tuning) from the APTOS-based model to a Mauritian-specific model.

In the future, more data, such as patient demographics, can be included to ensure clinical correlation. In addition, the Mauritian cohort can be analyzed to determine whether the data are demographically representative of the population and also the extent to which they are similar to those of the APTOS cohort. Our research shows the need for a precursor software to identify normal retinal images.

Author Contributions: Conceptualization, Z.M.-D., M.H.-M.K. and C.P.-R.; methodology, Z.M.-D., M.H.-M.K. and C.P.-R.; software, K.J., Z.M.-D., M.H.-M.K. and C.P.-R.; validation, B.N.B., N.Z.M.; formal analysis, K.J., Z.M.-D., M.H.-M.K., B.N.B., N.Z.M. and C.P.-R.; investigation, K.J., Z.M.-D., M.H.-M.K., B.N.B., N.Z.M. and C.P.-R.; resources, N.Z.M. and B.N.B.; data curation, K.J.; writing—original draft preparation, K.J., Z.M.-D., M.H.-M.K. and C.P.-R.; writing—review and editing, K.J., Z.M.-D., M.H.-M.K., B.N.B., N.Z.M. and C.P.-R.; funding acquisition, Z.M.-D., M.H.-M.K., N.Z.M. and C.P.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Higher Education Commission (HEC) under grant number T0714 and the H3ABioNet. H3ABioNet is supported by the National Institutes of Health Common Fund under grant number U41HG006941. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health and of the Higher Education Commission.

Institutional Review Board Statement: Ethical clearance to collect existing fundus images from local hospitals was obtained from the Ministry of Health and Wellness of Mauritius.

Informed Consent Statement: Informed consent was obtained from all subjects involved.

Data Availability Statement: Due to the confidentiality of the data, the dataset has not been made publicly available.

Acknowledgments: Authors acknowledge the support of the Ministry of Health and Wellness, the University of Mauritius, the H3ABioNet and Sherali Zeadally.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. GBD 2019 Blindness and Vision Impairment Collaborators; Vision Loss Expert Group of the Global Burden of Disease Study. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight: An analysis for the Global Burden of Disease Study. *Lancet Glob. Health* **2021**, *9*, E144–E160. [[CrossRef](#)]
2. Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A.A.; Ogurtsova, K.; et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* **2019**, *157*, 107843. [[CrossRef](#)] [[PubMed](#)]
3. Shah, A.R.; Gardner, T.W. Diabetic retinopathy: Research to clinical practice. *Clin. Diabetes Endocrinol.* **2017**, *3*, 9. [[CrossRef](#)] [[PubMed](#)]
4. Lam, C.; Yi, D.; Guo, M.; Lindsey, T. Automated Detection of Diabetic Retinopathy using Deep Learning. *AMIA Jt. Summits Transl. Sci. Proc.* **2018**, *2018*, 147–155.
5. Oltu, B.; Karaca, B.K.; Erdem, H.; Özgür, A. A systematic review of transfer learning based approaches for diabetic retinopathy detection. *arXiv* **2021**, arXiv:2105.13793.

6. Alyoubi, W.L.; Shalash, W.M.; Abulkhair, M.F. Diabetic retinopathy detection through deep learning techniques: A review. *Inform. Med. Unlocked* **2020**, *20*, 100377. [\[CrossRef\]](#)
7. Kandel, I.; Castelli, M. Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review. *Appl. Sci.* **2020**, *10*, 2021. [\[CrossRef\]](#)
8. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
9. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
10. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning transferable architectures for scalable image recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710.
11. Masood, S.; Luthra, T.; Sundriyal, H.; Ahmed, M. Identification of diabetic retinopathy in eye images using transfer learning. In Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 5–6 May 2017; pp. 1183–1187.
12. Li, X.; Pang, T.; Xiong, B.; Liu, W.; Liang, P.; Wang, T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 14–16 October 2017; pp. 1–11.
13. Challa, U.K.; Yellamraju, P.; Bhatt, J.S. A Multi-class Deep All-CNN for Detection of Diabetic Retinopathy Using Retinal Fundus Images. In *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17–20, 2019, Proceedings, Part I*; Deka, B., Maji, P., Mitra, S., Bhattacharyya, D.K., Bora, P.K., Pal, S.K., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11941, pp. 191–199.
14. Kaggle. APTOS 2019 Blindness Detection | Kaggle. Available online: <https://www.kaggle.com/c/aptos2019-blindness-detection/> (accessed on 15 February 2022).
15. Kassani, S.H.; Kassani, P.H.; Khazaeinezhad, R.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Diabetic retinopathy classification using a modified xception architecture. In Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, 10–12 December 2019; pp. 1–6.
16. Khalifa, N.E.M.; Loey, M.; Taha, M.H.N.; Mohamed, H.N.E.T. Deep transfer learning models for medical diabetic retinopathy detection. *Acta Inform. Med.* **2019**, *27*, 327–332. [\[CrossRef\]](#)
17. Hagos, M.T.; Kant, S. Transfer Learning based Detection of Diabetic Retinopathy from Small Dataset. *arXiv* **2019**, arXiv:1905.07203. [\[CrossRef\]](#)
18. Sikder, N.; Chowdhury, M.S.; Shamim Mohammad Arif, A.; Nahid, A.-A. Early blindness detection based on retinal images using ensemble learning. In Proceedings of the 2019 22nd International Conference on Computer and Information Technology (ICIT), Dhaka, Bangladesh, 18–20 December 2019; pp. 1–6.
19. Shaban, M.; Ogur, Z.; Mahmoud, A.; Switala, A.; Shalaby, A.; Abu Khalifeh, H.; Ghazal, M.; Fraiwan, L.; Giridharan, G.; Sandhu, H.; et al. A convolutional neural network for the screening and staging of diabetic retinopathy. *PLoS ONE* **2020**, *15*, e0233514. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Mushtaq, G.; Siddiqui, F. Detection of diabetic retinopathy using deep learning methodology. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1070*, 012049. [\[CrossRef\]](#)
21. Thota, N.B.; Umma Reddy, D. Improving the Accuracy of Diabetic Retinopathy Severity Classification with Transfer Learning. In Proceedings of the 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 1003–1006.
22. Gangwar, A.K.; Ravi, V. Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1*; Bhateja, V., Peng, S.-L., Satapathy, S.C., Zhang, Y.-D., Eds.; Advances in Intelligent Systems and Computing; Springer: Singapore, 2021; Volume 1176, pp. 679–689.
23. Dai, L.; Wu, L.; Li, H.; Cai, C.; Wu, Q.; Kong, H.; Liu, R.; Wang, X.; Hou, X.; Liu, Y.; et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **2021**, *12*, 3242. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Benson, J.; Maynard, J.; Zamora, G.; Carrillo, H.; Wigdahl, J.; Nemeth, S.; Barriga, S.; Estrada, T.; Soliz, P. Transfer learning for diabetic retinopathy. In *Medical Imaging 2018: Image Processing*; Angelini, E.D., Landman, B.A., Eds.; SPIE: Bellingham, WA, USA, 2018; p. 70.
25. Söderberg, S.; Zimmet, P.; Tuomilehto, J.; de Courten, M.; Dowse, G.K.; Chitson, P.; Gareeboo, H.; Alberti, K.G.M.M.; Shaw, J.E. Increasing prevalence of Type 2 diabetes mellitus in all ethnic groups in Mauritius. *Diabet. Med.* **2005**, *22*, 61–68. [\[CrossRef\]](#)
26. Housing and Population Census. Available online: <https://web.archive.org/web/20121114114018/http://www.gov.mu/portal/goc/cso/file/2011VoliIPC.pdf> (accessed on 15 February 2022).
27. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Cardoso, M.J., Arbel, T., Carneiro, G., Syeda-Mahmood, T., Tavares, J.M.R.S., Moradi, M., Bradley, A., Greenspan, H., Papa, J.P., Madabhushi, A., et al., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2017; Volume 10553, pp. 240–248.
28. Lopez-Nava, I.H.; Valentín-Coronado, L.M.; Garcia-Constantino, M.; Favela, J. Gait Activity Classification on Unbalanced Data from Inertial Sensors Using Shallow and Deep Learning. *Sensors* **2020**, *20*, 4756. [\[CrossRef\]](#)

29. Zhou, Y.; Wang, B.; He, X.; Cui, S.; Shao, L. DR-GAN: Conditional Generative Adversarial Network for Fine-Grained Lesion Synthesis on Diabetic Retinopathy Images. *IEEE J. Biomed. Health Inform.* **2020**, *26*, 56–66. [\[CrossRef\]](#)
30. Agustin, T.; Utami, E.; Fatta, H.A. Implementation of data augmentation to improve performance CNN method for detecting diabetic retinopathy. In Proceedings of the 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 November 2020; pp. 83–88.
31. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning—ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III*; Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 11141, pp. 270–279.
32. Da Rocha, D.A.; Ferreira, F.M.F.; Peixoto, Z.M.A. Diabetic retinopathy classification using VGG16 neural network. *Res. Biomed. Eng.* **2022**. [\[CrossRef\]](#)
33. Mule, N.; Thakare, A.; Kadam, A. Comparative analysis of various deep learning algorithms for diabetic retinopathy images. In *Health Informatics: A Computational Perspective in Healthcare*; Patgiri, R., Biswas, A., Roy, P., Eds.; Studies in Computational Intelligence; Springer: Singapore, 2021; Volume 932, pp. 97–106.
34. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. DenseNet Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
35. Taormina, V.; Cascio, D.; Abbene, L.; Raso, G. Performance of Fine-Tuning Convolutional Neural Networks for HEp-2 Image Classification. *Appl. Sci.* **2020**, *10*, 6940. [\[CrossRef\]](#)
36. Zhang, C.-L.; Luo, J.-H.; Wei, X.-S.; Wu, J. In defense of fully connected layers in visual representation transfer. In *Advances in Multimedia Information Processing—PCM 2017*; Zeng, B., Huang, Q., El Saddik, A., Li, H., Jiang, S., Fan, X., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2018; Volume 10736, pp. 807–817.
37. ElBedwehy, M.N.; Behery, G.M.; Elbarougy, R. Face recognition based on relative gradient magnitude strength. *Arab. J. Sci. Eng.* **2020**, *45*, 9925–9937. [\[CrossRef\]](#)
38. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [\[CrossRef\]](#)
39. Khan, A.H.; Cao, X.; Li, S.; Katsikis, V.N.; Liao, L. BAS-ADAM: An ADAM based approach to improve the performance of beetle antennae search optimizer. *IEEE/CAA J. Autom. Sinica* **2020**, *7*, 461–471. [\[CrossRef\]](#)
40. Keras Transfer Learning & Fine-Tuning. Available online: https://keras.io/guides/transfer_learning/ (accessed on 2 March 2022).
41. Peng, P.; Wang, J. How to fine-tune deep neural networks in few-shot learning? *arXiv* **2020**, arXiv:2012.00204.
42. Ismail, A. View of Improving Convolutional Neural Network (CNN) Architecture (MiniVGGNet) with Batch Normalization and Learning Rate Decay Factor for Image Classification. Available online: <https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/4558/2976> (accessed on 29 March 2022).
43. Usha Ruby, A. Binary cross entropy with deep learning technique for Image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 5393–5397. [\[CrossRef\]](#)
44. Song, H.; Kim, M.; Park, D.; Lee, J.-G. How does Early Stopping Help Generalization against Label Noise? *arXiv* **2019**, arXiv:1911.08059. [\[CrossRef\]](#)
45. Jayalakshmi, G.S.; Kumar, V.S. Performance analysis of Convolutional Neural Network (CNN) based Cancerous Skin Lesion Detection System. In Proceedings of the 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 21–23 February 2019; pp. 1–6.