# Diabetic Retinopathy Detection: A Deep Learning Approach to Identifying Early-Onset Diabetes


Frank Siyung Cho
*Queen's University*
20fsc@queensu.ca

Ola Elmaghraby
*Queen's University*
23lcd@queensu.ca

Ali Zidan
*Queen's University*
20az12@queensu.ca

Aimee Langevin
*Queen's University*
19al30@queensu.ca

Karim Ali
*Queen's University*
21kafm@queensu.ca



***Abstract*—This study addresses the pressing issue of Diabetic Retinopathy (DR), a leading cause of vision impairment among diabetic patients worldwide. With the prevalence of diabetes escalating, timely and accurate detection of DR is crucial. Deep Learning algorithms were utilized to predict early-onset DR across a dataset enhanced through various image processing techniques. Specifically, comparing the performances of a custom Convolutional Neural Network (CNN) model against pre-trained models InceptionV3 and ResNet-50. The methodology included experiments with different combinations of image enhancement techniques to assess their impact on model accuracy. The findings reveal that our custom CNN model which had been tailored to the nuances of DR detection, consistently outperformed the pre-trained models, especially when utilizing combined image enhancements. This research demonstrates the potential of using specialized machine learning models in improving diagnostic accuracy for DR, offering significant implications for early detection and intervention strategies.**


## I. Introduction

Diabetes is a systemic condition characterized by elevated blood sugar levels affecting more than 537 million adults globally. This value is predicted to rise to more than 783 million adults by 2045. [1]. It was responsible for 6.7 million deaths in 2021, costing healthcare expenditures of almost 966 billion USD in the last 15 years [1]. Thus, timely interventions are needed to control its associated symptoms [1]. Much of this burden falls on low-income and middle-income countries where 80% of people with diabetes reside, including 75+ million people in India and 140 million in China [2], [3]

Diabetic Retinopathy (DR), a condition affecting the eye, affects more than one third of adults over 40 with diabetes and is characterized by two major forms, non-proliferative (NPDR) and proliferative (PDR) [4]. Although both conditions can result to vision loss, NPDR is the early stage of DR, individuals may be asymptomatic or experience mild symptoms. Initial stages of NPDR include capillary microaneurysms (weakening of capillary walls), dot and blot hemorrhages (the escape of blood from blood vessels (BV)), soft and hard exudates (accumulations of fluid and cellular debris from BV) and swelling in the retina (responsible for receiving light and converting it into neural signals for visual recognition by the brain). NPDR can escalate into PDR, which is when inadequate blood supply to the retina is compensated for by the formation of new blood vessels which are fragile and prone to bleeding, potentially leading to vision loss [5].

Medical professionals can help slow down the progression of DR for patients by implementing interventions such as photocoagulation, intravitreal injections, and controlling blood-glucose levels and blood pressure [6].

### A. Problem Definition

Considering this, machine learning (ML) algorithms could enable early identification of patients at highest risk of vision loss, allowing timely referral to retina specialists. Countries with higher incidence have a smaller number of primary care physicians and retinal specialists than what is needed to tackle DR. This disparity can be seen in countries such as China where there are only 1.5 ophthalmologists per 50,000 citizens along with only 1.4 million family physicians for a population of over 1.4 billion [7]. Likewise, the United States, has approximately 18,000 ophthalmologists for their population of over 330 million [8]. It is imperative that the innovation of new strategies such as the use of Machine Learning algorithms bridge this gap to create accessible resources to supplement this disparity. Previously, the use of pretrained models (i.e. InceptionV3 and ResNet) has been employed for predicting DR. However, given that a preponderance of research articles within the last decade have focused on utilizing pre-trained models, building novel CNN architecture has varied in progress. This begs the question, can custom models provide better results than pre trained models. In this paper, we will compare both the use of pretrained models and custom models along with several image enhancement techniques in order to tailor the ML pipeline for predicting DR. Our Custom Model will look to increase the robustness of this concept across a diverse population to increase its generalizability, focus on optimizing aspects regarding computational infrastructure overlooked in other studies, and use an overlooked dataset to obtain novel results. Furthermore, the creation of a free and openly-accessible website that deploys the tested ML algorithms, will help bridge the gap in accessibility that currently exists between patients with limited accessibility to affordable healthcare.

## II. Related Work

Khalifa et al. [9] discusses different techniques and the current developments in image analysis for diabetic retinopathy fundus images. The approaches are organized by technique for imaging, image preprocessing, parasite detection

and cell segmentation, feature computation, and automatic cell classification. The paper investigates AlexNet, ResNet18, SqueezeNet. VGG16, VGG19, and GoogleNet. Similar to the previous paper, this research introduces pre-trained models that can be used and compared when building our own model. Specifically, the use of ResNet is helpful to implement and compare the pretrained model. The paper emphasizes AlexNet, which had a 97.9% total testing accuracy for the five classes combined (no DR, mild nonproliferative retinopathy, moderate nonproliferative retinopathy, severe nonproliferative retinopathy, proliferative diabetic retinopathy). AlexNet also had the least number of layers with only 8 layers, which led to fewer calculations and less complexity. The researchers also tested precision, recall, and F1 score and found that AlexNet achieved the best results for these metrics as well.

Das et al. [10] propose a deep learning model that contributes to deep feature extraction and image classification of Diabetic Retinopathy fundus images. The proposed model constitutes 26 pre-trained deep learning models, including ResNet50 and InceptionV3. The paper aims to determine the most suitable network for image classification. The dataset images are pre-processed for identification and differentiation of image details, which is implemented in a black-box style and cannot be visualized. The composition of models was trained on 27,446 fundus images and tested on 7680 images belonging to five classes (no DR, mild DR, moderate DR, severe DR, proliferative DR). Training was done on images of size 224x224 with a batch size of 32, the Adam optimizer, a learning rate of 0.0001, the categorical cross entropy loss function, and 50 epochs. They rejected the null hypothesis which claims to choose the best model and they accept their alternative hypothesis of finding and engineering a better model. This paper introduces many pre-trained models and assesses the accuracy of each one, which is useful to consider for our implementation of a pre-trained model. This has informed the decision to use ResNet50 and InceptionV3. EfficientNetB4 is highlighted for having a training accuracy of 99.37% and a validation accuracy of 79.11%, which is the best performance among the models.

## III. Methodology

### A. Dataset

In this study, retinal fundoscopy images were collected from 147 hospitals, between 2016 to 2018, covering 23 provinces in China [11]. The dataset(DDR) used was from researchers based out of Nankai University, China, led by Tao Li [11]. All images were collected from fundoscopy examinations, which is an examination that uses a magnifying lens and light to check the eye fundus, from 9598 patients. The images were originally categorized as either NDR, mild NPDR, moderate NPDR, severe NPDR, PDR, and ungradable. Given (a) the sample number of images used and (b) the objective of attaining generalized classifications between NDR, NPDR, and PDR, the images were divided into these classes as categorizations 0, 1, and 2, respectively. Initially, there were 13,666 images in the dataset, of which, 6263 were classified as NDR, 5341 as NPDR, 912 as PDR and 1,060 as ungradable. Figure 1, demonstrates examples for the three categorizations of images in the dataset used [12]. The initial study by the research team included extensive lesion segmentation and lesion detection [11]. Yet, the discussion and results of the paper suggest that this was only moderately successful as a foundation for more complex imaging annotations, hence, lesion segmentation was determined to be beyond the scope of the DR detection used in this paper [11] .

### B. Data Preprocessing

During data pre-processing, images were relabeled as 0, 1, and 2 for No DR, Non-Proliferative DR, and Proliferative DR. This combined the three instances of Non-Proliferative DR labels from the original labels: mild, moderate, and severe. It also discarded the ungradable images. This assumes that the model will perform better with less classes and that the model will still be applicable with the combined labels. Although the initial data was not balanced, after pre-processing and balancing the dataset for most of the CNN's trained, 579 image of each class was used. Data generators from TensorFlow were used to balance the data as well as implement data augmentation techniques. Therefore, it is assumed that the skew will not affect the results and the model will provide accurate predictions.

Images that were treated for DR were removed from the dataset because these images had additional circular features that were not indicators for DR, but the model may have mistaken them for symptoms such as hemorrhages, exudates, or microaneurysms. Other images were removed if they were corrupt or had intense brightness, as seen in Figure 1.

The images were cropped to remove any black background, resulting in square images that only contain the circular eye shape within the borders. This was done by removing all black rows and columns from each image. From there, the images were cropped to 512x512 pixels. The images were also normalized to have values between 0 and 1.

Before feeding the images into the CNN models, the images were augmented with horizontal and vertical flips. Finally, the images were split into training, validation, and testing data sets with 70%, 15%, and 15% respectively.

### C. Image Enhancement Techniques

Many approaches have been used in the literature on funduscopy images in order to reduce the noise and improve the contrast and quality of images to facilitate the detection and segmentation of inherit features. Three techniques were employed in this project. The integration of these techniques in our pipeline can be viewed in Figure 2.

*1) Isolation of Green Channel:* In funduscopy images, the green channel of the RGB images has the best contrast between the overall background and features such as exudates and blood vessels, making it easier to identify theses features [13], [14]. Therefore, The green channel was extracted to provide better quality and also a smaller image size by reducing the amount of channels in the image. This will also

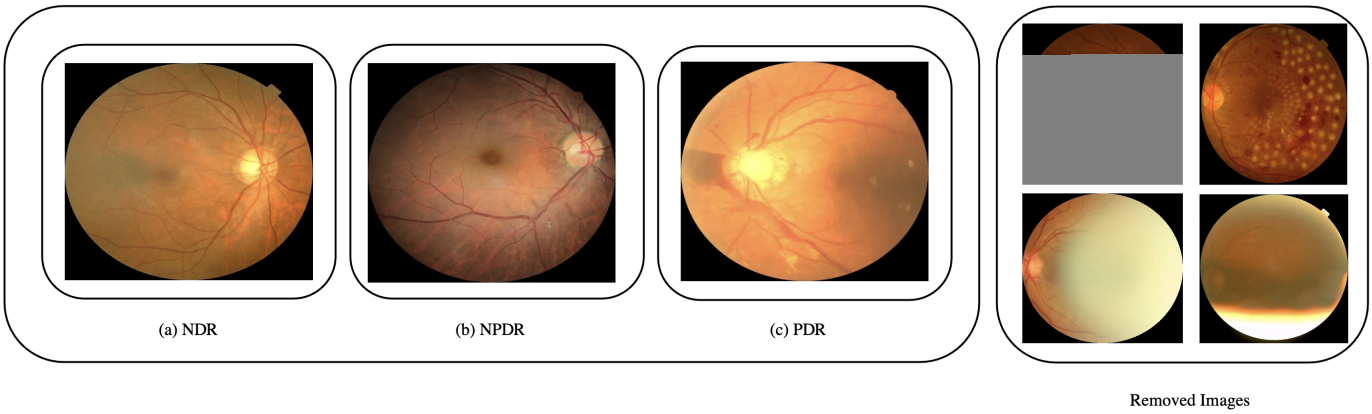


Fig. 1. Image pertaining to each class as well as ungradable images.

help facilitate the training speeds and reduce computational resources. Figure 3 shows a comparison of the 3 channels.

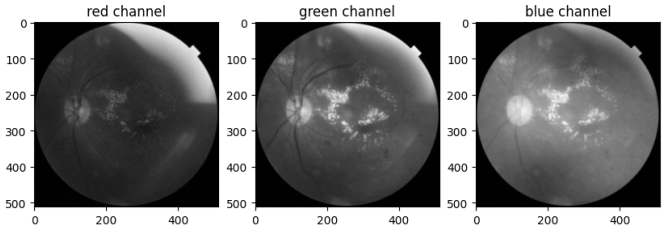


Fig. 3. comparison between the 3 color channels in a funduscopy image

*2) Median Filter:* is a nonlinear filter in which each pixel value in a kernel is computed as the median value of the pixels in the kernel which unify the values in each neighbourhood to reduce salt and pepper noise. The median filter is applied with a small kernel size of 3 to denoise the image and slightly blur it as advised in [13].

*3) Contrast Limited Adaptive Histogram Equalization (CLAHE):* The implementation of Contrast Limited Adaptive Histogram Equalization (CLAHE) stands out as a pivotal technique in enhancing the visual quality of retinal images. CLAHE, known for its ability to improve the contrast of images while preventing noise amplification, is particularly beneficial in medical imaging where the distinction of subtle features is crucial. By applying CLAHE, we aim to enhance the visibility of critical DR indicators such as microaneurysms, hemorrhages and exudates in retinal fundus images. This enhancement facilitates the ML model's ability to accurately classify the stages of DR, contributing significantly to the early detection and intervention efforts. The images were brightened before the contrast was adjusted using a Contrast Limited Adaptive Histogram Equalization (CLAHE) method. Insufficient or uneven luminance can obscure the details of the retinal imaging, so it is important to increase brightness while preserving the original colour of the pixels. Based on [15], the images were transformed to Hue Saturation Value (HSV) colour space to isolate and normalize (divide by 255) the value channel. Gamma correction was applied to the value channel with a gamma value of 1/2.2 based on the success of [15]. After the power transformation was applied, the value channel was denormalized (multiplied by 255) and converted back the to BGR colour space.

To further enhance the contrast, the CLAHE method was applied, which divides the images into small regions and equalizes the histogram on each tile [15]. Contrast limiting is applied which clips and uniformly distributes pixels if they are above the specified contrast limit [16]. The RGB image is converted to the L*a*b* space and the L* component is enhanced before converting back to RGB [15]. Using [15] as a starting point and experimenting with the given dataset, the tile size was 4x4 pixels and the clip limit was set to 3. The integration of CLAHE and their effects on images can be seen in Figure 4.

### *D. Proposed Models*

A comprehensive analysis was performed by employing an approach that involved the comparison of a Custom Convolutional Neural Network (CNN) model, alongside two pretrained models: InceptionV3 and ResNet-50. The selection of InceptionV3 and ResNet-50 as candidate models was informed by an extensive review of existing literature, which highlighted their efficacy and robust performance in similar applications.

The architecture of our custom CNN model can be seen in Figure 5 and was engineered to address the specific nuances of our dataset and classification task. Specifically, the model's structure encompasses four convolutional layers, each followed by batch normalization and max pooling layers, designed to sequentially extract and refine feature representations from the input data. Subsequently, the data undergoes

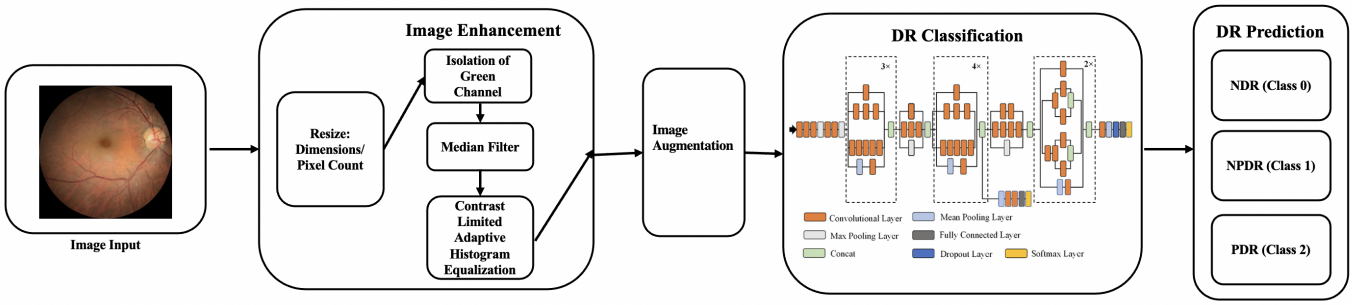


Fig. 2. Image Enhancement integration into pipeline.

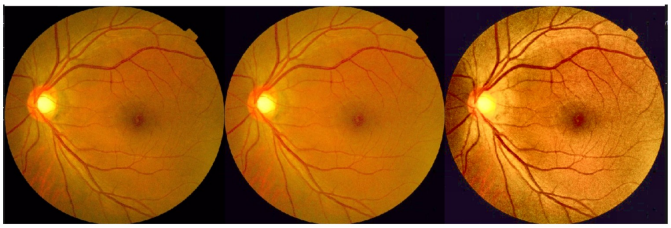


Fig. 4. a) Original image. b) Image enhanced with luminance only. c) Image enhanced with luminance and CLAHE.

dimensionality reduction through a flatten layer, preparing it for processing by a densely connected neural network. This network comprises two hidden layers, along with two dropout layers, to mitigate overfitting and enhance the model's generalization capabilities across unseen data.

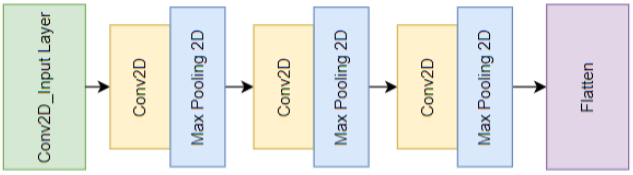


Fig. 5. Custom CNN Architecture.

InceptionV3 architecture shown in Figure 6, leverages multiple kernel sizes within the same layer to capture information at various scales, thereby enhancing its ability to recognize patterns with minimal computational resources. The model itself an evolution of the GoogleNet architecture was engineered by researchers from Google, and was originally trained on the ImageNet dataset comprised of over 14 million images of hand written digits. The Inception V3 architecture includes the use of asymmetric convolutions, factorization into smaller convolutions, and auxiliary classifiers to mitigate the vanishing gradient problem, thereby ensuring efficient training and high accuracy.

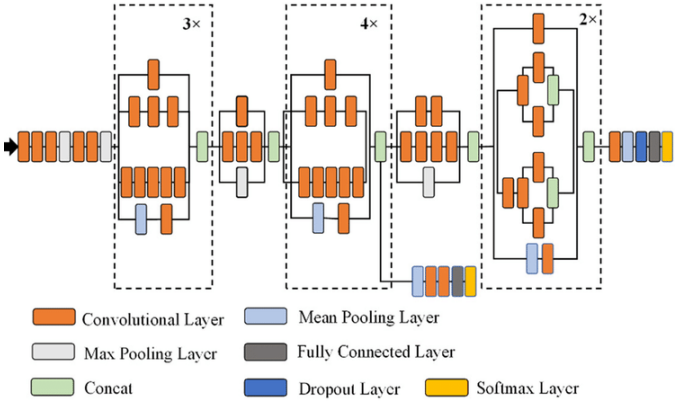


Fig. 6. InceptionV3 Architecture.

ResNet-50 architecture shown in Figure 7,implements the concept of residual learning to address the degradation problem inherent in deep networks. By incorporating residual blocks with skip connections that allow the flow of gradients directly through the network, ResNet-50 facilitates the training of substantially deeper models without a corresponding decline in performance. Similar to the InceptionV3 model, Resnet-50 was also trained on the ImageNet dataset. This architecture's ability to learn residual functions with reference to the layer inputs, as opposed to learning unreferenced functions, significantly enhances its learning capability and generalization.

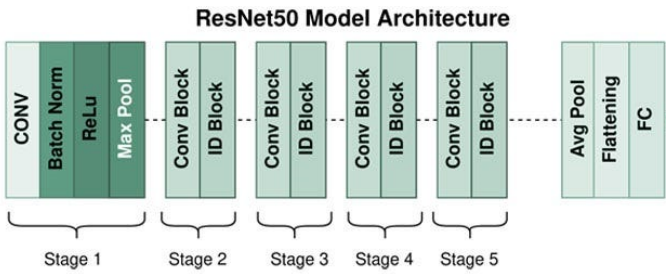


Fig. 7. ResNet-50 Architecture.

Throughout the strategic implementation of these models, our team aims to harness the distinct advantages each model offers, thereby optimizing performance and establishing a new benchmark in DR classification research.

### *E. Experiment setup*

After preprocessing the data and enhancing it, our 3 models InceptionV3, ResNet50, and custom CNN were trained on 6 experiments to test the effect of the different image enhancement techniques on a balanced dataset as follows:

1) combination of the 3 image enhancement techniques on 1737 grayscale images.
2) 1737 RGB images a sample of original images and 1737 enhanced images using median filter and CLHAE without the extraction of green channel. This experiment was done to compare the effectiveness of grayscale vs RGB images. 3474 images in total.
3) 1737 grayscale version of the original images and 1737 fully enhanced images. 3474 images in total.
4) 1737 RGB images with the median filter only
5) 1737 RGB images with the CLAHE only
6) 1737 grayscale images of the isolated green channel

The grid search technique is a systematic and organized way to choose the best values for the hyperparameters and facilitate tuning during training. That is performed by specifying a grid of hyperparameter values and then training the model multiple times, each time with a different combination of hyperparameters from the grid. Then the best model with the combination of hyperparameters that produces the best performance on a validation set is saved. The grid of parameters and values used in this project is shown in Table I. Overall, a total of 36 different parameters were tested along with three cross validation metrics, meaning that 108 total fits of the model were tested for in every test.

### *F. Evaluation Methods*

Choosing the evaluation metrics plays a critical role in judging if the deep learning model is working as intended or not. Depending on the task and sensitivity of the results different evaluation metrics are used. When it comes to multi-class classification the preferred metrics are:

TABLE I
HYPERPARAMETERS

| Parameter | Values |
|---|---|
| batch_size | 32<br>64 |
| epochs | 25<br>30<br>35 |
| optimizer | SGD<br>Adagrad<br>Adadelta<br>Adam<br>Adamax<br>Nadam |

*1) Accuracy:* gives an overall measure of the percentage of right prediction across all classes.

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

where:

$$TP : \text{True Positives}$$
$$TN : \text{True Negatives}$$
$$FP : \text{False Positives}$$
$$FN : \text{False Negatives}$$

*2) Precision:* percentage of how many instances of a certain class were classified correctly. for multi-class classification, the target class is considered positive and all other classes are negative. In this case, the precision is calculated for each class.

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

*3) Recall:* percentage of how many instances of some predicted class were actually from that class. for multi-class classification, the target class is considered positive and all other classes are negative. In this case, the recall is calculated for each class.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

*4) F1-score:* It is the harmonic mean of precision and recall.

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

In healthcare tasks such as this project where it is the worst case scenario to have PDR cases classified as NDR (false negatives are more concerning), it's crucial to prioritize recall over precision. This is because false negatives in this case can result in serious consequences for patients.

Confusion matrix is also used to facilitate evaluation by visualizing the results and assess what classes are confused together or if there is a bias.

## IV. Results and Discussion

### A. Quantitative Evaluation

The experimental analysis encompassed six designed tests to assess the models under varying conditions, as well as a systematic hyperparameter optimization using the grid search technique. These tests were partitioned into two categories: Tests 1-3 evaluated the collective impact of multiple image enhancement techniques, whereas Tests 4-6 focused on the efficacy of individual image enhancement strategies. The average performance metrics for Tests 4-6 are displayed below, offering insight into the distinct capabilities and limitations of each model in response to specific image enhancements.

TABLE II
Average performance metrics for the Custom Model across Tests 4-6.

| Test | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 4 | 0.65 | 0.59 | 0.58 | 0.60 |
| 5 | **0.67** | **0.63** | **0.65** | **0.65** |
| 6 | 0.55 | 0.49 | 0.46 | 0.49 |

TABLE III
Average performance metrics for the InceptionV3 Model across Tests 4-6.

| Test | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 4 | 0.46 | 0.35 | 0.26 | 0.36 |
| 5 | **0.57** | 0.46 | 0.47 | 0.47 |
| 6 | 0.54 | **0.61** | **0.61** | **0.61** |

TABLE IV
Average performance metrics for the ResNet-50 Model across Tests 4-6.

| Test | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 4 | 0.52 | 0.37 | 0.32 | 0.37 |
| 5 | **0.63** | **0.67** | **0.66** | **0.62** |
| 6 | 0.53 | 0.45 | 0.43 | 0.46 |

### B. Image Enhancement Performance and Discussion

The investigation into the individual impact of image enhancement techniques (Tests 4-6) revealed differing performances across the models. This differentiation can demonstrate the interaction between model architecture and specific image processing methods.

The Custom Model exhibited robustness in Tests 4 and 5, with a notable performance decrement in Test 6, suggesting its optimized calibration for CLAHE-enhanced imagery. This inference is supported by its peak accuracy of 0.65 in Test 5, highlighting the model's proficiency with CLAHE preprocessing.

Conversely, the InceptionV3 Model's performance was varying across tests, with its poorest results in Test 4. Its superior performance in Test 6, achieving a 0.61 accuracy, suggests an affinity for isolated green channel enhancement, potentially due to this technique's emphasis on structural and vascular features inherit to diabetic retinopathy images.

The ResNet-50 Model mirrored InceptionV3's variability, yet its optimal performance was displayed in Test 5 (CLAHE), similar to the Custom Model. This parallel demonstrates CLAHE's effectiveness in enhancing diagnostic features across different architectures.

A critical observation across the models is the suboptimal performance with median filtering (Test 4), indicating its potential to obfuscate rather than enhance diagnostic features within the retinal images.

### C. Comprehensive Analysis and Synthesis

Focusing on the integrated approach of Tests 1-3, the models' responses to combined image enhancements were evaluated, revealing a progressive improvement in performance metrics. This trend demonstrates the synergistic potential of employing multiple preprocessing strategies and image enhancement techniques to refine model accuracy and generalization.The average performance metrics for Tests 1-3 are displayed below.

TABLE V
Synthesized performance metrics for the Custom Model across Tests 1-3.

| Test | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 1 | 0.66 | 0.67 | 0.66 | 0.67 |
| 2 | 0.68 | 0.68 | 0.68 | 0.69 |
| 3 | **0.70** | **0.71** | 0.68 | **0.71** |

TABLE VI
Synthesized performance metrics for the Inception V3 Model across Tests 1-3.

| Test Num. | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 1 | **0.74** | 0.44 | 0.54 | **0.67** |
| 2 | 0.56 | 0.60 | **0.56** | 0.62 |
| 3 | 0.54 | **0.62** | 0.54 | 0.63 |

TABLE VII
Synthesized performance metrics for the Resnet-50 Model across Tests 1-3.

| Test Num. | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| 1 | **0.75** | 0.55 | 0.46 | 0.36 |
| 2 | 0.55 | 0.60 | 0.58 | 0.62 |
| 3 | 0.66 | **0.66** | **0.66** | **0.66** |

*1) Performance Evaluation of the Custom Model:* The Custom Model's linear improvement across Tests 1-3 exemplifies its adaptability to a differing preprocessing scheme. This robustness to both grayscale and RGB image inputs is particularly displayed in Test 3, where the integration of grayscale original images with fully enhanced images achieved the highest performing metrics. This outcome not

only demonstrates the model's flexibility but also suggests that leveraging a diverse mix of preprocessing techniques can enhance diagnostic accuracy. This is crucial for applications in medical imaging, where precision and recall balance is vital. In addition, this tendency can be seen through the consistent performances in F1 Scores across tests.

*2) InceptionV3 and ResNet-50: Variability and Optimization:* The InceptionV3 and ResNet-50 models exhibited variable responses, with specific enhancements yielding notable impacts on performance. This variability displays the need for tailored preprocessing pipelines, aligned with the architectural design of each model.

The InceptionV3 model's performance, while demonstrating a notable improvement in recall from Test 1 to Test 3, suggests a sensitivity to the combined preprocessing strategies. In particular, a sensitivity to color information could be infered. This is shown by a decline in performance metrics when transitioning from grayscale to RGB images in Test 2. This variability demonstrates the necessity of optimizing the preprocessing pipeline to attain the full potential of the InceptionV3 architecture, particularly in enhancing its precision without decreasing recall. The observed performance also suggests that optimizing the InceptionV3 model for specific diagnostic tasks may require a specific consideration of image enhancement techniques, particularly those affecting color and contrast.

The ResNet-50 model's performance demonstrates a significant improvement in accuracy from Test 1 to Test 3. This highlights the model's capacity for adaptation. There were notable struggles in Test 1, which utilized enhanced grayscale images. However, the model demonstrated an improvement in Tests 2 and 3, suggesting its potential effectiveness when exposed to a specific set of preprocessing techniques. However, the disparity in performance across the tests displays the critical need for a tailored approach to preprocessing. By aligning the image enhancement techniques with the inherent design principles of the ResNet-50 architecture, it is possible to maximize its diagnostic performance.

## V. Conclusion

In conclusion, these findings demonstrate a critical pathway towards optimizing CNN models for diabetic retinopathy classification through tailored image enhancement strategies. The findings advocate for a model-specific approach to preprocessing, emphasizing the importance of aligning image enhancements with the inherent architectural features and classification needs of each model. The Custom Model designed in this research paper was both consistent and outperformed those of pretrained variants; particularly under the use of combined image enhancements, this emphasizes the merit of these findings.

## VI. Ethical Considerations

Accessing retinal specialists for those in underserved communities may affect their ability to secure fundoscopy images that can be used to train and validate models such as the one in our study. This may affect the generalizability of our findings considering that they were likely to be from those with access to specialists in more affluent locales. Further, integrating this machine learning technology into existing healthcare frameworks may perpetuate existing healthcare inequities by improving resource accessibility to those who already have access to retinal specialists. On a larger scale, these technologies are also more likely to be integrated in developed countries with the disposable income and tangible assets to invest in these upstream interventions. This may differ from healthcare bodies in developing countries who may use the available funding on downstream interventions to address more time-sensitive issues.

## VII. Limitations

In the ophthalmology field, although fundoscopy images are considered more widespread in use, more detailed images from techniques such as optical coherence tomography are available which provide clearer indicators for DR characteristics which could have provided better input for our model. Further, in balancing the dataset, we had to significantly limit the number of images used for all three retinal image classifications, which reduced the images we used for both training and validation stages. With an inability to access the widely used EYEPACS dataset, this constraint may have affected our model performance and ability to learn distinctive features of retinal images This was perpetuated as the decision to remove treated DR images manually reduced our sample size and the generalizability of results to treated patients. Internally, our model's success was curtailed by long-wait times during dataset training which detracted from time that may have been used to improve the model parameters. Having tested 3 different models, this hurdle may have contributed to reduced accuracy results in our final model.

## VIII. Future Work

Acknowledging the limitations encountered in our research, future research should emphasize expanding the dataset to encompass a broader demographic spectrum, enhancing the model's generalizability. The exploration of advanced deep learning architectures and the integration of multimodal data, including genetic markers and patient histories, could provide a more comprehensive diagnostic framework. Furthermore, the incorporation of lesion segmentation techniques is paramount for the identification of NPDR indicators such as microaneurysms, hemorrhages, hard exudates, and cotton wool spots. Developing algorithms that can accurately segment these lesions from retinal images will enable the model to precisely pinpoint early features of DR. Training the model to recognise these segmented lesions as indicators of disease progression will be crucial in advancing the diagnostic capabilities of DR. This approach enhances the specificity of DR detection and contributes to more sophisticated models that can provide timely medical interventions.

## IX. User Interface

A user interface was also created with the final models created. Users are capable of inputting an image of a fundoscopy

and choose the model they would like to test. Various metrics about the results and model is then displayed along with the model's final prediction and probability values. Informational pages were also created in order to guide users and aid in any knowledge gaps that may be pertinent. A snippet of the website can be view in Figure Figure 8. The website can be accessed through www.retinaradar.com. **Disclaimer: The use of this website complies with all governing bodies; however, it should not be used as a substitute for professional medical advice, diagnosis, or treatment. We are not liable for any warranty or liability for your use of this information. Images uploaded are not saved and will not be collected.**

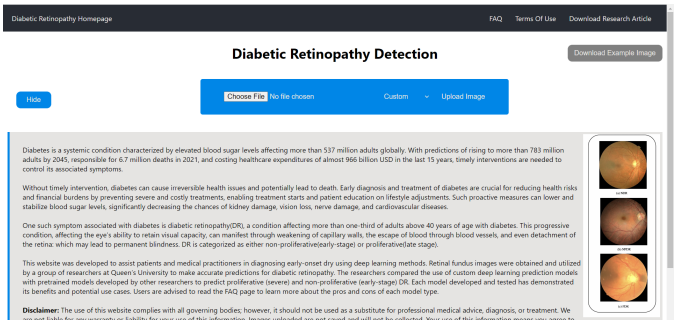


Fig. 8. Example of Website Frontend

## REFERENCES


[1] International Diabetes Federation, "IDF Diabetes Atlas - Diabetes around the world in 2021," 2021. [Online]. Available: https://diabetesatlas.org/. [Accessed: Feb. 27, 2024].

[2] The Lancet, "The Prevalence of Diabetes Mellitus: estimates for 2023," 2023. [Online]. Available: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(23)00381-9/fulltext. [Accessed: Feb. 27, 2024].

[3] Statista, "Diabetes in China - Statistics and Facts." [Online]. Available: https://www.statista.com/topics/6556/diabetes-in-china/#topicOverview. [Accessed on: Feb. 27, 2024].

[4] Centers for Disease Control and Prevention, "Vision Impairment and Blindness." [Online]. Available: https://www.cdc.gov/visionhealth/pdf/factsheet.pdf. [Accessed on: Feb. 27, 2024].

[5] Korean Diabetes Association, "Effects of Alcohol on Diabetes Mellitus," *Diabetes and Metabolism Journal*, vol. 42, no. 4, pp. 287-295. [Online]. Available: https://www.e-dmj.org/journal/view.php?doi=10.4093/dmj.2018.0182. [Accessed on: Feb. 27, 2024].

[6] A. Klein et al., "Brain responses to food-cue distractors in overweight adolescents: an fMRI study," *Nature Communications*, vol. 12, no. 1, p. 5549, 2021. https://doi.org/10.1038/s41467-021-23458-5.

[7] the Opthalmologist, "Accessibility and affordability of cataract surgery has improved in China," 2024. [Online]. Available: https://theophthalmologist.com/subspecialties/the-china-study. [Accessed on: Feb. 27, 2024].

[8] S. Dang, H. Pakchanian, E. Flynn, and R. Raiker, "Estimating Patient Demand for Ophthalmologists in the United States using Google Trends," *Investigative Ophthalmology and Visual Science, 2021.*

[9] N. E. M. Khalifa, M. Loey, M. H. N. Taha, and H. N. E. T. Mohamed, "Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection," *Acta Informatica Medica*, pp. 327–332, 2019. https://doi.org/10.5455/aim.2019.27.327-332.

[10] D. Das, S. K. Biswas, and S. Bandyopadhyay, "Detection of Diabetic Retinopathy using Convolutional Neural Networks for Feature Extraction and Classification (DRFEC)," *Multimedia Tools and Applications*, 2022. https://doi.org/10.1007/s11042-022-14165-4.

[11] T. Li, Y. Gao, S. Guo, and H. Liu, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, 2019. https://doi.org/10.1016/j.ins.2019.06.011.

[12] M. W. Nadeem, H. G. Goh, M. Hussain, S.-Y. Liew, I. Andonovic, and M. A. Khan, "Deep Learning for Diabetic Retinopathy Analysis: A Review, Research Challenges, and Future Directions," *Sensors*, vol. 22, no. 18, Article 6780, 2022. https://www.mdpi.com/1424-8220/22/18/6780.

[13] M. Monemian and H. Rabbani, "Exudate identification in retinal fundus images using precise textural verifications," *Scientific Reports*, vol. 13, no. 1, p. 2824, 2023. https://doi.org/10.1038/s41598-023-29916-y.

[14] R. Sarki et al., "Image Preprocessing in Classification and Identification of Diabetic Eye Diseases," *Data Science and Engineering*, vol. 6, no. 4, pp. 455–471, 2021. https://doi.org/10.1007/s41019-021-00167-z.

[15] M. Zhou et al., "Color Retinal Image Enhancement Based on Luminosity and Contrast Adjustment," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 521–527, 2018. https://doi.org/10.1109/TBME.2017.2700627.

[16] OpenCV, "Histograms - 2: Histogram Equalization." [Online]. Available: https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html. [Accessed: Feb. 28, 2024].