Optimized Deep Learning Model For Diabetic Retinopathy Screening

MSc Research Project
Data Analytics

Kaushik Kalyanaraman
Student ID: 21122211

School of Computing
National College of Ireland

Supervisor: Aaloka Anant

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Kaushik Kalyanaraman
Student ID:	21122211
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Aaloka Anant
Submission Due Date:	15/12/2022
Project Title:	Optimized Deep Learning Model For Diabetic Retinopathy Screening
Word Count:	5977
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th December 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimized Deep Learning Model For Diabetic Retinopathy Screening

Kaushik Kalyanaraman
21122211

Abstract

Diabetic retinopathy(DR) is a disorder caused by uncontrolled diabetes over an extended period of time. It is the most serious eye problem triggered by diabetes, affecting more than one-third of all diabetics. Even with the attempts of governments to expand screening, a dearth of skilled individuals for screening has made early detection even more challenging. This results in delayed treatment of this progressive disease which might lead to loss of vision, hence necessitating the use of an automated screening technique. This research seeks to implement an approach for screening using a custom optimized Convolutional Neural Network (CNN) for a binary classification of DR and No DR using retinal fundus images. This suggested implementation aims to provide an efficient model that may be used as an automated tool to screen diabetic retinopathy, allowing for early diagnosis and treatment of afflicted individuals. Sensitivity, specificity, F1 score and accuracy have been used as the evaluation metrics to measure the performance of the model. By using the same images, an analysis is done between the implemented Logistic Regression, ResNet 50 and custom optimized CNN is done where the custom optimized CNN exceeds the performance of the other two models on all metrics with a sensitivity of 92.9%, specificity of 93.4% and a test accuracy of 93.2%.

1 Introduction

Type II Diabetes, also known as Diabetes Mellitus, is a chronic disease that affects people of all ages worldwide. It is influenced by the human body's inability to produce enough insulin, which raises blood sugar levels. As a patient battles diabetes for a long time, blood vessels are often damaged, and when this affects the retina and/or optic nerves around the eye, the retina impact condition is commonly known as diabetic retinopathy. This is still one of the most common causes of vision loss, and doctors believe it is critical to diagnose it at the onset or as soon as possible to avoid large-scale consequences contributing to vision loss.

Primitive diagnosis is critical in detecting DR, which is aided by technological advancements in today's world. However, physical tests such as acuity testing, pupil dilation, optical coherence topography, and so on are time consuming and can harm the eye. Furthermore, manual checks are prone to human error, particularly in the early stages of DR. As a result, artificial intelligence or machine learning comes into play in this domain, as the algorithm can play an important role with both expediting diagnosis and providing accurate reports on the condition of damage, thereby ensuring that the eye is treated in the early stages.The algorithm must be trained to interpret a significant number of retinal fundus images in order to accurately diagnose even minor changes that we, despite our vast expertise in the field, frequently miss. The development of an innovative custom optimized CNN is covered in this study.

1.1 Background & Motivation

Diabetes has an harmful effect on an individual's body over a period of time and affects vital organs of the body like eyes, nerves, heart and kidneys. While the effect on other organs can be understood by person by looking out for the easily noticeable symptoms, the effect it has on eyes may not be understandable by the individual. Hence it is necessary for diabetic patients to get their eyes tested at regular intervals so that these scans can indicate any deterioration of the eye. Since the number of people having diabetes is increasing exponentially (Singh Gautam et al.; 2019), there is an increasing need to have automated systems which can help with the prediction so that the workload remains manageable and the patients would not need to wait for longer periods for follow-ups.

These implications have encouraged the author to implement the optimized innovative CNN model which will be able to detect DR efficiently. Even though marketing of AI has been permitted for detection of diabetic retinopathy[1], the model would need to be strongly evaluated for its metrics of sensitivity and specificity along with others so that the robustness of the model can be checked. The following sections of the report include an assessment of existing work in the form of a literature review, methodology, implementation, evaluation using metrics and discussion, all of which will help in the attempt to answer the research question posed below.

1.2 Research Question & Objectives

1.2.1 Research Question

To what extent can the implemented custom optimized CNN improve the evaluation metrics in the detection of diabetic retinopathy with fundus images?

1.2.2 Research Objectives

- Review and analysis of existing work done in the field of diabetic retinopathy
- Collection of fundus images, preprocessing and transformation of these images as part of data preparation of model.
- Implementation of custom optimized CNN, ResNet 50 and Logistic Regression models for DR detection.
- Evaluation of implemented custom optimized CNN, ResNet 50 and Logistic Regression models.
- Discussion and comparison of implemented models.

[1]US FDA: `https://rb.gy/fraud1`

2 Related Work

The health sector has continued to improve its research with the advent of newer methods in the area of supervised and unsupervised learning techniques. In the health sector, the model development process must be reliable and robust enough to reduce the possibility of any errors. In this section, the author attempts to provide an analysis of the past work in DR screening using machine learning, deep and transfer learning techniques in subsection 2.1 and subsection 2.2.

2.1 Diabetic Retinopathy Screening Using Machine Learning Approaches

Preprocessing plays an important role in the model building and needs to be carried out according to the dataset and the requirements. (Maneerat et al.; 2020) have used an approach of k-means to classify DR and Non-DR images. The dataset has been prepared by normalizing the images, adjusting the contrast and background color after which k-means algorithm has been applied to this transformed dataset. The research has chosen sensitivity and specificity as the evaluation metrics along with the accuracy which seems valid due to the importance of understanding the results of the model. But with the number of images in the dataset fairly low, the achieved accuracy of 73% might not give the ideal picture. (Kubde and Mohod; 2021) have similarly attempted to completely transform the images by by applying contrast to the images in the dataset, removing the optical disc and detecting the exudates from the images after which they have applied the Support Vector Machine(SVM) and Naive-Bayes algorithm to classify the images accordingly. With the maximum accuracy of 53% achieved by SVM classifier, it might indicate that the model would need to be tuned and optimized further to achieve better results for this dataset using the above transformation processes.

The use of ensemble methods in machine learning has vastly improved the model building process and has helped in gaining better results by helping reduce the variance. (Reddy et al.; 2020) have attempted to use Random Forest, Decision Tree, Logistic Regression, Adaboost and KNN as a part of the proposed ensemble method with the use of grid search for hyperparameter tuning of the models which has enabled them to further enhance their model. While this model has achieved an accuracy of 81%, the use of a larger dataset and experimentation with different parameter tuning techniques might help get an accurate model that can be used. Using the information gain attribute and wrapper subset methods for feature selection, (Odeh et al.; 2021) have used the combination of Random Forest, neural network and SVM as an ensemble to train the model for DR classification. The use of the original dataset has achieved the highest accuracy of 75.1% over the other attempts using the feature selection methods, which indicate that the feature selection techniques have either been not utilized properly to attain the best features of the images, or that these techniques might not have been suited for this particular dataset.

The use of machine and deep learning will only be successful in the real world if the developer takes into consideration each and every scenario. (Paradisa et al.; 2020) have made the use of only a thousand images after augmentation which consists of DR and non-DR images which is further split into the required train and test sets. This has then been used with different algorithms which has resulted in high accuracy for some of the models. This is an example of a model which cannot be used practically as it has not been tested with the different scenarios that might be available in the case of diabetic retinopathy.

2.2 Diabetic Retinopathy Screening Using Deep & Transfer Learning Approaches

The introduction of transfer learning models have provided the developers to apply a readily optimized model on their dataset. (Chen and Chang; 2022) have preprocessed and transformed the dataset by resizing and augmentation after which they have made use of InceptionV3 and EfficientNet models whilst adjusting the learning and dropout rate of the models to fine-tune them. Amongst the different experiments conducted with the fine-tuning of the models, EfficientNet B0 has achieved the highest accuracy of 86.3% which can be further increased by experimenting the more pre-processing and transformation techniques which will help the model get the best possible images as input which will help in the efficient training of the model. (Ebin and Ranjana; 2020) have utilized VGG16 and InceptionV3 to classify the DR screening images. In this dataset, it can be seen that there is a high imbalance in the number of images the each category which has not been rectified properly. This has resulted in a poor model performance for both models. It is highly essential to maintain the balance in the dataset so that the distribution helps the model to train and predict correct classes accordingly. This kind of issue has been covered by (Kumari et al.; 2022) in their research wherein they have analyzed the dataset post which augmentation has been done in a manner that all the classes have the same number of images. Their research also includes preprocessing by cropping the images to remove noise and segmentation after which this transformed data has been given as input to different transfer learning models. The performance of the models could have been better if the research had included more training time for better understanding of the images by the respective models.

(R et al.; 2022) have made use of a modified InceptionV3 in their attempt to implement DR classification and have included other models like CNN, DenseNet 121, VGG 16 and ResNet 50 to understand the performance of their model with other models. It can be seen that their model outperforms other models and has an accuracy of 80%. Their results might get better with more tuning of their proposed transfer learning model. By modifying ResNet 50, (Elswah et al.; 2020) have attempted using it along with a feed-forward neural network for classification purposes. To add to this, another ResNet 50 model has been added to improve the efficiency of the DR classification. The maximum result achieved is 86.67% which might further improve by taking a higher ratio of training images as compared to the 70:30 split in the implemented model. (Nurmalasari et al.; 2022), on the other hand, have implemented multiple transfer learning CNN based models like AlexNet, DenseNet 121, InceptionV3, ResNet 50, VGG 16 and Xception. The comparative study includes implementing these models with and without Stochastic Gradient Descent, RMSprop and Adam optimization techniques. The Adam optimization for VGG16 achieves the best result with 77% accuracy. The low accuracy might be due to the uneven distribution of images of different classes.

(Qian et al.; 2021) have dealt with a highly imbalanced dataset for which they have included data augmentation as part of the data preparation techniques along with which they have normalized, resized and applied contrast to the images. These transformed images are then used a combination of DenseNet and ResNet which will aid in extracting the features better. They have also utilized the attention mechanism due to which the learning capability of the model will increase significantly. They have fine-tuned the learning rate and have trained the model using the Adam optimizer along with dropout rates. This model has an accuracy of 83.2%. They have converted it into a comparative study by implementing two other research work as well and comparing the efficiency. While their model seems to have performed better than the other two models, this result can be increased with the help of further optimization of the model used. (Guo et al.; 2022) have attempted to fine-tune their ResNet 34 model developed for DR classification. They have followed prior data preprocessing and transformation by resizing, balancing and normalizing the dataset. They have further attempted to tune the model by freezing the layers and using them once the training accuracy reaches a certain point. While this technique sometimes helps in improving the performance of the model, factors like bias and weights could also have been tuned in order to improve the overall structure of the network.

2.3 Conclusion

The above analysis highlights the critical modelling components that were overlooked, resulting in disparate sets of results. When dealing with a critical healthcare application, one must not only strive for greater accuracy, but also meticulously follow each and every process of data preparation and modelling to avoid errors caused by negligence. As we are dealing with patients who will be administered treatment based on the results of the implemented model, the data must be carefully preprocessed to enhance the features while avoiding information loss, and the models that may be chosen for implementation must be carefully evaluated and tuned to achieve the best possible evaluation metric scores. It can also be inferred that a dataset containing less data will possibly result in a skewed performance from the model and hence would be advisable to use a large enough dataset for proper training and evaluation of the model. The next section discusses the methodology followed for the implementation.

3 Methodology

The use of prediction systems for health related issues must be approached taking all the necessary steps and precautions due highly sensitive task that the model will be performing. In case of DR detection, false negatives are a prime area of concern. The models analyzed in the above section do not possess a high accuracy due to which their models might predict multiple false negatives and this process will hinder the timely treatment process of the patient. This research attempts to build a model considering the issues with the models used and the requirements in the DR detection that might need to be addressed. For this purpose, this research utilizes KDD methodology (Fayyad et al.; 1996) which will provide a step-wise framework for achieving the research objectives.This involves data collection, data preprocessing, data transformation, data modeling and evaluation metrics phases which when followed uses 9600 images for modeling phase after transformation phase completion. This section will discuss the applied methodology in detail.

3.1 Data Collection

There are numerous sources online which have fundus images available for use. But, it is of utmost importance to use data from verified sources so that the data used for the research will be authentic. For this purpose, the fundus images are obtained from a verified repository in Kaggle[2]. This repository contains 3662 colored fundus images of five distinct categories of no DR, mild, moderate, severe, proliferate and as illustrated in Figure 1. This is a public dataset and has been licensed to use for any purposes.

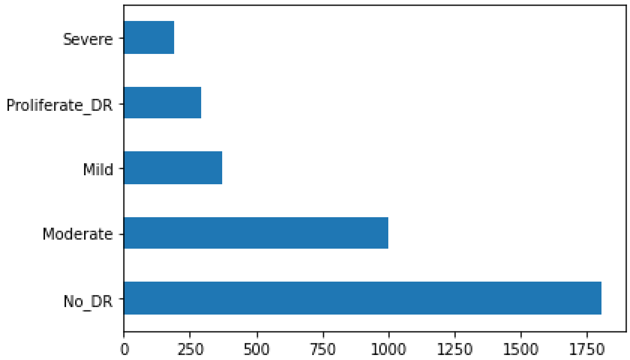

Figure 1: Data Collection

The labels for each of these figures are available in a csv file which contains the image filenames and the associated labels from 0 to 4 respectively.

3.2 Data Preprocessing

This step is necessary to understand the data. The existing images in the repository are loaded and mapped according to their labels. Another column is added which contains binary labels that will classify the images into two categories namely DR and No DR.

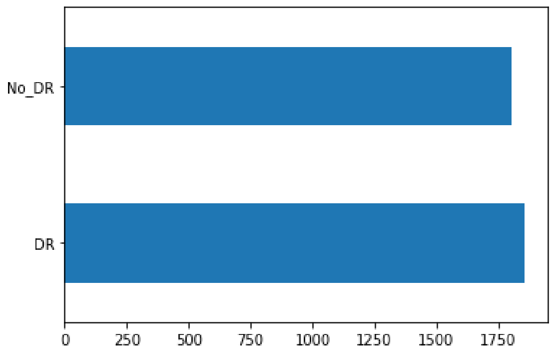

Figure 2: Preprocessing Phase

This will not only in achieving the aim of classifying the images into DR or No DR category but will help to solve the imbalance in the different classes that can be observed in from Figure 2. For this purpose, the images with labels that had the binary label DR

[2]Dataset: `https://rb.gy/cvauju`

were merged together to form a single DR folder. No filters are applied to the images so that the model can be built in such a way that it should be able to detect images even with a little noise. This will enable the model to be able to test the images that will be provided directly from the scan prints.

3.3 Data Transformation

Once the images have been labeled into DR and No DR categories, the images in the different classes are moved into these two folders according to their labels. The ability of deep learning can be truly harnessed only if there is adequate data for the model so that it will be able to train itself properly. The technique of augmentation is used in this scenario (Alam et al.; 2021) through which the total number of images is increased to 9600 which have equal number of images in both the categories using a variety of techniques like width shift, height shift, horizontal flip and zoom factor. These variations in the images and their sizes will provide more learning data for the models to learn. Sample images after augmentation are as illustrated in Figure 3.

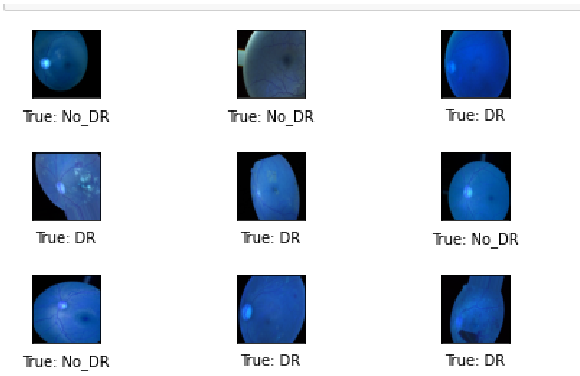

Figure 3: Transformation Phase

3.4 Data Modeling

Once the data is ready to be used, the modeling result should be such that it should be able to classify the fundus images with a high accuracy. From section 2, it can be inferred that while the number of layers might improve the overall performance of the model, it might not be a valid solution until all the parameters of the existing layers have been optimized and tuned to the best level possible. Hence, for this purpose, this research will implement a customized CNN with as few layers as possible which are optimized at each and every level to extract the maximum performance of the model and so that a robust and efficient model is developed in this process. Along with this model, this research implements ResNet 50 and Logistic Regression models which are again tuned so that these models can be compared for their strengths, weaknesses and their overall efficiency in predicting the fundus images.

3.5 Model Evaluation

Evaluation of models will help in understanding the true ability of the model in its process of achieving the desired outcome. In health sector, it is absolutely essential to use the metrics of sensitivity and specificity (Moccia et al.; 2018) as they truly determine the model's ability to predict the disease. Apart from the above, other metrics like confusion matrix, F1 score will also be used to evaluate the models which will reflect the model's balanced capacity to both identify and accurately anticipate positive cases.

4 Design Specification

4.1 Project Flow Design

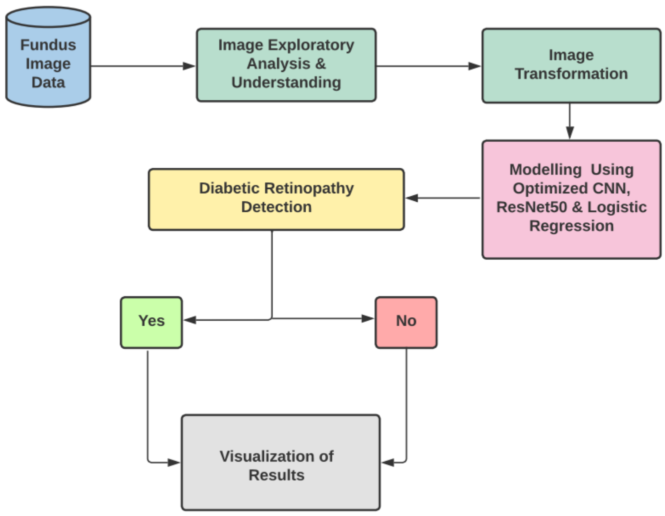

Figure 4: Project Flow Design

Figure 4 depicts the overall flow of the project. The steps of fundus image data collection, data understanding and preprocessing and transformation are discussed in the above section 3.Modeling using custom optimized CNN, ResNet 50 and Logistic Regression and their model evaluations are discussed in the below section 5.

4.2 Custom Optimized CNN Architecture Design

The introduction of CNN has highly improved the ability of models to detect or predict the images and provide the required outputs after processing the images through its layers. The use of multiple layers in a CNN has been seen as a way to improve the performance of the models but with increasing layers, issues like vanishing gradient come up which often result in decrease in the model performance. Hence it is essential to explore CNN completely along with its parameters and tune the existing layers before adding further layers. (Ouyang et al.; 2019), in their research of SAR Image Ground Object Recognition Detection, have advocated the need to tune the layers properly by using the parameters of weight, activation functions and other parameters which might enhance the efficiency of the model. Figure 5 illustrates the custom optimized design of the CNN architecture

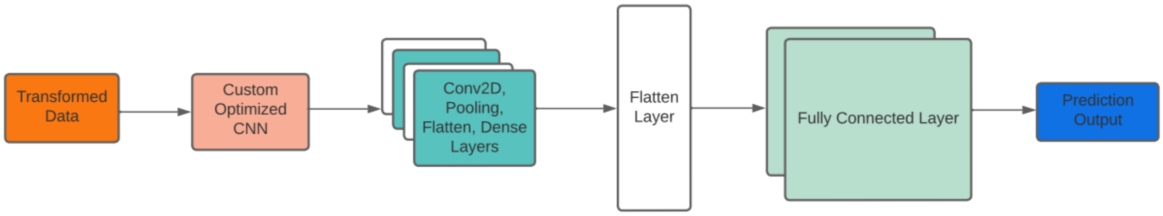

Figure 5: Custom Optimized CNN Architectural Design

whose implementation is described in the below section. This CNN has been built with a minimal 3 convolutional layers, whose optimized parameters of weight and bias have been used in each of the layer with its values changing dynamically according to the input received from the previous layer. Apart form these, the other parameters used have also been tuned according to the efficiency exhibited by the model's performance during the multiple runs leading to the final design. These are discussed in detail in the below section 5.

5 Implementation

This research is being run on a system with a 3.2 GHz quad-core Intel Core i5 processor, 16 GB of 3200 MHz DDR4 RAM, and a 2 GB Nvidia GeForce MX450 graphics card with 512 GB SSD storage. This system configuration is suitable for this research due to the high computational power that is required for image processing and for running CNN, ResNet 50 and Logisitic Regression models considering the complexity involved in processing the images with the above architectures. Anaconda has been used as the tool for running the code with a separate environment setup for the project with required libraries so that there is no issue of other conflicting libraries in the environment. In this section, the custom made optimized CNN, ResNet 50 and Logistic Regression implementations and evaluations will be discussed in detail.

5.1 Implementation of Designed CNN Architecture

In this CNN architecture, the preprocessed and transformed fundus images which are stored in train and test folders according to the labels are read and stored with image size to be of 128 pixels. The validation set for the model is taken for 0.2 ratio of the training set due to its large enough size. Based on the understanding of existing research and knowledge of the model, the complete architecture is fine tuned at each and every step of the model and has been run multiple times with different tuning parameters with the model returning the best possible result by to check the efficiency post which a final model has been created and is described below.

- **Input Shape:** The shape of the input data remains same throughout as (128,128,3) with three as the number of channels as the image is colored. This shape has been finalized after checking the effects of different images dimensions on the performance of the model.

- **Convolutional Layer:** The CNN architecture has different elements in the convolutional layer which when tuned correctly can give better results. In this implementation, there are a total of 3 convolutional layers. Since it is a colored image, 3 channels have been used. The weights at first are calculated according to the shape and are not randomly distributed but in the 95% confidence interval of the standard deviation. In the subsequent 2 layers, the weights are calculated dynamically based on the input provided by the previous layer. The biases are also calculated during the model run-time in a similar fashion. The first two layers have 32 filters with each filter of size 3 and the last layer has 64 filters also of the size 3. The stride value for the model is set at 1 so that the image is thoroughly analyzed with one pixel movement at a given time.

- **Pooling Layer:** The use of max pooling is done with an aim to reduce the overfitting in any case and will additionally help in reducing the overall time taken for the model run. Here, the stride size of 2 is taken so that the most significant characteristics of the image are recorded.

- **Rectified Linear Unit Activation Function(ReLU):** To enhance the image understanding capacity of the model, ReLU is being used. ReLU is also being used to keep the resultant output to non-zero values so that the existing negative values are converted to zero and hence their impact is voided. Also, early levels will benefit from picking up crucial knowledge when the network is back propagating.

- **Flatten Layer:** This layer is used to convert the above output of n-dimensional array to a flat array. This flattened array will serve as the input to the next layer.

- **Fully Connected Layer:** There are two fully connected layers at the end which will help in getting the prediction as DR or No DR. 128 neurons have been used and ReLU is used in the first fully connected layer to remove discrepancies. In both the layers, weights are again calculated dynamically based on the input that is given to it and will multiply the input layer by the weights to calculate each layer as a matrix.

- **Adam Optimizer:** This optimizer has been used due to its ability to handle sparse gradients on images with noisy data.

With the above layers with different tuned parameters and the cross entropy function to help remove the minute errors and the early stopping feature so that the model would be able to stop the process if the validation scores were not improving. The model has been tested with multiple runs having different parameters of batch size like 32, 64 and 128, the different optimizer functions, input image sizes and calculation strategies for weights and biases. The above final model proved to have the most impact on getting the good results for the custom optimized CNN model which is evaluated in the below sub-section.

5.2 Evaluation of Designed CNN Architecture

The selection of the final model design was done after the model was run numerous times with different set of model parameters while monitoring the evaluation metrics of each run of the model. For this purpose, the confusion matrix, validation accuracy, training accuracy and testing accuracy have been calculated in between intervals to keep track of

the model performance. The inclusion of the early stopping parameter will help in getting best possible result of the model as it will exit when the model score is not improving. After the first epoch, the training, validation and test accuracy stand at 62.5%, 40.6% and 51.4% respectively with the validation loss at 0.738. It can be seen that the the validation loss has gradually decreased to 0.386 when it exits the model at 53 epochs.
Figure 6 depicts the confusion matrix of the model.

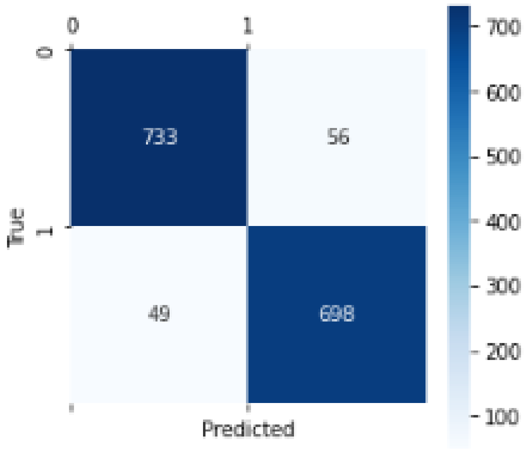

Figure 6: CNN Confusion Matrix

The final training accuracy is 96.9%, and the accuracy for validation and test sets are 90.6% and 93.2% respectively. The sensitivity of the model is scored at 92.9% while the specificity is 93.4%. The model achieves an F1 score of 0.93. The F1 score suggests that the model is highly dependable in predicting both DR and No DR cases.

5.3 Implementation of ResNet 50 Architecture

CNN architectures have been used as a base to create multiple pre-trained models. ResNet 50 is sophisticated residual network model which has the ability to train the deeper neural networks. This variant of the residual network model contain 50 layers and uses the concept of skip connections which helps with the removal of the vanishing gradient issues that arise in a model. This transfer learning model does not require much modifications. For this implementation, the ResNet 50 model will take the input as the fundus images of shape(64,64,3) and a batch size of 64. Along with these parameters, max pooling layer is used. As the model uses pre-trained weights, it is not necessary to train all the layers. Also, for the classification purposes of DR and No DR, a fully connected layer is used instead of the last layer. The use of residual networks has achieved good classification scores as discussed in the related work. The model has been tuned for different parameters before the above parameters were finalized due to better performance. (Ying; 2022), have used a similar technique of implementing ResNet with modification of using BCNN as well in an attempt to enhance the model, they have achieved an accuracy of 71.11%. The evaluation of this implemented model is discussed in the next subsection.

5.4 Evaluation of Implemented ResNet 50 Architecture

This transfer learning model implemented with above parameters was run multiple times for different epochs. The same number of images are used for this implementation as CNN so that the models can be analyzed for their abilities with each other. The runs which had larger number of epochs resulted in decrease of overall accuracy and performance of the model. While the model starts fluctuating after 4 epochs, it was observed that the model only gains till 10 epochs before its accuracy drops again. Figure 7 depicts the model loss and the accuracy graphs for train and test runs. At the end of the run,

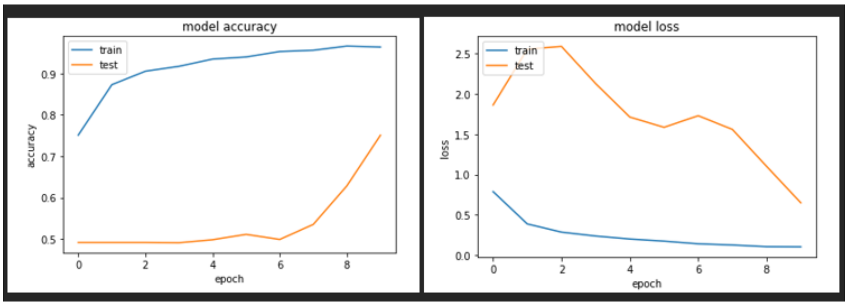

Figure 7: Loss & Accuracy Graph

the training accuracy of the model is 96.37%, while the validation and test accuracy are 75.07% and 77.19% respectively are lower when compared to the training accuracy. The scores of sensitivity, specificity and F1 are computed to be 69.26%, 85.6% and 0.75 respectively. These low scores can be understood with the confusion matrix represented by Figure 8 which depicts the actual and predicted numbers of the model. The specificity

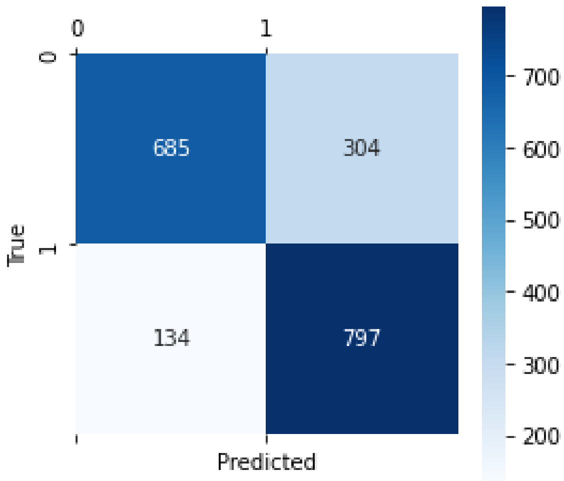

Figure 8: ResNet 50 Confusion Matrix

score indicates that the model performs well enough to correctly predict No DR cases up to an extent.

5.5 Implementation of Logistic Regression

Machine learning algorithms often excel with text data. In case of image processing, the machine learning algorithms need meticulous planning for input and the parameters using which the model is trained. In this design for the implementation, the input fundus images are taken in the shape of (224,224,3) and are converted into grayscale images due to the ability of the algorithm to handle grayscale images is better. For linear transformation, the weights and bias are initiated to zero at the start. Sigmoid function is used so that it can act as an input for the linear transformation of the features. Forward propagation is used here for calculate the predicition value is calculated with the help of a given weight bias and dataset. After this the cost function is calculated for which gradient descent is used to minimize it. After propagating backwards for the calculation of the derivatives, the weight and bias are updated accordingly. With the help of two functions, the above four operations are carried out. Using the final values, different parameters of iterations and learning rate are used to run the model. The final implemented model runs 200 iterations at learning rate of 0.0006. With this learning rate, the model seems to perform better than with other values. (Emon et al.; 2021) have implemented a similar Logistic Regression model with a smaller dataset from UCI repository and have achieved an accuracy of 75% with their implemented model. The evaluation of this implemented model is discussed in the below subsection.

5.6 Evaluation of Implemented Logistic Regression

The use of different parameters helped gain an understanding of the tuning the parameters. With the 200 iterations having learning rate 0.0006, it can be seen from Figure 9 that the cost has minimized through the complete run of iterations with the help of which the ability of the model for prediction would have increased. With the use of forward

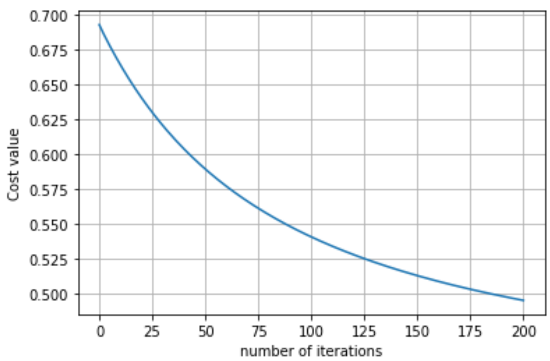

Figure 9: Cost Function

and back propagation along with gradient descent for cost minimization and tuned learning rate, training accuracy of the model is 80.07% while the testing accuracy obtained is 78.21%. Figure 10 illustrates the confusion matrix of the implemented model. The computed sensitivity is 67% and specificity is 89%. The F1 score calculated from the matrix is 0.76. Considering the 75% accuracy achieved by (Emon et al.; 2021), it can be inferred that this model performs better considering the fact that it has achieved a higher accuracy rate and is also based on a large dataset. The strong specificity score indicates the strength of the model in predicting most of the No DR cases successfully.

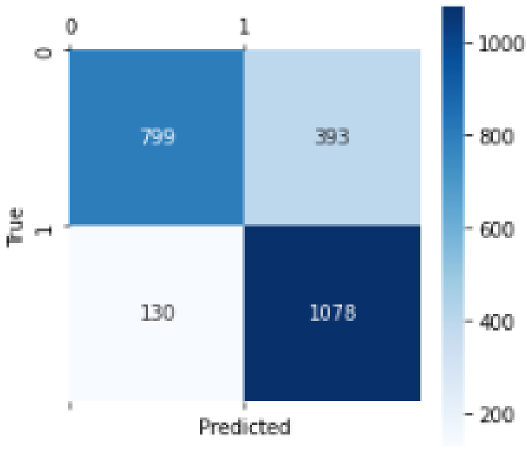

Figure 10: Logistic Regression Confusion Matrix

6 Results & Discussion

This section discusses the implemented models of CNN, ResNet 50 and Logistic Regression. One of the objectives of this research is to discuss and compare the three implemented models based on their performance considering that the same number of images has been passed through each of these models. The models have been evaluated using the metrics like accuracy, sensitivity, specificity and F1 score. These evaluation metrics have been chosen because of the need to evaluate a model developed for the health sector where it is critical to understand the model's ability to correctly identify a patient with a disease and one who does not have a disease. Figure 11 represents the performance of the three models. It can be seen that the specificity of the three models

Figure 11: Performance Comparison

are towards a higher end which means that in majority of the cases, the three models would be able to correctly classify the patient without DR with the custom optimized CNN achieving the highest score of 93.4%. On the other hand, the sensitivity levels of ResNet 50 and Logistic Regression models are way too low to be considered for use in healthcare sector. Here, the custom optimized CNN proves to be useful and has a score of 92.9% and has the highest sensitivity amongst all. The testing accuracy of ResNet 50 is 77.19% and Logistic Regression is 72.21%. In this case as well, the custom optimized CNN model excels by achieving an accuracy of 93.2% for the test set. Observing the F1 scores, it can be seen that the score of CNN model 0.93 which is the highest amongst the three models. It can be concluded from the above comparative analysis that the custom optimized CNN model exceeds the ResNet 50 and Logistic Regression models in each of the critical evaluation metrics and has a better performance overall.

7 Conclusion and Future Work

The increasing number of diabetic patients and the workload on the ophthalmology departments on the nation's health infrastructure has been a cause of concern for governments worldwide. This research aimed at creating a novel approach of implementing a custom optimized CNN architecture which will improve the overall performance of the available DR prediction models. For this purpose, this research has used the standard practices of data collection, preprocessing, transformation and model building using techniques of normalization, data balancing using augmentation and the intricate processes of each of the model building phases. The implemented CNN model has 3 convolutional layers whose weights and biases are dynamically calculated for each layer depending on the previous input with fine tuning the other available parameters of the model like pooling, activation function, fully connected layer. This model is evaluated with the critical metrics like sensitivity, specificity, F1 score and accuracy of the model. The model has achieved a highest accuracy of 93.7% while having a specificity of 94.6% , specificity of 88.7% and a F1 score of 0.91. While achieving these metrics, it has managed to achieve a superior performance when compared to the Logistic Regression and ResNet 50 models. Thus, the main objective of the study has been accomplished by evaluating degree to which the evaluation metrics of the developed custom optimized CNN's to identify diabetic retinopathy using retinal fundus pictures.

The use of robust machine and deep learning models can definitely help reduce the workload on the health infrastructure of a country, and can be a benefit to the patients as they might be able to start the treatment process for diabetic retinopathy earlier without waiting for their long queue for a follow-up so that they can be referred to a eye specialist. Considering the ethical and human concerns of the medical organizations in implementing the use of such models in the healthcare sector, it is necessary to create a model which can be measured up to the high quality standards of the medical field. This could be achieved by using such efficient models combined with the input from the medical research team on the latest medical advancement in this field so that this can together help create a model that can be deployed for diabetic retinopathy detection. Using this idea, in the future, a system which scans for the retinal fundus images can be integrated with the implemented model . Here, the images will be directly taken in as input from the scans and tested by the best performed model so that the automated prediction for DR screening can be done instantly.

8 Acknowledgement

Without the tremendous assistance of my supervisor, Aaloka Anant, this implementation and the research supporting it would not have been possible. His zeal, expertise, and meticulous attention to detail have inspired me and guided my work from the beginning of the research process to the final manuscript.

References

Alam, M. S., Wang, D. and Sowmya, A. (2021). Image data augmentation for improving performance of deep learning-based model in pathological lung segmentation, *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–5.

Chen, C.-Y. and Chang, M.-C. (2022). Using deep neural networks to classify the severity of diabetic retinopathy, *2022 IEEE International Conference on Consumer Electronics - Taiwan*, pp. 241–242.

Ebin, P. and Ranjana, P. (2020). An approach using transfer learning to disclose diabetic retinopathy in early stage, *2020 International Conference on Futuristic Technologies in Control Systems Renewable Energy (ICFCR)*, pp. 1–4.

Elswah, D. K., Elnakib, A. A. and El-din Moustafa, H. (2020). Automated diabetic retinopathy grading using resnet, *2020 37th National Radio Science Conference (NRSC)*, pp. 248–254.

Emon, M. U., Zannat, R., Khatun, T., Rahman, M., Keya, M. S. and Ohidujjaman (2021). Performance analysis of diabetic retinopathy prediction using machine learning models, *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, pp. 1048–1052.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). The kdd process for extracting useful knowledge from volumes of data, *Communications of the ACM* **39**(11): 27–34.

Guo, W., Guo, R., Du, X., Zhao, Y., Li, X. and Toe, T. T. (2022). Diabetic retinopathy detection method based on improved convolutional neural network using fine-tuning, *2022 4th International Conference on Robotics and Computer Vision (ICRCV)*, pp. 113–117.

Kubde, A. and Mohod, S. (2021). Automated computer aided detection of diabetic retinopathy using machine learning hybrid model, *2021 Sixth International Conference on Image Information Processing (ICIIP)*, Vol. 6, pp. 218–223.

Kumari, C. U., Hemanth, A., Anand, V., Kumar, D. S., Naga Sanjeev, R. and Sri Harshitha, T. S. (2022). Deep learning based detection of diabetic retinopathy using retinal fundus images, *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, pp. 1312–1316.

Maneerat, N., Thongpasri, T., Narkthewan, A. and Kimpan, C. (2020). Detection of hard exudate for diabetic retinopathy using unsupervised classification method, *2020 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1–4.

Moccia, S., De Momi, E., El Hadji, S. and Mattos, L. S. (2018). Blood vessel segmentation algorithms — review of methods, datasets and evaluation metrics, *Computer Methods and Programs in Biomedicine* **158**: 71–91.

Nurmalasari, R. R., Irawan, K., Suwita, F. S., Hadi, A., Khomisah and Sumiati (2022). Retinal fundus images classification to diagnose the severity of diabetic retinopathy using cnn, *2022 8th International Conference on Wireless and Telematics (ICWT)*, pp. 1–5.

Odeh, I., Alkasassbeh, M. and Alauthman, M. (2021). Diabetic retinopathy detection using ensemble machine learning, *2021 International Conference on Information Technology (ICIT)*, pp. 173–178.

Ouyang, Y., Wang, K. and Wu, S. (2019). Sar image ground object recognition detection method based on optimized and improved cnn, *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Vol. 1, pp. 1727–1731.

Paradisa, R. H., Sarwinda, D., Bustamam, A. and Argyadiva, T. (2020). Classification of diabetic retinopathy through deep feature extraction and classic machine learning approach, *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, pp. 377–381.

Qian, Z., Wu, C., Chen, H. and Chen, M. (2021). Diabetic retinopathy grading using attention based convolution neural network, *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Vol. 5, pp. 2652–2655.

R, H., B, G. and P, P. (2022). Retinopathy detection on diabetic patients through transfer learning based inceptionv3 algorithm, *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 299–306.

Reddy, G. T., Bhattacharya, S., Siva Ramakrishnan, S., Chowdhary, C. L., Hakak, S., Kaluri, R. and Praveen Kumar Reddy, M. (2020). An ensemble based machine learning model for diabetic retinopathy classification, *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pp. 1–6.

Singh Gautam, A., Kumar Jana, S. and Dutta, M. P. (2019). Automated diagnosis of diabetic retinopathy using image processing for non-invasive biomedical application, *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 809–812.

Ying, Z. (2022). Classification of diabetic retinopathy based on b-resnet, *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*, pp. 570–575.