A Novel Deep Learning Framework for Diabetic Retinopathy Detection Integrating Ben Graham and CLAHE Preprocessing

MSc Research Project
Data Analytics

Namrata Shrishail Tarade
Student ID: X22238867

School of Computing
National College of Ireland

Supervisor: Vikas Tomer

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Namrata Shrishail Tarade		
Student ID:	X22238867		
Programme:	MSc. Data Analytics	**Year:**	2024
Module:	MSc. Research Project		
Supervisor:	Vikas Tomer		
Submission Due Date:	16/09/2024		
Project Title:	A Novel Deep Learning Framework for Diabetic Retinopathy Detection Integrating Ben Graham and CLAHE Preprocessing		
Word Count:	8902	**Page Count:**	25

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:	Namrata Shrishail Tarade
Date:	16th September 2024

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):	

A Novel Deep Learning Framework for Diabetic Retinopathy Detection Integrating Ben Graham and CLAHE Preprocessing

Namrata Shrishail Tarade
x22238867

Abstract

Diabetic retinopathy (DR) is a severe complication of diabetes that can lead to blindness if not detected and managed appropriately. The disease progresses through various stages from the initial mild non-proliferative abnormalities to advanced proliferative retinopathy, where new blood vessels develop on the retina. Early detection and treatment of DR are critical to preventing vision loss and improving patient outcomes. Traditional approaches to DR detection have relied on manual examination by ophthalmologists, which has been criticized for being time- consuming and vulnerable to human error. Automated detection has been attempted using machine learning approaches, such as Support Vector Machines (SVMs) and Random Forests, but these approaches often find difficulty with large and complex datasets. This study brings forth a new deep learning-based approach to DR detection, which utilizes three advanced models, Custom CNN (Convolutional Neural Network), EfficientNetB7, and NasNet. The research applied sophisticated image preprocessing techniques such as Ben Graham and CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve image quality. While all models produced the same accuracy, NasNet produced higher precision, recall, and F1-score measures, an advance in the quality of detection. The novelty in this research is in the evaluation and comparison of the models' performance using various metrics and advanced image preprocessing, offering a more reliable and accurate detection framework for diabetic retinopathy.

1 Introduction

Diabetic retinopathy (DR) has been identified as a complicated condition of the retina from diabetes mellitus. It causes lesions within the retina and adversely impacts vision. Within the healthcare landscape, the scope to treat complex diseases has become more effective and sophisticated, thus enabling an efficient detection of these diseases in the early stages. Diabetes can be identified as a life-threatening condition that increases the blood glucose level due to the insufficiency in insulin production by beta cells. As per the statistical report, diabetes affects more than 425 million people globally Alyoubi et al. (2020). While diabetes itself is a life-threatening and complex, disorder, the underlying effect on the retina, heart, and other organs typically underpins its multifaceted condition. One such disease caused by Type II diabetes mellitus T2DM) is diabetic retinopathy which has been identified as a complication causing a protruding condition in retinal blood vessels and thus, leaking fluids and blood Dai et al. (2021).

The impact of diabetic retinopathy can lead to the blinding condition of the affected person if it is not treated in the early stage. As per the statistical report, approximately 2.6% of vision impairment is caused by diabetic retinopathy and the possibility increases with the extreme suffering of diabetic patients Alyoubi et al. (2020). Therefore, regular screening of the retina is important for these patients to observe and understand the condition. The detection of DR is typically enhanced by observing the condition of the lesion in retinal images which appear in different types - microaneurysms, hemorrhage conditions, and soft/hard exudates Pires et al. (2019). The traditional processes which were manual observation of fundus images by optical experts were time-consuming, effort-intensive, and costly. Also, conventional means of DR detection were not reliable and could provide ineffective screening outcomes. Therefore, an improvement in the detection process using state-of-the-art methods has been considered through which improved performance is achieved. The automated DR detection process is cost and time-effective compared to the manual process and even prevents misdiagnosis. In this study, the application of the deep learning method has been examined that utilizes the fundus images as a dataset to classify and detect the DR condition. Deep learning is a disciplinary domain of machine learning, which involves hierarchical deep layers with non-linear data processing stages Goodfellow et al. (2016). This method applies to learning both unsupervised features and classifying patterns from the datasets. It is widely applicable in medical diagnosis and serves potential applications in image analysis, classification, segmentation, and retrieval of those images. Thus, this study has used this model and fundus images as a dataset to classify and detect diabetic retinopathy.

1.1 Motivation

The motivation of this research comes from the critical necessity to advance early detection of diabetic retinopathy (DR), being one of the leading causes of blindness in today's diabetic patient population around the world. Traditional methods of detection involve manual examination a method that is time-consuming with the potential for human error and as a result, a delayed prevention and diagnosis of the disease. With the emergence of deep learning, however, there is potential to develop a solution that carries out automated detection with a higher level of accuracy in less time. Automating this process would undoubtedly reduce the workload of healthcare professionals and offer a chance for a timelier intervention for their patients. Here, with models such as NasNet, EfficientNetB7, Custom CNN, and more advanced image preprocessing methods, we aim to develop a new stage of DR detection and reduce the burden for HCPs, while augmenting patient care and ideally contributing to the health of their vision.

1.2 Research Objectives

To accurately detect diabetic retinopathy following research objectives are derived.

- To develop a deep learning-based framework for the detection of diabetic retinopathy using advanced models such as Custom CNN, EfficientNetB7, and NasNet

- To apply and evaluate the effectiveness of image preprocessing techniques, including the Ben Graham method and CLAHE, in enhancing the quality of retinal images for more accurate detection.

- To compare the performance of Custom CNN, EfficientNetB7, and NasNet in terms of accuracy, precision, recall, and F1-score, identifying the most suitable model for diabetic retinopathy detection.

1.3 Research Question

- How can advanced deep learning models and image preprocessing techniques be optimized to improve the accuracy and reliability of diabetic retinopathy detection?

2 Related Work

2.1 Chapter Overview

In this part of the study, evidence-based information has outlined the condition of diabetic retinopathy and accordingly presents insights into the way the disease can be detected in the early stage to prevent vision impairment. Different models and automated methods have been explored while comparing the methods in terms of performance accuracy and other metrics.

2.2 Diabetic Retinopathy Detection Using Fundus Image Datasets

Diabetic Retinopathy (DR) is a multifaceted condition of patients with diabetes mellitus, which is highly complicated. The medical diagnosis of this disease is challenging since in most cases, the diagnosis process failed to detect the symptoms in its early stages. While understanding this, medical experts have stated the relevance of fundus images, which are digitally processed and used in manual and advanced DR detection. According to the information presented by Pires et al. (2019), the digital imaging of fundus images has enabled a computerized screening process that deals with and documents information on retinal diseases. Generally, the development and acquisition of fundus images are formed by capturing images through a camera consisting of a low-power microscope that can capture the inside details of the retina. According to the information presented by García et al. (2017), it is important to understand the minute details of the disease where fundus images of the retina enable experienced clinicians to gather such information. In the above study, the data used consisted of 35,126 high-resolution fundus images captured from different angles and various conditions. It is based on the dataset's significance that upon training the model, it provides a specificity of nearly 93.65%, while the accuracy is 83.68% respectively.

Author-Year	Model used	Performance Metrics	Findings	Limitation
Dai et al. (2021)	DeepDR	AUC to detect microaneurysms, exudates, haemorrhages, and cotton-wool spots	A DL method – Deep DR trained with fundus images for lesion detection provides suitable results with an accuracy of 94.3%, 95.5%, and 97.2% and proving the efficiency of the automated system in DR grading.	Accurate microaneurysm detection is a consistent issue for the DL system
Tymchenko et al. (2020)	CNN	Sensitivity, Specificity, Kappa score	The concerned study has presented the significance of the CNN model to help in the diagnosis of DR. In this regard, a labeled dataset has been used and further classified as well as segmented images for better detection.	Inconsistency in performance of medical experts with these models and expensive labeled datasets
Qummar et al. (2019)	ResNet50, Inceptionv3, Xception, Dense121, and Dense169	Recall, Specificity, Accuracy, Precision, and F1-score	Introduced the significance of 5 CNN models in this study, which achieves highly accurate performance. As a result, it can be stated that each CNN model performs better than the existing state-of-the art models when utilizing a similar dataset.	Mainly focused on ensemble outcome rather than the model specific detection accuracy for different stages.
Qureshi et al. (2019)	Computer Assisted Diagnosis (CAD)	AUC and accuracy	As per the information presented by the study, it has been identified that a CAD system introduced in the detection process.	Improvement in the accuracy level is required to ensure a better diagnosis result.
Pires et al. (2019)	CNN	Area under ROC curve, and error rate, accuracy	A data-driven approach has been presented to extract a highly significant and powerful representation typically retrieved from direct retinal images. The study shows that a deep neural architecture that has been used to classify the images based on data augmentation and feature extraction serves patient-based analysis, improving the accuracy level.	Lack of focus on computational as well as implementation efficiency to serve the detection benefits across all stages of diagnosis

Pradeep Kumar et al. (2021)	Microaneurysm’s detection algorithm.	Accuracy to detect the microaneurysms.	A novel systematic approach has been developed for identifying the infection in the retinal images for diagnosing diabetic retinopathy with results using the Messidor dataset.	The complex structure of the algorithm might be an issue if implemented into real-time and the efficiency of the method should be in check with other databases.
Naveen et al. (2019)	Adaptive Histogram Equalization (AHE) and Contrast Limited Adaptive Histogram Equalization (CLAHE)	Specificity, Sensitivity and Accuracy	Proposed an Image processing technique with CLAHE for the detection of DR with high accuracy of 98% in the identification of DR from fundus images.	The method may magnify noise in areas with relatively low intensity differences and more work can be done to generalize the method successfully for a diverse set of retinal images.
Zaman & Bashir et al. (2016)	Convolutional Neural Network (ConvNet)	Accuracy	As per the information presented by the study, investigated a method of ConvNet for DR with the goal of diagnosing multiple stages of the retinopathy and emphasized on the abilities of DL in the field of analyzing medical image.	The absence of specific performance metric and thorough analysis of each and every proposed model which is important in aiming to authenticate the efficacy of proposed model.
Argade et al. (2015)	K-Nearest Neighbor Algorithm, Decision Tree, Ellipse Fitting.	Specificity: 80.20%, Sensitivity: 70.66%	An automated technique in the detection of Retinal abnormality through the use of image processing and data mining approach for DR detection, particularly on features that may be indicative of early DR development.	Sensitivity and specificity values are less from the standard values which may lead to some problem while in clinical use.

Ganguly et al. (2014)	Adaptive Threshold Algorithm	Specificity and Sensitivity metrics were used.	Proposed technique for segmentation of red lesions in the fundus images in particular microaneurysms and hemorrhages that are the early signs of DR.	Lack of accuracy metrics which is essential in determining the reliability of the method.
Elbalaoui et al. (2016)	Hessian multi-scale enhanced filter, vessel filter	Metrics mention is missing	In this study, a technique was used for identification and characterization of retinal blood vessels needed for detection of DR, glaucoma, and hypertension.	No specific performance measures are given and the detection of vessels without taking into account other characteristics of DR such as exudates and microaneurysms.
Bansal et al. (2016)	Multilevel threshold, Automated intensity method	Precision, Specificity and Sensitivity	A simple technique used for macula identification in the retinal images that are crucial for DR identification and yielded good results in preliminary works.	More extensive experiments on the different size data sets can be conducted in order to prove the efficiency of the proposed method.
Narkhede et al. (2015)	KNN, Histogram Threshold and Decision Tree	Sensitivity and Specificity	Proposed the screening for DR using the automated system that involve the image processing, especially focusing on the two major performance indicators – sensitivity and specificity.	It lacks optimal sensitivity and specificity required in screening for DR which is important for most of the false positive and false negative results.
Minar et al. (2015)	Fundus image analysing, Laplace method.	Precision and recall for vessel Detection.	In this study, it is observed that a different approach used for the blood vessel detection applying the Laplace operator to increase the level of accuracy during the examination of the fundus image.	It only focuses on vessel extraction ignoring other DR-related features hence only useful in diagnosis of DR.

Table 1: Summary of Diabetic Retinopathy Detection Studies

2.3 Diabetic Retinopathy Detection Based on Traditional Approaches

Image processing by performing suitable operations is a mandatory procedure in the effective detection of diabetic retinopathy. The condition of diabetic retinopathy (DR) is a severe medical complication that has come to attention in recent years because of the rising dilemma in medical diagnosis. Based on the explanation provided by Naveen et al. (2019), disease diagnosis using image processing techniques has become a common procedure where different methods have been applied over the years. In the above study, one such method has been used to understand the condition of DR and particularly focuses on patients with diabetes. In this study, an “adaptive Histogram Equalization” method has been used using the “Contrast Limited Adaptive Histogram Equalization” (CLAHE) algorithm. The framework has been implemented for DR detection based on which the result obtained shows an accuracy of 98%, indicating the method’s exclusiveness and specificity in image processing. Al Hazaimeh et al. (2018) in their study explained that the risk of vision impairment due to DR is effectively studied using image processing techniques and further evaluated through various performance metrics. As per the evidence, it has been identified that an automatic screening process of DR images has been used where a software simulation I performed based on MATLAB using the DIARETDB1 dataset.

The experimental result presented by the study shows that the method is validated through a comparison with the manual process performed by clinical ophthalmologists. As the result progressed, it was identified that the technique provided an effective detection outcome with enhanced accuracy, specificity as well as sensitivity. The process of diagnosing diabetic retinopathy is done through automatic screening which is a precautionary development in a way that it will not be diagnosed wrongly, but will produce a better result. The interference with the retina as described by Argade et al. (2016) requires efficient interpretation to avoid this impact. In this regard, regular screenings help medical experts in the easy detection of the diabetic patient’s condition. The study, upon understanding the situation has emphasized retinal image determination through effective data mining procedures. Based on the observation, it can be stated that image processing as well as data mining provides a typical comparison between normal and affected images, thus ensuring proper DR diagnosis.

Pradeep Kumar et al. (2021) explained that the foremost detection of diabetic retinopathy through fundus images involves practitioners who can recognize small topographies with the detail of grading procedures. The identification of diabetic retinopathy (DR) through biomedical image processing and analysis has become the most compatible procedure among all other clinical diagnoses and activities. The utilization of fundus images in diabetic retinopathy helps practitioners gather relevant insights into the proper diagnosis of retinal diseases, especially diabetic retinopathy Pradeep Kumar et al. (2021). As the study progressed, the image processing technique, preferably an algorithm, was tested using fundus images that further characteristics image variants of DR as microaneurysms and the severity in the type. As the result indicated, the algorithm serves a better outcome upon testing through the dataset. It has been further observed that the dataset introduced by ophthalmologists has provided an excellent result with a promising solution. The complications due to diabetes have become a global concern for years. As per the evidence provided by many studies, it has been identified that diabetic retinopathy, which is a multifaceted condition in patients with diabetes impacts small retinal vessels and leaks blood and fluids, which therefore causes vision impairment.

Upon understanding the condition, Abreu et al. (2021) introduced a software - RapidMiner and CRISP-DM methodology. Accordingly, various classifiers have been developed under different circumstances, which were tested and trained using DR-based datasets. Indicating the result, it has been observed that the software provides nearly 76.90% accuracy, 85.92% precision rate, and 67.40% sensitivity when combined with an algorithm - Logistic Regression (LR). It can be stated more improvement is required in the accuracy rate of detection to ensure a better outcome.

2.4 Diabetic Retinopathy Detection Based on State-of-the-Art Methods

The diagnostic procedures and potential treatments for diabetic retinopathy have shown an evolution over the years. Bestowing the effect, it has on patients with the underlying impact of diabetes has urged the need to define and explore improved measures for successful outcomes. Amid this understanding, the contribution of state-of-the-art methods has been studied that has enhanced the grading process in DR detection. In this regard, one of the studies presented by Qureshi et al. (2019) explained that computer assisted diagnosis (CAD) has positively enabled the control and effective treatment of diabetic retinopathy. Previously, the manual detection process was highly challenging, and the development of the CAD technique has become a fundamental progress for ophthalmologists in recognizing both inter and intra-variations. Indeed, the result suggested an improved outcome; however, the study exclaimed that a need for improvement in accuracy with the CAD system is essential. Therefore, future studies are needed on this matter. Contrastingly, another study presented by Malhi et al. (2023) explained that an automated process of grading the variations of DR based on fundus images is a significant improvement in the detection mechanism where real-world datasets are used.

In the above study, an automated feature extraction process using machine learning (ML) models has progressed. Different models are used to make a comparison based on grading parameters and observed that K-Nearest Neighbor (KNN)shows the highest accuracy estimated to be 92.1% for microaneurysms while decision tree (DT) has presented a high-level accuracy of 99.9% in terms of the disease severity prediction. The information presented by another study by Selvachandran et al. (2023) explained that computer-aided diagnosis is an advanced process of disease detection that has efficiently reduced the disease burden on ophthalmologists. Understandably, the above study has introduced some state-of-the-art methods such as “artificial neural network” (ANN) and “support vector machine” (SVM) that show suitability in DR detection. Diabetic retinopathy detection has undergone evolution and become more advanced with time. In this regard, Bilal et al. (2021) explained that the integration of state-of-the-art methods in detecting the non-proliferative condition of DR describes an immense observation of exudates, signs of hemorrhages, and the presence of microaneurysms. The study further stated that by using artificial intelligence (AI), the screening process has become more acceptable compared to previous challenges identified with manual approaches.

From the above research perspectives, it is an informative consideration that understanding the prolonged complications of DR requires enhanced forecasting as well as evaluation procedures. Understandably, many studies have presented explorable solutions where feature extraction from exclusive datasets based on fundus images is ensembled with various supervised ML models. Amid this consideration, the information presented by Bhardwaj et al. (2021) identified the importance of applying ensemble methods containing two or multiple classifiers. Considerably, a model named the "Prominent Feature-Based Transfer Learning" algorithm has been used where statistically optimized features are utilized that has yielded an accuracy estimated to be 90.51%. The model has utilized the Inception V3 method and provides an improved detection feature through parametric observations.

2.5 Diabetic Retinopathy Detection Based on Deep Neural Methods

In previous sections, empirical evidence on data mining procedures using MATLAB and image processing based on Computer-Assisted Diagnosis (CAD) has been established that provide somewhat accurate results in terms of DR detection. Contrastingly, evidence-based insights are further investigated on improved methods where state-of-the-art supervised machine learning models have proven their effectiveness in the efficient detention of the complication of diabetic retinopathy (DR) disease. However, while observing their accuracy level, the implication indicates a need for further performance improvement to ensure a better medical diagnosis. Understandably, a focus on advanced deep neural networks has been attained which provide extensive detection results. In the study presented by García et al. (2017), the information introduced an improved neural architecture a convolutional neural network that has identified exudates, microaneurysms as well as hemorrhages within fundus retinal images. The model has been trained using a database that contains nearly 35,126 retinal images of high-resolution and provides an accuracy level of 83.68%, with specificity is 93.65%, respectively. In another study presented by Qummar et al. (2019), it was identified that the complication with encoded features from the Kaggle dataset containing retinal images provided a low accuracy with traditional state-of-the-art methods.

On the contrary, when the above dataset is used to train five deep CNN models that are in ensemble form encode rich features and further improve the classification accuracy of DR at different stages. As per the result obtained it has been identified that, unlike the existing state-of-the-art methods, the CNN ensemble when trained with the Kaggle dataset provides a highly accurate performance Qummar et al. (2019). The identification of diabetic retinopathy has become a potential priority of medical experts, particularly ophthalmologists through enhanced diagnostic performance. As per the information provided by Tymchenko et al. (2020), the accurate identification of DR has become increasingly challenging with human interpretations using fundus images. Therefore, it has become an important consideration for researchers and medical experts to simplify the detection process to help people worldwide. Previously, it has been identified that diabetic retinopathy is prevalent in patients with diabetes mellitus, which has increased persistently across the world as a major noncommunicable chronic illness. Hence, the detection of the complications of DR disease by classifying the stages through fundus images is important.

Understandably, Dai et al. (2021) introduced a DeepDR model that has been trained using a local dataset for real-time assessment of images and shows effective detection of the disease. As per the grading process followed in the study, it has been observed that the model has effectively detected lesions based on approximately 466,247 fundus images that have been collected from nearly 121,342 diabetes patients. Based on the result obtained, it has been identified that the AUC provides a range of consideration between 0.916 and 0.970, which validates the system. Further, the grading process shows mild moderate, and severe as well as proliferative AUCs, which are estimated to be 0.943, 0.955, 0.960, and 0.972 respectively, thus indicating the model efficiency in diabetic retinopathy detection. The overall information has identified the necessity of the early diagnosis of diabetic retinopathy where retinal screening plays an important role. As per the information presented by Dai et al. (2021), the estimation from recent research has forecasted that approximately six hundred million individuals will suffer from diabetes by 2040, while nearly 1/3rd of the population is expected to experience retinal disease, precisely diabetic retinopathy. It is already aware that DR, particularly mild and non-proliferative DR (NPDR) is one of the factors causing vision impairment and even characterized by microaneurysms presence. Therefore, the detection of the disease is imperative to prevent the risk of vision loss. At the same time, it has also been determined that early intervention through the control of glycemic index and blood pressure can reduce the risk of diabetic retinopathy, while in the late-stage intervention, the impact of photocoagulation (also known as intravitreal injection) can further reduce the chance of vision loss.

3 Methodology

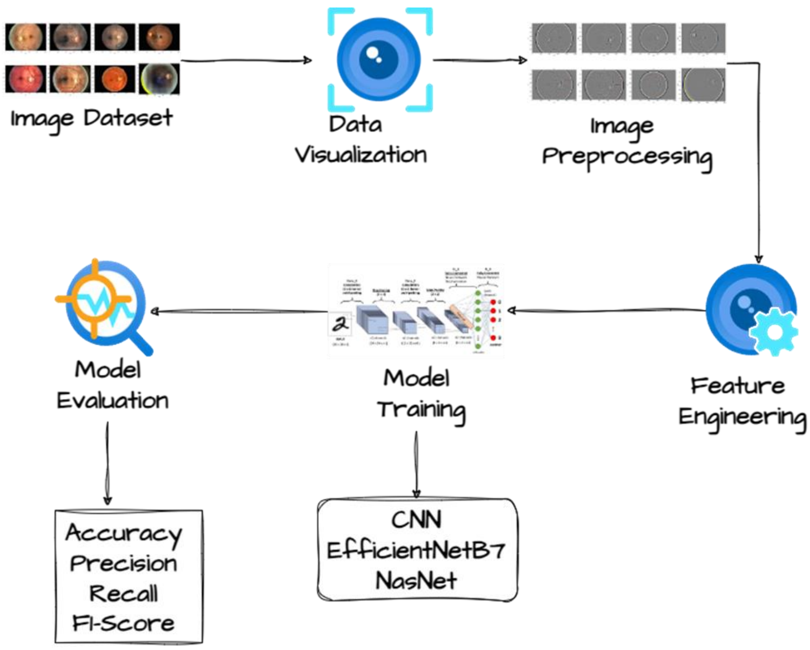

Figure 1: Schematic Methodology diagram for the detection of Diabetic Retinopathy

Diabetic Retinopathy is the commonly occurring cause of new cases of blindness across adult in the age year 20-74. For the first two decades of this disease, all the patients with diabetes of type 1 and greater than 60% of patients with diabetes of type 2 suffer from retinopathy. Diabetic Retinopathy increased from mild nonproliferative abnormalities, which is identified by enhancing vascular permeability. Diabetic retinopathy progresses from mild nonproliferative abnormalities, accompanied by the resulting increment in vascular permeability to moderate and severe non-proliferative diabetic retinopathy. It is worth mentioning that although NPDR is accompanied by vascular closure, NPDR can progress to develop into PDR. Our methodology follows the step-by-step process to accurately detect diabetic retinopathy disease; a schematic diagram of the methodology is shown in Figure 1.

3.1 Data Description

In this research, the dataset is extracted from Kaggle which gives a large set of retina high-resolution images that is taken in several imaging situations *Diabetic Retinopathy Detection — kaggle.com* (n.d.). The size of the dataset is more than 88.29GB. The dataset consists of retinal images which are labeled with subject ID either for right or left eye. The dataset corresponds to five classification classes named No DR, Mild, Moderate, Severe, and Proliferative DR. The images are captured from different camera models and angles by which the visual appearance of the images is affected as left vs. right. Some images in the dataset show the retina anatomically for which on the left side is the macula, and on the right of the eye is the optic nerve[1].

3.2 Data Analysis and Visualization

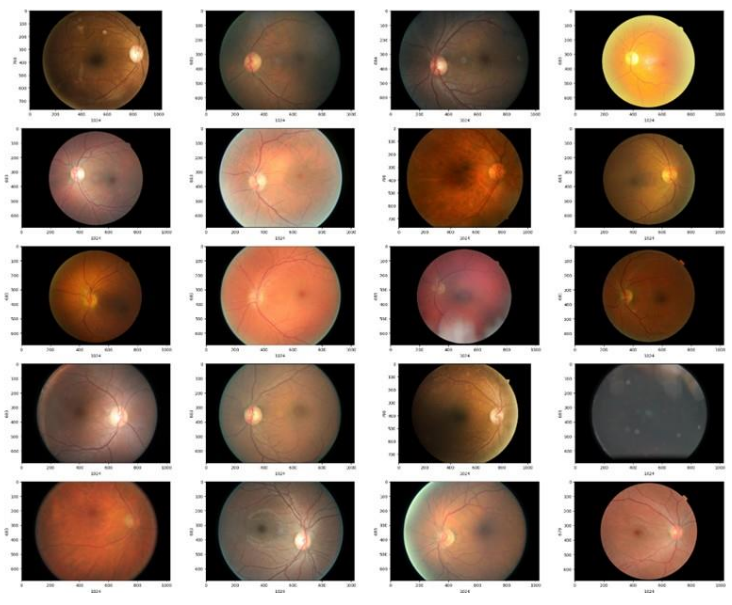

Figure 2: Images of random Retinas captured from different camera Angles.

[1] *https://www.kaggle.com/c/diabetic-retinopathy-detection/data*

Data analysis and visualization are crucial in handling image data as they enable clearer insights into the underlying patterns and features of the data. Visualization helps in effectively interpreting complex image data and results, which makes it easier to identify trends anomalies, and areas requiring further analysis. The images of random eyes that are captured from different camera models are visualized in Figure 2. Since the data contains five different classes, which makes this a multiclass classification task, thus it is necessary to understand the data distribution of each class for a better understanding of the data and this helps in identifying further steps.

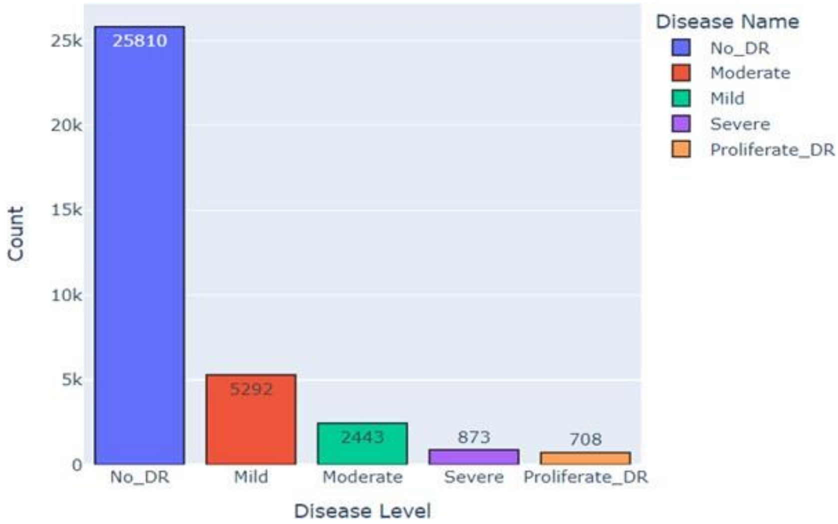

Figure 3: Distribution of Disease Levels

The bar chart which shows the distribution of disease levels in Figure 3 represents the count of images in each class is not uniformly distributed thus raising the issue of data imbalance, so there is a need to balance the dataset in each class.

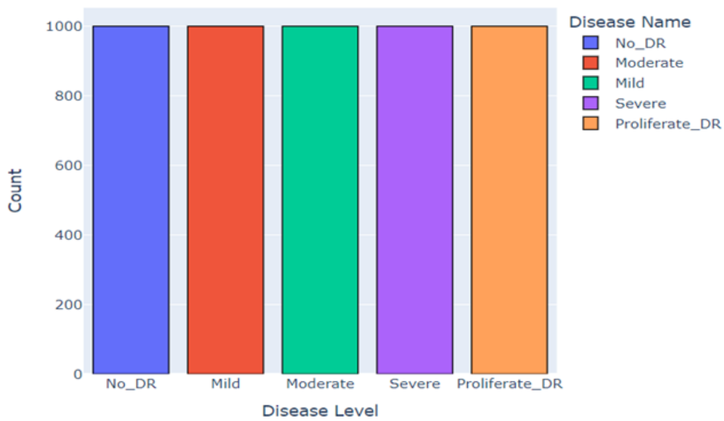

Figure 4: Distribution of disease levels after balancing the classes

A balanced distribution of classes is visualized in Figure 4. The classes are balanced by augmenting the dataset of images for diabetic retinopathy. The data is separated into majority and minority classes and augmentation methods are applied such as flipping, rotation, and adjustment of the brightness of the images.

3.3 Data Preparation and Image Preprocessing

Image processing plays a crucial role in enhancing the quality and usability of image data by improving clarity and contrast. It helps in extracting meaningful features from images, facilitating better analysis and interpretation. Additionally, image preprocessing techniques, such as noise reduction and normalization, are essential for accurate model training and prediction in various computer vision tasks. The Ben Graham function is used to enhance the features of images. The preprocessed image usually has increased edges and contrasts which makes it suitable for further analysis. The adaptive Histogram Equalization method is used to improve the contrast of the images which specifically uses contrast limited adaptive histogram equalization (CLAHE) technique that enhances the visual quality of images, especially in cases where the lighting conditions are poor or there is a low dynamic range in the image which results in the areas that are highly dark and bright as shown in Figure 6. A few images are visualized after applying the Ben Graham method as shown in Figure 5.

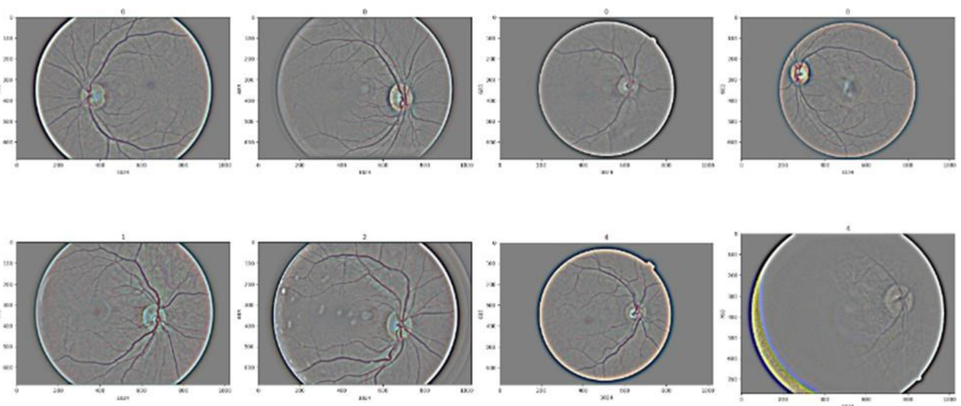

Figure 5: Images after applying Ben Graham method.

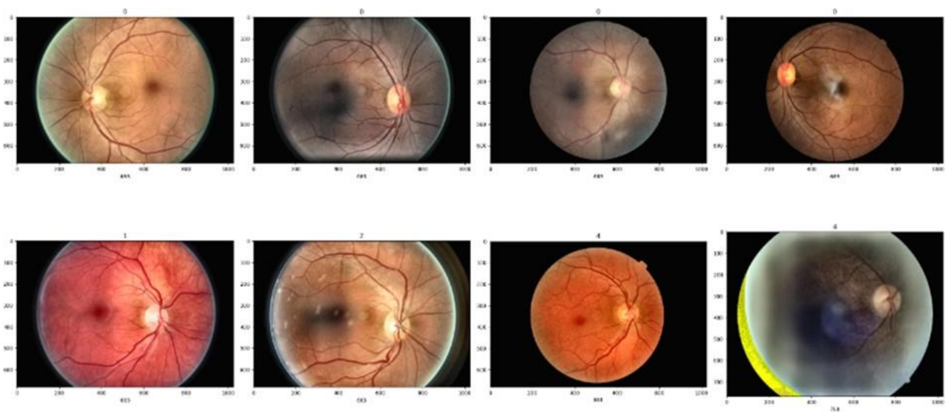

Figure 6: Images after applying CLAHE method.

3.4 Feature Engineering

In image processing, feature engineering is used to extract, select, and transform raw image data into meaningful and informative features for training the models. The features that are captured have the important characteristics of the images that are most relevant for detecting diabetic retinopathy. In this research, feature engineering involves splitting the data into training and testing sets with a training size of 80%. Following the splitting of data, the images are resized, and the Ben Graham method is applied to the images. As for training the deep learning algorithms, a large amount of data is required to generate more data. Image Augmentation technique is applied to generalize the training dataset. In image augmentation the images are rotated with some axis, rescaled in the range 0 and 1, shifting of width and height with some range, images are flipped horizontally, and the images are sheared with some ranges.

3.5 Model Training

Model training is the crucial step in predicting the data. In the model training phase, the model captures and maps features to targets by adjusting the internal parameters. Model training helps to predict the result, helps in capturing the complex patterns, and improves the accuracy of the model. This research is carried out by leveraging the power of three deep-learning algorithms to detect diabetic retinopathy. The three models are custom Convolution Neural Networks (CNN) in which the model is defined, and the other two models are pre-trained CNN models that is EfficientNetB7 Model and NasNet Model.

3.6 Model Evaluation

The last stage in working with a particular algorithm is the evaluation of outcomes after the data has been trained on the algorithm. Therefore, there is a need to evaluate the performance of the algorithm to understand how good the algorithm is in handling the data. Four performance measures are computed for the algorithms specified above. The metrics adopted are accuracy applied when classes are balanced and it gives the ratio of instances that are classified correctly, precision – highlights the ratio of correct positive predictions, recall which indicates the ratio of actual positives that are correctly identified and used where high false negatives are recorded, F1- score is the harmonic average of precision and recall. These metrics are computed by using the test data on the algorithm and its output is then compared with the actual results from the test data set. Thus, it can be stated that using these indicators, the better performance of the given algorithm is indicated.

4 Design Specification

The research focuses on the classification task for detecting diabetic retinopathy which is carried out by implementing three deep learning algorithms: Custom Convolutional Neural Network (Custom-CNN), EfficientNetB7 and NasNet. A detailed explanation of these algorithms is given below.

4.1 Custom Convolutional Neural Networks (Custom-CNN)

Convolutional Neural Networks (CNNs) are the family of deep neural networks that are used for analyzing visual data. CNN is implemented on the image data to learn the spatial hierarchies of the features from the input images by leveraging the power of convolutional layers, pooling layers, and fully connected layers. CNNs help detect diabetic retinopathy because of their ability to learn and capture the features automatically from the retinal images, accurately and efficiently. In this research, we have used the custom CNN, which contains the three convolutional layers, 3 MaxPooling layers, and 2 BatchNormalization layers. After extracting the feature maps from the convolutional layers, a flattened layer is added for converting the feature map into a 1D vector. Three dense layers are added with ReLU activation and a dropout for regularization and the last layers uses softmax activation for the classification task. An architectural diagram of the CNN is shown in Figure 7.

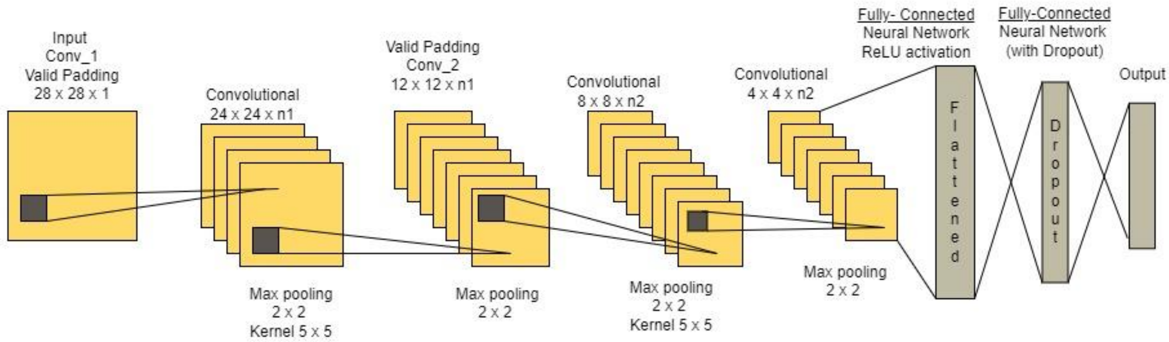

Figure 7: Architectural diagram of Custom CNN

The fundamental mathematical processing in a Custom-CNN include convolution, activation functions, pooling, and full connected layers. Here's a breakdown of the key components:

- **Convolutional Layer:**

$$z_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{c-1} w_{m,n,c,k} x_{i+m,j+n,c} + b_k$$

where Zi,j,k is the output feature map, Xi,j,c is the input, Wm,n,c, are filter weights, and bk is the bias.

- **Activation Function:**

After the convolution layer, an activation function such as ReLU (Rectified Linear Unit) is applied

$$A_{i,j,k}=\text{ReLU}(Z_{i,j,k})=\max(0,Z_{i,j,k})$$

- **Pooling Layer:**

Pooling, often max pooling, reduces the dimensionality of the feature maps.

$$P_{i,j,k=(m,n)\in window} \max A_{i+m,j+n,k}$$

- **Fully Connected Layer**: The output of the final convolutional layer is flattened and passed through one or more fully connected layers

$$y = \text{softmax}\,(W_{fc} \cdot x + b_{fc})$$

where Wfc and bfc are the weights and biases of the fully connected layer, and x represents the flattened input. The logits obtained are passed through the softmax function, a process common in classification problems Goodfellow et al. (2016).

4.2 EfficientNetB7

EfficientNet-B7 is the largest variant model in the EfficientNet family with a series of CNNs implemented to attain high performance efficiently and accurately. Developed by Google for the balanced approach to scale up model size by keeping the efficiency in terms of both accuracy and computational cost. In the research, Efficient-B7 is used for detecting diabetic retinopathy due to its high accuracy, and ability to handle high resolution images. The architecture of this model is suitable for capturing the complex 13 features that are present in retinal images, which makes it a powerful tool for the accurate detection of diabetic retinopathy. A schematic diagram of the EfficientNetB7 model is shown in Figure 8.

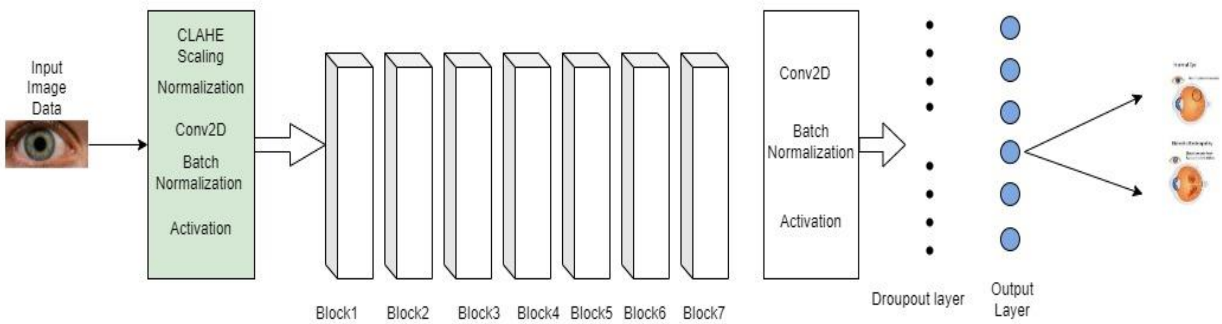

Figure 8: Schematic diagram of EfficientNetB7

The Brief working of each layer in the EfficientNetB7 model which explains how the input data is being processed through each and every layer is explained below:

(1) Global Max Pooling:

$$Z_{GMP} = GlobalMaxPool(Z_{EfficientNetB7})$$

- $Z_{EfficientNetB7}$: It is included as the output of last convolutional layer of the EfficientNetB7 Model which is pretrained.
- Z_{GMP}: It is actually the feature map of the last layer which is compressed and generated after the exclusion of other layers such as dropout layer and max-pooling layer as described by Tan et al. (2019).
- GlobalMaxPool: t is an operation that brings an increment in the connectivity between the neurons and decreases the height and width of the feature map; by taking the maximum value in each feature map.

(2) Fully Connected Layers:

After the pooling, the dense layer is applied for the process to extract the features

$$Z_{dense1} = ReLU(W_{dense1} Z_{GMP} + b_{dense1})$$

A Dropout Layer is applied, so that it can prevent the Overfitting of model

$$Z_{droupout1} = Dropout(Z_{dense1}, p)$$

(3) Final Output Layer (Softmax):

$$\hat{y} = Softmax((W_{out} Z_{dropout2} + b_{out}))$$

The Output Layer computes the final processed features into probabilities using the Softmax function through which the model makes its prediction on the 5 classes of the output layer.

4.3 NasNet

NasNet Mobile represents a significant advancement in the field of neural architecture search (NAS), designed particularly for mobile devices. NasNet Mobile utilizes reinforcement learning methods to automatically discover efficient neural network architectures optimized for mobile platforms. NasNet Mobile utilizes neural architecture search to explore a vast space of possible architectures and identify the optimal configurations for mobile image recognition tasks. Due to the ability to optimize the neural network of NasNet, it is used to detect diabetic retinopathy with high performance, flexibility, and efficient design which makes it ideal for analyzing retinal images. An illustrated diagram of NasNet is depicted in Figure 9.

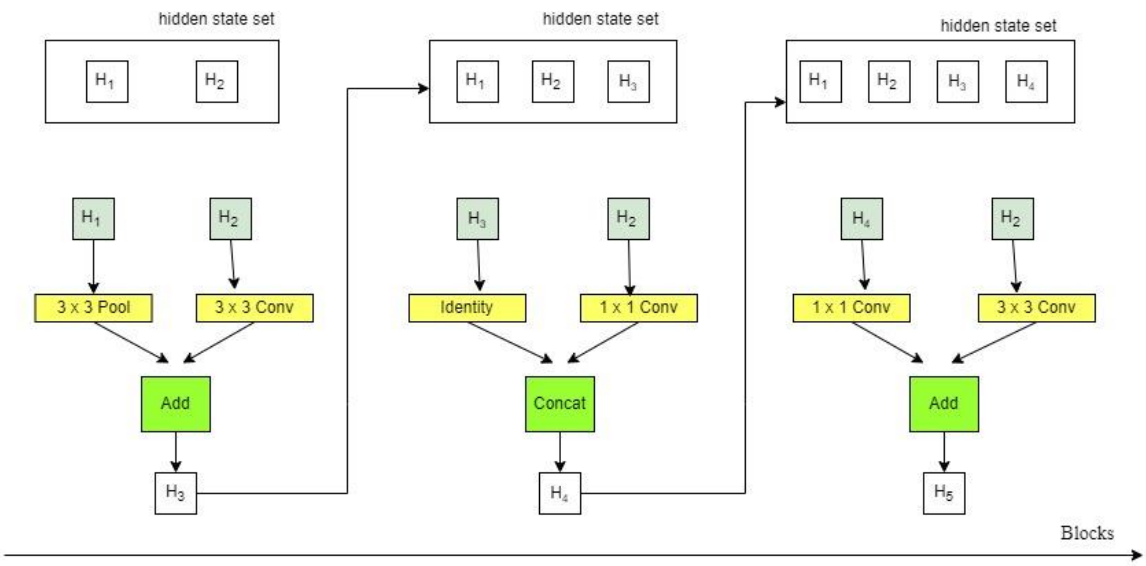

Figure 9: Schematic diagram of NasNet search area

The operations are explained in Algorithmic mathematical terms which gives an idea how the input data is processed at every stage of the model. The model's pre-trained layers are used for initial feature extraction, with additional dense layers for final classification:

Step 1: Input: Image x of the shape 128 x 128 x 3.

Step 2: Passing x through the NASNetMobile pretrained model which extracts the feature maps $Z_{NASNETMobile}$

Step 3: Applying Global Max Pooling to $Z_{NASNETMobile}$

It applies compression on the spatial dimensions and pool the feature maps to create vector Z_{GMP}

$$Z_{GMP} = GlobalMaxPool(Z_{NASNETMobile})$$

Step 4: Applying the Fully Connected Layers:

After the max pooling the pooled features are then passed through in a series of dense layers as used in NasNet Zoph et al. (2018)

- First Dense Layer: Vector multiplied by weights, add bias and applying ReLU activation.

$$Z_{dense1} = ReLU(W_{dense1} Z_{GMP} + b_{dense1})$$

Step 5: Applying the Dropout Layer to prevent the model's overfitting:

- Dropout:

$$Z_{droupout1} = Dropout(Z_{dense1}, p)$$

Step 6: Applying the more Dense Layers after dropout:

- Second Dense Layer:

$$Z_{dense2} = Linear(W_{dense2} Z_{dropout1} + b_{dense2})$$

- Dropout:

$$Z_{droupout2} = Dropout(Z_{dense2}, p)$$

- Third Dense Layer:

$$Z_{dense3} = ReLU(W_{dense3} Z_{dropout2} + b_{dense3})$$

- Dropout:

$$Z_{droupout3} = Dropout(Z_{dense3}, p)$$

Step 7: Output Layer (Softmax):

$$\hat{y} = Softmax((W_{out} Z_{dropout3} + b_{out}))$$

After iterating through every dense layer, class probabilities are produced by the output layer.

Step 8: Output produced class probabilities $\hat{y}$.

5 Implementation

This research is conducted in Google Colab with 12 GB of RAM. To train the high dimensional images Google T4 GPU has been used. The entire implementation of the project has been performed using Python programming language, which focuses on predicting diabetic retinopathy from the retinal images. The implementation is carried out by collecting the data from the Kaggle which is a repository for datasets. The implementation involves libraries such as NumPy, Pandas for reading the data, matplotlib for visualizing the data, and scikit-learn and Tensorflow for image processing and model training. The data is loaded, and the classes are mapped with each image. The retinal images are visualized to understand the dataset. As the number of classes is imbalanced, the classes are balanced with the help of the IEA library by augmenting the images such as Fliplr, and Affline for rotation of images, and the brightness of the images is also increased. The color histograms of the images are plotted by selecting some sample images that help in identifying the distribution of pixel intensities across the red, green, and blue (RGB0 color channels). The data is prepared by augmenting the data by resizing the images using the resize function provided by the cv2 library, and the weights are added using the cv2 library for implementing the Ben Graham and CLAHE for preprocessing the images. For feature engineering, the image data is split into training and testing sets using the train test split method in scikit-learn with a training size of 80%. For effective model training, more images are generated using ImageDataGenerator by the TensorFlow library. The images are augmented using rescaling the images, rotating the images shifting the width and height of the images, and shearing the images. Three deep learning algorithms are implemented on the training data to capture the complex pattern from the images to detect diabetic retinopathy. Following model training, the algorithms are evaluated on four key performance metrics: accuracy, precision, recall, and f1-score.

Component	**Specification**
GPU	Google T4 GPU
RAM	12 GB
Storage	1TB SSD
Operating System	Windows 10
Python Version	3.10
Image Processing	Rescaling, Rotation, Shifting width and height, Shearing, Flipping, Ben Graham preprocessing
ML libraries	Pandas, NumPy, Scikit-learn, TensorFlow, iaa, matplotlib, seaborn, NasNetMobile, EfficientNetB7
IDE	Google Colab
Performance Metrics	Accuracy, Precision, Recall, F1-Score

Table 2: System Requirement and Resource Details

6 Evaluation

The research is carried out to implement three deep learning algorithms on the model: Custom CNN, NasNet, and EffiecientNetB7. The problem is to detect diabetic retinopathy with five classes as No DR, Mild, Moderate, Severe, and Proliferate DR. The model is evaluated on Four key performance metrics: Accuracy, Precision, F1-Score, and Recall.

These metrics then compare with each other to find which model is suitable for predicting the target.

6.1 Evaluation Based on Accuracy

It indicates the proportion of classification to the total number of cases found. It is used when the classes are balanced; The proportion of the links connecting the two classes is equivalent on either side. Higher values of these parameters imply that the given model is behaving optimally in the contextualized environment, and it gives the assertion that the model is performing well. The accuracy achieved by the custom model is 73.3% and the accuracy achieved by NasNet is also the same at 73.3% and there is no change in the performance of the EfficientNetB7 model also with the same accuracy as 73.3% which represents that all the models in this research are equally good at correctly predicting the target. Accuracy can be the same across models because accuracy is reflective of the proportion of correct predictions overall, which could be the same if all models are equally effective in correctly classifying most cases regardless of how they handle individual class distinctions. The performance of these algorithms is visualized in Figure 10.

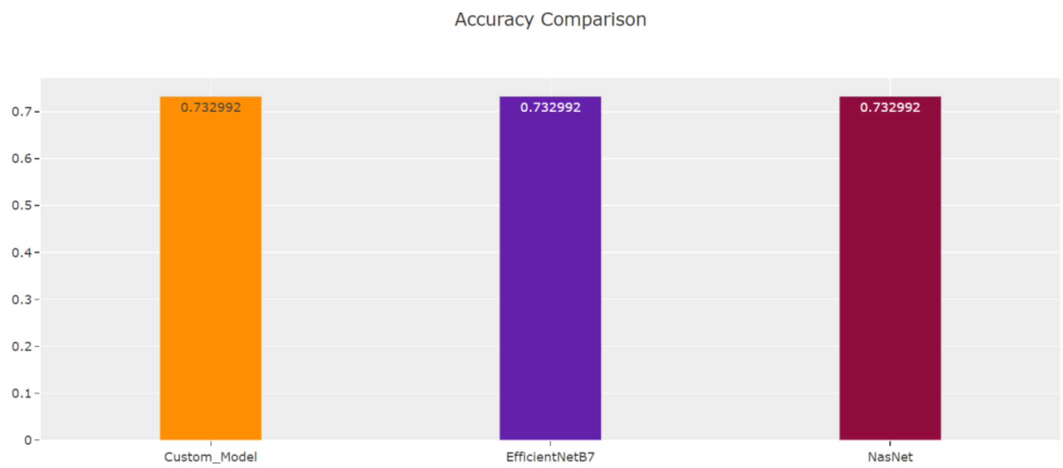

Figure 10: Accuracy comparison of the models for Diabetic Retinopathy Detection

6.2 Evaluation Based on Precision

The ability of the model to accurately detect positive instances of all classes in a multiclass classification scenario is known as 'precision.' In our examination, a significantly lower precision of 53.72% is calculated using our Custom CNN model, compared to NasNet (73.3% precision). The lowest precision is delivered by EfficientNetB7 at 14.65%, resulting in an increase in false positives across all classes. NasNet provides the highest level of precision out of all three models, emphasizing its ability to accurately discern the different stages of diabetic retinopathy and correctly identify the existence of the disease in all classes. A bar chart is visualized to compare the precision of these models as shown in Figure 11.

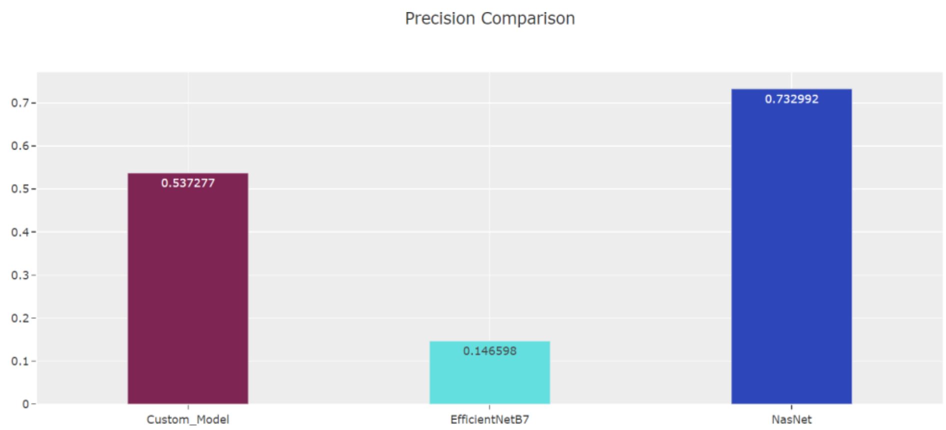

Figure 11: Comparison of Precision Score of Models for Diabetic Retinopathy Detection

6.3 Evaluation Based on Recall

The True Positive Rate, also known as Recall or Sensitivity, is the ability of the model to accurately identify all instances of one class out of all instances that belong to that class. It is particularly important in scenarios where missing any class would have severe consequences. In our multi-class challenge, the True Positive Rate was used to evaluate the ability of the models to correctly predict an instance across all classes. The Custom CNN model has a substantially lower True Positive Rate of 20%, missing many instances across the different stages of diabetic retinopathy. In contrast, both NasNet and EfficientNetB7 performed well on our multi-class challenge dataset, with a True Positive Rate of 73.3%, possessing the ability to accurately identify instances across all classes, and thereby, fewer cases will be missed. A horizontal bar chart is plotted for the comparison of recall within each model and is shown in Figure 12.

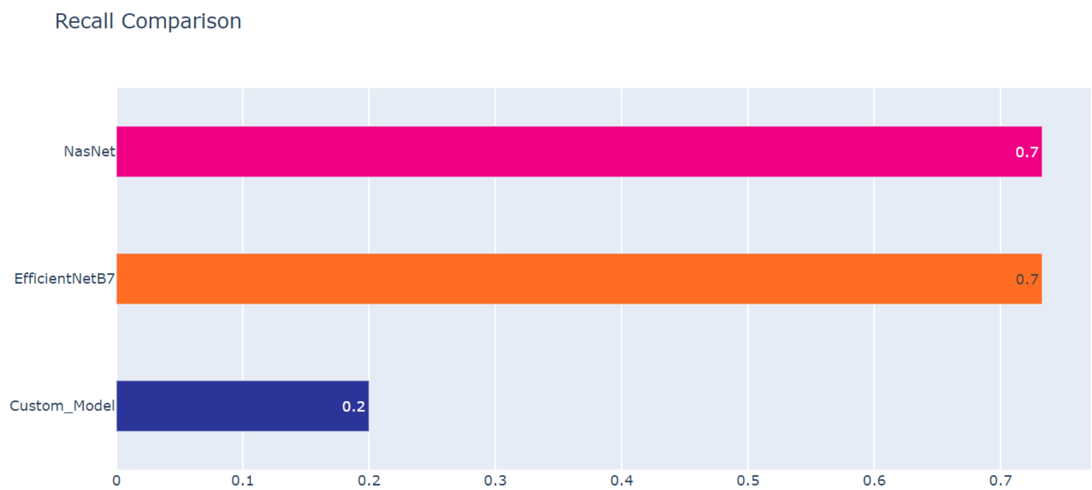

Figure 12: Comparison of Recall Score among models for Diabetic Retinopathy Detection

6.4 Evaluation Based on F1-Score

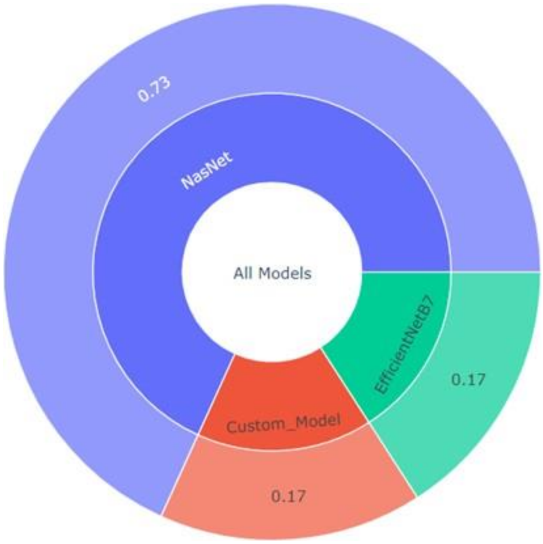

Figure 13: Comparison of F1-Score between the models for Diabetic Retinopathy Detection

In this experiment, the F1-score, which is the harmonic mean of the precision or recall gives a single value for each that balances between them. Custom model and EfficientNetB7 models give the same F1 Score of 16.91% which is much lower than the NasNet with the F1-Score of 73.3%. The lower F1-Score of custom models and EfficientNetB7 indicates a poor balance between precision and recall although NasNet shows higher values which indicates it performs better in identifying the balance between these metrics. A sunburst chart is plotted for the comparison of F1-Score between the models as shown in Figure 13.

6.5 Discussion

This study used three deep learning algorithms on retinal image data to detect diabetic retinopathy in five target classes: The terms used to classify different levels of DR are: No DR, Mild, Moderate, Severe, and Proliferative DR. The models were evaluated using four key performance metrics: They include Accuracy, Precision, Recall and F1 score. NasNet performed the highest for precision, recall, and the F1-score under the three models, making it accurate in identifying diabetic retinopathy from the images of the retina. Recall scores on EfficientNetB7 were good with relatively poor precision, which led to a low F1-score, and as a result, it seems that the model was good at distinguishing the true instance from multiple classes but had a high number of making wrong classifications. The Custom CNN had relatively okay precision levels but a very low recall rate, which lead to low F1-score. This means that whereas it may properly categorize some instances, it consistently misclassifies others, and this occurs for all the five classes thereby leading to the misclassification of the disease progression. The relatively lower performance of Custom CNN, indicates that its simple architecture hampers its capability of capturing complicated features and intra-class variations, hence there is a need to employ better architecture like NasNet. This is the reason why NasNet has shown better results in comparison to other networks because of its deep learning architecture of the network that enables it to learn patterns as well as variations within the respective classes across images. NasNet was specifically highly effective in distinguishing among multiple classes of DR hence leading to a better precision, recall and F1 scores.

One important observation in this study is that, while all the models had the same accuracy, there were statistically significant differences in their precision, recall, and F1-scores. This implicitly demonstrates that accuracy alone is not enough to tell us whether a model is effective. This is especially true when working with a multi-class classification problem like diabetic retinopathy detection. Accuracy provides a “top level” correctness, but masks problems with the class-specific performance in tasks that are difficult to classify (minority classes). While the models performed similarly, NasNet demonstrated higher precision and recall than the other models which suggests it can correctly classify the different stages of the disease, while minimizing the number of false positives and false negatives. This is a reason looking at performance metrics in combination, rather than only looking at accuracy, can allow for a better understanding of how well a model is working and how much the model’s performance can be trusted. Our example indicates that the NasNet model, on average, does a better job, and is more dependable in a meaningful way, than the other models in the challenging multi-class real-world medical task of detecting diabetic retinopathy.

7 Conclusion and Future Work

Diabetic Retinopathy (DR) affects the retina’s blood vessels which potentially leads to vision impairment and blindness if it is not treated. It is important to identify vision loss at early stages followed by necessary action. However, the diagnosis of diabetic retinopathy is a difficult process due to the early signs of the disease are hard to identify usually for an eye check from an eye specialist. Deep learning algorithms for detecting DR have also proved effective in early diagnosis, accessibility, and overall reduction of costs and load on the healthcare system. The primary challenge to diagnosing diabetic retinopathy lies in the accurate identification of various pathological signs such as microaneurysms and exudates in retinal images. In the experiment, three models: the custom CNN model, EfficientNetB7, and NasNet trained on the retinal data to detect diabetic retinopathy at early stages. These models are evaluated on four key performance metrics: accuracy, precision, recall, and f1-score. Although all the models resulted in the same accuracy of 73.3%, NasNet outperformed better compared to other algorithms with the highest precision, recall, and F1 Score. This shows that NasNet not only predicts images accurately but also maintains a strong balance between identifying true positives and reducing false positives. The identical accuracy but varied precision, recall, and F1-score of our models further emphasize the strengths of each model in different areas of the classification task. While all models are expertly capable and correctly predict most cases, reaching the same overall accuracy, the different precision, recall, and F1-score reveal how the strength of each model may lie in different areas. The future scope of this research is to increase the volume of training data to improve the robustness of the model, use of hybrid and the strengths of the different models can be combined to get better performance. A user-friendly interface can be developed for telemedicine platforms that could make these models accessible to remote areas enabling early detection and timely intervention for diabetic retinopathy.

References

Abreu, A., Ferreira, D., Neto, C., Abelha, A. and Machado, J. (2021) 'Diagnosis of diabetic retinopathy using data mining classification techniques', *Advances in Intelligent Systems and Computing*, vol. 1352, Springer, pp. 198–209

Al Hazaimeh, O. M., Nahar, K. M. O., Al Naami, B. and Gharaibeh, N. (2018) 'An effective image processing method for detection of diabetic retinopathy diseases from retinal fundus images', *International Journal of Signal and Imaging Systems Engineering*, 11(4), p. 206.

Alyoubi, W. L., Shalash, W. M. and Abulkhair, M. F. (2020) 'Diabetic retinopathy detection through deep learning techniques: A review', *Informatics in Medicine Unlocked*, 20, p. 100377.

Argade, K. S., Deshmukh, K. A., Narkhede, M. M., Sonawane, N. N. and Jore, S. (2016) 'Automatic detection of diabetic retinopathy using image processing and data mining techniques', *Proceedings of the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 517–521.

Bhardwaj, C., Jain, S. and Sood, M. (2021) 'Transfer learning based robust automatic detection system for diabetic retinopathy grading', *Neural Computing and Applications*, 33(20), pp. 13999–14019.

Bilal, A., Sun, G. and Mazhar, S. (2021) 'Survey on recent developments in automatic detection of diabetic retinopathy', *Journal Francais d'Ophtalmologie*, 44(3), pp. 420–440.

Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., Liu, R., Wang, X., Hou, X., Liu, Y. et al. (2021) 'A deep learning system for detecting diabetic retinopathy across the disease spectrum', *Nature Communications*, 12(1), pp. 1–11.

Diabetic Retinopathy Detection — kaggle.com (n.d.) Available at: *https://www.kaggle.com/c/diabetic-retinopathy-detection/data*

García, G., Gallardo, J., Mauricio, A., López, J. and Del Carpio, C. (2017) 'Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images', *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10614 LNCS, pp. 635–642.

Hinz, T., Barros, P. and Wermter, S. (2016) 'The effects of regularization on learning facial expressions with convolutional neural networks', pp. 80–87.

Khade, S., Gite, S. and Pradhan, B. (2022) 'Iris liveness detection using multiple deep convolution networks', *Big Data and Cognitive Computing*, 6(2).

Malhi, A., Grewal, R. and Pannu, H. S. (2023) 'Detection and diabetic retinopathy grading using digital retinal images', *International Journal of Intelligent Robotics and Applications*, 7(2), pp. 426–458.

Naveen, R., Sivakumar, S. A., Maruthi Shankar, B. and Keerthana Priyaa, A. (2019) 'Diabetic retinopathy detection using image processing', *International Journal of Engineering and Advanced Technology*, 8(6 Special Issue), pp. 937–941.

Pires, R., Avila, S., Wainer, J., Valle, E., Abramoff, M. D. and Rocha, A. (2019) 'A data-driven approach to referable diabetic retinopathy detection', *Artificial Intelligence in Medicine*, 96, pp. 93–106.

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V. and Meriaudeau, F. (2018) 'Indian diabetic retinopathy image dataset (idrid): A database for diabetic retinopathy screening research', *Data*, 3(3), p. 25.

Pradeep Kumar, K. G., Karunakara, K., Thyagaraju, G. S. and Dixit, S. (2021) 'A data mining inspired methodology towards the identification of diabetic retinopathy', *IOP Conference Series: Materials Science and Engineering*, vol. 1022, p. 012082.

Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U., Khan, I. A. and Jadoon, W. (2019) 'A deep learning ensemble approach for diabetic retinopathy detection', *IEEE Access*, 7, pp. 150530–150539.

Qureshi, I., Ma, J. and Abbas, Q. (2019) 'Recent development on detection methods for the diagnosis of diabetic retinopathy', *Symmetry*, 11(6), p. 749.

Selvachandran, G., Quek, S. G., Paramesran, R., Ding, W. and Son, L. H. (2023) 'Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods', *Artificial Intelligence Review*, 56(2), pp. 915–964.

Shah, F., Khan, M., Sharif, M., Tariq, U., Khan, A., Kadry, S. and Thinnukool, O. (2021) 'A cascaded design of best features selection for fruit diseases recognition', *Computers, Materials and Continua*, 70, pp. 1491–1507.

Tymchenko, B., Marchenko, P. and Spodarets, D. (2020) 'Deep learning approach to diabetic retinopathy detection', *ICPRAM 2020 - Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods*, pp. 501–509.

Goodfellow, I., Bengio, Y. and Courville, A., (2016)**.** *Deep Learning*. MIT Press.

Tan, M. and Le, Q., (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*, PMLR 97:6105-6114.

Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V., (2018). Learning Transferable Architectures for Scalable Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.8697-8710.