DIABETIC RETINOPATHY DETECTION USING ADVANCED DEEP LEARNING ALGORITHMS

MSc Research Project
Data Analytics

Mrunali Khokale
Student ID: X22138561

School of Computing
National College of Ireland

Supervisor: Abdul Qayum

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Mrunali Narayan Khokale
Student ID:	x22138561
Programme:	MSc Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Abdul Qayum
Submission Due Date:	31/01/2024
Project Title:	DIABETIC RETINOPATHY DETECTION USING ADVANCED DEEP LEARNING ALGORITHMS
Word Count:	7523
Page Count:	18

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:	Mrunali Narayan Khoakle
Date:	31/01/2024

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):	

DIABETIC RETINOPATHY DETECTION USING ADVANCED DEEP LEARNING ALGORITHMS

Mrunali Khokale
X22138561

Abstract

Current methods of diabetic retinopathy involve manual inspection by trained professionals, which is time-consuming and requires expertise. This study aims to detect diabetic retinopathy, a major cause of blindness in people with diabetes, using advanced deep-learning algorithms. The focus of the study is on creating and testing a Transformer-CNN hybrid model against well-known architectures such as EfficientNet, ResNet50, InceptionNet, Efficient Net. Models were trained and tested using the Diabetic Retinopathy Comptetion Dataset, which included retinal images with different levels of retinopathy, to see how well these models compare based on accuracy. The results show that the ResNet50 model did better than the others with an accuracy of 80%, indicating its prowess in the classification of complex retinal images. Transformer-CNN, EfficientNet, and InceptionNet, on the other hand, had similar accuracy rates of about 73%. Selecting the right deep learning architectures for specific medical imaging tasks is very important, as shown by this variation in accuracies. The novelty of the study lies in the implementation of the Transformer-CNN hybrid model. The model did not perform significantly better than traditional architectures in this application but achieved comparable accuracy to other models, maybe owing to its smaller architecture failing to extract deep features. Research shows that advanced deep-learning techniques could be used to detect diabetic retinopathy early. This could lead to better diagnostic procedures and better patient care in the field of ophthalmology

1 Introduction

Diabetic retinopathy is an ocular complication of diabetes. It is brought on by arterial damage to the photosensitive tissue at the back of the eye (retina) Alyoubi et al. (2020)(Alyoubi et al, 2020). Diabetic retinopathy may initially manifest with mild vision impairments or no symptoms at all. It may, nevertheless, result in permanent blindness. Diabetic cases are concurrently contributing to an increase in the worldwide prevalence of diabetic retinopathy Abràmoff et al. (2010) The prevalence of diabetic retinopathy is substantial among the approximately 422 million people worldwide who have diabetes, as reported by the World Health Organization. This underscores the criticality for implementing efficacious strategies for detection and treatment (World Health Organization, 2021). Preventing vision loss requires early detection of diabetic retinopathy Alyoubi et al. (2020)(Alyoubi et al., 2020). Conventional detection approaches are labour-intensive, expensive, and susceptible to human error, as they require proficient clinicians to examine retinal images manually Alyoubi et al. (2020)(Alyoubi et al., 2020). Developing deep learning algorithms presents a potentially advantageous alternative by enabling automated, precise and efficient diagnosis Chetoui et al. (2018)(Chetoui et al., 2018). Enhanced patient care and a paradigm shift in the field of ophthalmology may result from incorporating these sophisticated algorithms into medical imaging Gao et al. (2018)(Gao et al., 2018).

The main objective of this study is:

- To design and authenticate an innovative approach utilising deep learning to identify and categorise retinal images as indicative of diabetic retinopathy.
- Concerning image analysis, the research specifically examines the potential benefits of a Transformer-CNN hybrid model, which merges the sophisticated functionalities of attention mechanisms based on transformers with the well-established effectiveness of convolutional neural networks.

The primary research question guiding this investigation is:

- "How does the performance of a hybrid Transformer-CNN model compare to that of established deep learning architectures, namely InceptionNet, EfficientNet, and ResNet50, in diabetic retinopathy detection?"

The potential transformative impact of this research on the diagnosis of diabetic retinopathy is its primary significance. In image analysis, sophisticated deep learning models provide improved precision and efficacy, including Transformer-CNN hybrids, InceptionNet, Efficient Net and ResNet50. These models can detect subtle alterations in retinal images that may elude human vision, allowing for earlier and more precise diagnoses. Detection at an early stage is crucial for preserving vision and preventing the progression of diabetic retinopathy.

By presenting an innovative approach for detecting diabetic retinopathy utilising a Transformer-CNN hybrid model, this study makes a valuable contribution to the field. This research investigates the performance of this model in comparison to well-established architectures such as InceptionNet, EfficientNet, and ResNet50. By investigating the potential synergies between convolutional neural networks (CNNs) and transformer-based attention mechanisms, the novelty is introduced. This is achieved through the utilisation of hierarchical feature extraction and sequential modelling. By establishing fresh standards in medical image analysis, this novel methodology intends to yield significant knowledge that will guide subsequent advancements in the domain.

The implications of this research go beyond the detection of diabetic retinopathy. Potential advances in the diagnosis and treatment of a wide range of diseases may result from the application of the methodologies and findings to additional domains of medical imaging. This research not only makes a scholarly contribution to the domain of ophthalmology but also presents novel opportunities for investigation in the wider field of medical technology through the expansion of established methodologies.

This chapter introduced the reader to the topic, the motivation of undertaking the study, and the contribution and significance of the study. The following section deals with the critical analysis of the state of the art of the study followed by the methodology adopted for the study. The architecture of the system follows and the results are discussed in the subsequent chapter before concluding the report.

2 Background

The prevalence of diabetes mellitus, in recent times, has achieved epidemic proportions, adding to a hike in related complications, significantly diabetic retinopathy (DR) (Pandova, 2019). DR, an eminent microvascular intricacy impacting the retina, proves to be an important cause of visual impairment, and blindness globally. "Diabetic Retinopathy (DR)" is a condition associated with the eye, as a result of chronic diabetes. "DR" is the raging reason behind blindness among adults who belong to working age, throughout the globe and it can impact more than 93 million individuals. During this health concern, "advanced deep learning algorithms" have risen as efficient tools for early and precise detection of "DR", exhibiting a paradigm shift in diagnostic methodsNguyen et al. (2020)(Nguyen et al., 2020).

Review of the existing literature

According to Nguyen et al. (2020)Nguyen et al. (2020) Over 93 million individuals worldwide suffer from "diabetic retinopathy (DR)", the primary cause of blindness in working-age persons with diabetes. The "digital retinal scans" should be manually examined early detection is of equal importance. This study presents an automatic classification method to analyse fundus photos with different fields of view and lighting using machine learning models such as "CNN", "VGG-16", and "VGG-19".

The automated method used in this effort was motivated by "diabetic retinopathy (DR)", a potentially fatal consequence of diabetes that is typically detected by ophthalmologists. A model was developed on a large "Kaggle dataset (APTOS)" of 3662 high-resolution fundus pictures using "Deep Learning", especially "DenseNet", to classify DR phases 0 through 4. An accuracy of 0.9611 and a "quadratic weighted kappa" score of 0.8981 were produced by the impressive performance of the "DenseNet Architecture" as stated by Brik et al. (2022) Mishra et al., 2020. The output is produced by the activation function of the model, and the method entails extracting features from patient fundus pictures.

This paper by Doshi et al. (2016) Gangwar et al., 2021, addresses the urgent problem of diabetic retinopathy, which is the primary cause of blindness in people between the ages of 20 and 65. Using a "deep learning" hybrid model that applies transfer learning to the already trained "Inception-ResNet-v2" as a unique method. A "bespoke block" of "Convolutional Neural Network (CNN)" layers built on top of "Inception-ResNet-v2" is included in the model. Tests on the APTOS 2019 "blindness detection (Kaggle dataset)" and the "Messidor-1" diabetic retinopathy dataset show better performance than previous findings.

The difficulties evaluated by Qummar et al. (2019) Qummar et al., 2019, associated with manual identification of the sight-threatening condition of "diabetic retinopathy (DR)" led to the invention of an automated detection technique. This study uses a collection of five deep "Convolutional Neural Network (CNN)" models such as "Resnet50, Inceptionv3, Xception, Dense121, and Dense169" to improve classification accuracy throughout the five phases of DR using a "Kaggle dataset".

This study by Alyoubi et al. (2020) Tsiknakis et al., 2021, looks at the worldwide problem of diabetic retinopathy, a major diabetes-related cause of blindness. The study thoroughly examines the diabetic retinopathy diagnosis pipeline utilising "deep learning techniques" on fundus retinal pictures with an emphasis on artificial intelligence solutions. Reviewing lesion localization, grading, preprocessing methods, models, datasets and actual clinical applications emphasising the value of early diagnosis and therapy.

This research by Tymchenko et al. (2020)Tymchenko et al., 2020, presents an autonomous deep-learning strategy using "convolutional neural networks (CNNs)" to address the crucial requirement for early identification in diabetic retinopathy. The strategy takes into account difficulties such as inter-doctor variability and dataset costs and uses a new multistage transfer learning methodology. With a quadratic weighted kappa score of 0.925466 on the APTOS 2019 Blindness Detection Dataset, it stands out among 2943 competing approaches and demonstrates its efficacy in handling the intricacy of diabetic retinopathy staging from a single fundus shot.

This work introduces an optical "coherence tomography (OCT)" image-based diagnostic method for early and automated identification of "diabetic retinopathy (DR)," spanning grades 0 and 1. DR is a global health concern as stated by Li et al. (2019) Li et al., 2019. To extract characteristics from raw OCT pictures and retinal layer data, a unique deep network called "OCTD_Net" is constructed. The network demonstrates effectiveness in early DR detection with an amazing accuracy of 0.92, sensitivity of 0.90, and specificity of 0.95. Significant alterations in the retinal layers are shown by analysis for individuals with grade 1 DR, highlighting the potential of "OCTD_Net" to support prompt and precise diagnosis, thereby lowering the risk of vision loss.

According to Bora et al. (2021) Bora et al., 2021, these developments validate two "deep-learning algorithms" to address the difficulty of increasing the screening rate for "diabetic retinopathy (DR)". These techniques use one-field or three-field colour fundus pictures to forecast the likelihood that diabetic patients will develop DR within two years. In internal validation, the one-field system can be used to achieve an AUC of 0.70, whereas the three-field system received an AUC of 0.79. Predictive accuracy was greatly increased when available risk indicators were combined with these deep learning systems. The study highlights how these systems can help patients with diabetes achieve better vision-related results at a lower cost by optimising screening intervals, improving risk stratification, and improving vision-related outcomes.

This study by Qiao et al. (2020) Qiao et al., 2020, focuses on "microaneurysms" in fundus imaging to solve the persistent problem of early "diabetic retinopathy (DR)" identification. Increased blood sugar levels are a factor in microvascular issues that result in permanent eyesight loss. The suggested system uses convolutional neural network methods that are accelerated by GPU for effective medical picture processing, using deep learning. Ophthalmologists may more easily grade fundus pictures as early "nonproliferative diabetic retinopathy (NPDR)", moderate NPDR, or severe NPDR with the use of semantic segmentation, which separates diseased and normal fundus images. With its improved efficiency and accuracy in NPDR prediction, this automated method presents a viable option for the early detection and prognosis of microaneurysms in diabetic retinopathy.

According to the research conducted by Gao et al. (2018) Gao et al., 2018, one of the major causes of blindness in diabetic individuals is "diabetic retinopathy (DR)", which needs prompt administration by routine fundus photography exhibition. The researchers created a collection of labelled "DR" fundus photos in response to the need for an effective exhibition. The researchers then used "deep convolutional neural network" models to automatically diagnose patients and offer treatment. The models, which graded "DR" severities with an accuracy of 88.72%, were put to use on a cloud platform to provide hospitals with prototype diagnostic methods. The establishment demonstrated a high consistency rate of 91.8% with ophthalmologists in the clinical evaluation, demonstrating the method's efficacy in improving the diagnosis and management of "DR" worldwide.

According to the research conducted by Kandel and Castelli (2020)Chen et al., 2018,"diabetic retinopathy (DR)" is a common complication of diabetes which is one of the leading causes of blindness globally. "DR" is hard to detect in the early stages the diagnostic procedure can be time-consuming and abundant expertise is required. The researchers proposed a computer-aided diagnosis method based on a "deep learning algorithm" to automatically diagnose "DR" and divide colour retinal fundus photographs into five levels. A novel pre-processing algorithm is adopted to improve the quality and uniformity of input retinal images and a transfer learning method to achieve better administration. The method is evaluated based on a test set with 7023 images an accuracy of 80.0% and a kappa score of 0.64 is acquired.

According to Chetoui et al. (2018)Chetoui et al., 2018, if left untreated, "diabetic retinopathy (DR)", a disease linked to diabetes that damages the retina, can result in blindness globally. Previous indicators that need prompt diagnosis include "haemorrhages", "hard exudates", and "micro-aneurysms (HEM)", irrespective of age. This job enhances upon the performance of traditional "Local Binary Pattern (LBP)" by introducing additional texture characteristics for "DR detection": "Local Ternary Pattern (LTP)" and "Local Energy-based Shape Histogram (LESH)", methods. The gathered histogram characteristics are classified using "Support Vector Machines (SVM)" techniques. "LESH" applies "SVM" with a "Radial Basis Function kernel" to get the greatest precision. "LESH" with "SVM-RBF is the best performance, according to the "ROC" curve study, with an excellent "Area Under Curve (AUC)".

According to the research conducted by Doshi et al. (2016) Doshi et al., 2016, the retina is harmed by "diabetic retinopathy", which has impacted up to 80% of diabetics for more than a decade. It is hard to diagnose this illness in places where technology and experience are not strong. This research presents a "GPU-accelerated deep convolutional neural network" for the automatic diagnosis and categorization of diabetic retinopathy into five severity stages, in contrast to previous work that concentrated on disease detection or manual feature abridgement. On the "quadratic weighted kappa metric", the ensemble of three identical models achieves an accuracy of 0.386, and enhances performance to 0.3996, hence proving their trait. This technique aims to diagnose diabetic retinopathy more quickly and automatically, particularly in situations with limited resources when it is desperately required.

Pires et al. (2019)Pires et al. (2019) investigated a data-driven technique to aggregate representations that are directly extracted from the retinal images. On top of this the authors implemented robust feature-extraction and augmentation techniques along with multi-resolution training in implementing deep learning algorithms. The authors tested a range of machine learning models in classifying the extracted features from the retinal images ranging from the Random Forest and Neural Networks to Transfer learning approaches. The evaluation of the method using AUC-ROC, shows a significant value of 98.2% on the Messidor-2 dataset.

According to the research conducted by Gargeya and Leng (2017) Gargeya and Leng, 2017, the international issue of "diabetic retinopathy (DR)" leading to avoidable blindness is the focus of this research. A "deep learning programme" analyses colour fundus photos to automate the "DR" exhibition. It performs this by differentiating between healthy and "DR-affected" instances for quick medical recommendations. The prototype, which was trained using fundus pictures of 75,137 diabetic patients, gets a cross-validation score of 0.97 "AUC" with 94% sensitivity and 98% uniqueness. "AUC" results of 0.94 and 0.95 are obtained via external validation on the "MESSIDOR 2" and "E-Ophtha" datasets. The data-driven “AI” approach works well for screening, providing a high degree of reliability in detecting instances that need to be evaluated by an ophthalmologist and perhaps lowering the overall amount of vision loss caused by “DR” ailments.

Li et al. (2019)Li et al. (2019) in their research implemented InceptionNet-V3 through transfer learning to classify the retinal photographs in order to automatically detect the diabetic retinopathy. They experimented with two sets of data. One set involves the retinal photographs acquired from 5278 adult patients amounting to 19233 images. The other set of data involves the Messidor-2 dataset acquired from Kaggle. With the incorporation of 10-fold cross validation technique, the technique achieved an impressive accuracy of 93.49

This research by Priya and Aruna (2013) Priya and Aruna, 2013, concentrates on the early diagnosis of “diabetic retinopathy (DR)”, an eye condition associated with the advancement of DR. Three samples including “Bayesian Classification, Support Vector Machine (SVM), and Probabilistic Neural Network (PNN)” are used to diagnose “non-proliferative (NPDR)” and “proliferative diabetic retinopathy (PDR)”. 350 fundus pictures are applied (100 for training, 250 for testing) to extract characteristics such as “blood vessels, haemorrhages for NPDR, and exudates for PDR” using image processing. The outcomes show that “SVM” is superior to “PNN” (89.6%) and “Bayesian Classifier” (94.4%), with an accuracy of 97.6%, as suggested by the results. With a precision of 95.38%, external validation on 130 pictures from the ”DIARETDB0” database confirms the effectiveness of “SVM” systems.

A study by Kalyani et al. (2023) Kalyani et al. (2023) involved the implementation of the capsule neural networks to detect the diabetic retinopathy from the retinal images. The authors implemented the system to detect the retinopathy using the Messidor dataset. The model was tested for classification of different stages of the diabetic retinopathy viz. healthy, stage 1, stage 2, and stage 3. The methodology achieved accuracy of 97.98% for healthy retina, 97.65% for stage 1 and stage 2 classification, and a better accuracy of 98.64% for the stage 3 fundus images.

According to the research conducted by Gulshan et al. (2016) Gulshan et al., 2016, to develop an algorithm for the automatic diagnosis of diabetic retinopathy and diabetic macular oedema in retinal fundus photos, this work uses “deep learning”, more especially a “deep convolutional neural network” samples. After being taught on an extensive dataset and verified on two distinct datasets, the system exhibits impressive sensitivity and uniqueness. It recommends an area under the receiver operating curve of 0.991 and 0.990 for the two datasets for the detection of referable “diabetic retinopathy” samples.

3 Methodology

This section of the report details the methodology adopted for the study based on the thorough literature review conducted related to the study. This methodology is a combination of different modules that work in tandem.

The methodology flow is depicted in Figure 1 below. The various modules and their working are explained further.

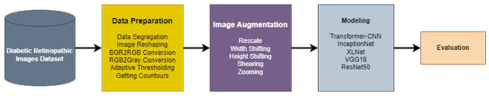


Figure 1: Methodology flow for the study

3.1 Data Collection

The dataset used for the study is a smaller version of the Diabetic Retinopathy Detection Competition Dataset Pires et al. (2019) (Pires et al, 2019) collected from the Kaggle repository. The dataset is divided into two parts: first is the collection of the Diabetic Retinopathic Images, and second is a CSV file containing the Training Labels Pires et al. (2019) (Pires et al, 2019). These labels provide the class of Retinopathy present in the particular image. Figure 2 below shows the sample images from the dataset.

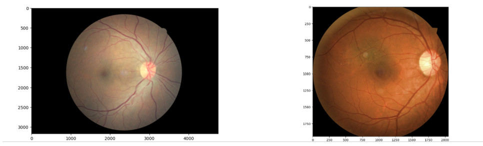
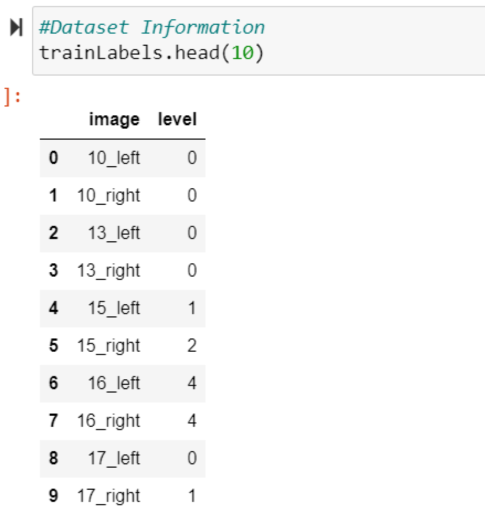


Figure 2: Sample images from the dataset

The Figure 2 illustrates the blackish area in the images represents the presence of Retinopathy in the eyes. The contents of the CSV file in the dataset are shown in Figure 3 below.



```
#Dataset Information
trainLabels.head(10)
```

	image	level
0	10_left	0
1	10_right	0
2	13_left	0
3	13_right	0
4	15_left	1
5	15_right	2
6	16_left	4
7	16_right	4
8	17_left	0
9	17_right	1

Figure 3: Contents of 'trainLabels.csv' file

The first column in the file represents the image name, whereas the level column represents the severity of the Retinopathy present in the corresponding image. The 'right' and 'left' text in the image name mean the position of the Retinopathic Features relative to the Retina.

The severity of the retinopathy can be identified in Table 1 below.

Table 1: Severity of the Retinopathy

Serial Number	Level	Severity
1	0	No Diabetic Retinopathy Present
2	1	Mild Retinopathy
3	2	Moderate Retinopathy
4	3	Severe Retinopathy
5	4	Proliferative Retinopathy

3.2 Data Preparation

Once the data is collected, it must be prepared for the modelling part. Several steps are performed in the process of Data Preparation. These steps are explained in brief below.

3.2.1 Data Segregation

The files in the dataset need to be read using the names in the CSV file. This makes obtaining the class names difficult. In this step, the images in the dataset are grouped based on the severity of the Retinopathy present in the images. This is done to obtain a better exploration of the data relative to the class that it belongs to. Once the segregation is complete, the data is subjected to resizing.

3.2.2 Image Resizing

The images present in the dataset are not of similar sizes as can be seen in Figure 2. For the deep learning models to learn over these images reliably, it is beneficial for these images to be of equal sizes, that is why the images are resized to be of shape 256x256. Once resized, these images can be processed further.

3.2.3 BGR2RGB Conversion

The images are read into the workspace using the CV2 library for Python. The CV2 library, by default, reads the images into the workspace in Blue-Green-Red colourspace. To process these images further, an intermediate conversion of the images to a Red-Green-Blue colourspace is essential. This conversion of the colourspace helps in visualising the images better.

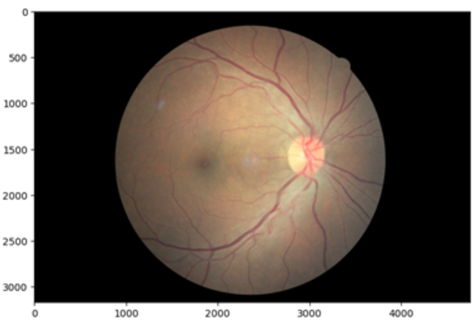


Figure 4: RGB converted image

3.2.4 RGB2Gray Conversion

Once the RGB version of the image is obtained, it is converted to a grayscale image, merging the three channels in an RGB into a single channel with a maximum value of 256. This allows for the thresholding of the image.

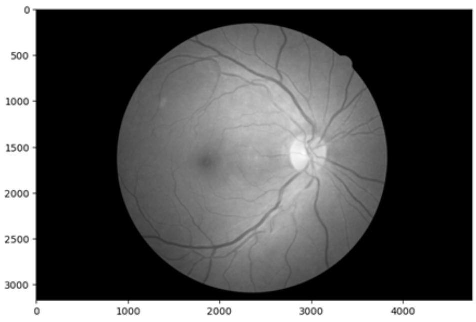


Figure 5: Grayscale image

3.2.5 Adaptive Thresholding

Adaptive thresholding is a technique of thresholding an image to identify the most prominent part of the image by assuming that the local regions of an image will be more uniform. Once the thresholding of the images is performed, the contours from the images can be obtained. These contours finally give a segmented image.

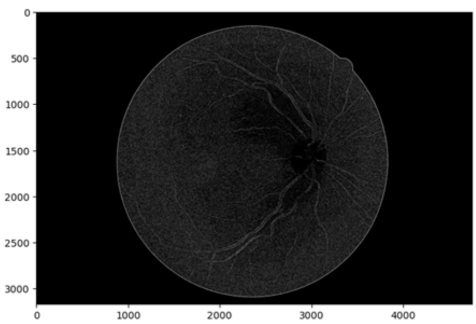


Figure 6: Image after adaptive thresholding

3.3 Image Augmentation

For implementing machine learning algorithms for data analysis, the data must be class-balanced to avoid any model bias (reference is missed).The dataset under study exhibits some class imbalance, as seen in Figure 7.

The figure shows that there are around 12000 images belonging to the No Diabetic Retinopathy (NoDR) class, 1100 images of the 'mild' class, about 2500 images of the 'moderate' class and a small number of images for the 'severe' and 'proliferative' class. It becomes essential to balance these classes to avoid biased data learning. This class balancing is achieved through several image augmentation techniques that generate additional

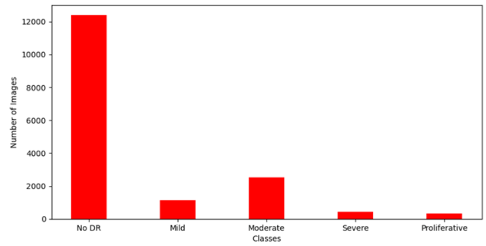


Figure 7: Class imbalance in the dataset

images from the dataset using spatial operations on the images Shorten and Khoshgoftaar (2019) (Shorten and Khoshgoftaar, 2019). Operations such as width shifting and height shifting are performed on these images to generate augmented images. The values of the augmentations performed are shown in Table 2 below.

Table 2: Image Augmentation Values

Serial Number	**Augmentation Technique**	**Values**
1	Width Shifting	0.2
2	Height Shifting	0.3

The width shifting involves shifting the image horizontally by the amount provided by a value whereas the height shifting shifts the image vertically. This process moves the contents in the image in respective axis helping the models learn the data from alternative perspective.

Figure 8 below depicts some of the augmented images.

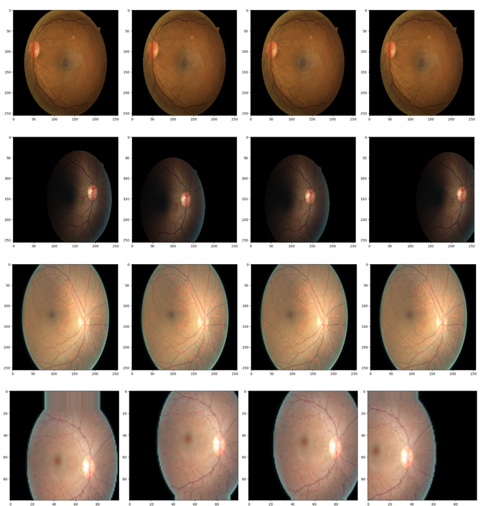


Figure 8: Augmented images

3.3.1 Conclusion

This chapter explained in detail the methodology adopted for the study undertaken with great emphasis on the data collection and data preparation phases. The models that are implemented in the study are also mentioned in the section. The following section of design and specifications presents an overview of the architecture of the system for the detection of the diabetic retinopathy from eye images

4 Design Specification

This section of the report deals with the architecture of the system developed for detection of the severity of the Diabetic Retinopathy. Through the augmentation of the retinopathic images two sets of data are obtained viz. training and testing. The training set of the data contains 16802 images whereas the testing set contains 1678 images after the augmentation. The models are trained on the training set of the data and tested on the testing set.

The overall system architecture consists of two tiers viz. client tier and business tier. The client tier is responsible for the presentation of the dataset files, this helps to observe the data properties such as size, image format etc. The second tier, the business tier is responsible for all the processing done in the study. It consists of reading the CSV file from the dataset from which the images in the dataset are read; the pre-processing steps of image resizing and data augmentation are also performed in the tier before modelling them using the models such as Transformer CNN hybrid, Efficient Net, ResNet50, and the Inception net. The implementation of the models is given in Chapter 5 of the report.

The design process flow of the system is shown in Figure 9 below

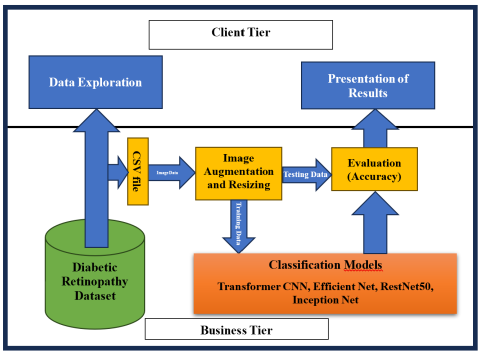


Figure 9: System Design Flow

5 Implementation

This section of the report delves into the implementation of the study. The presented study is implemented using the Python programming language using the Jupyter Notebook environment in Anaconda for Windows. Prominent libraries such as Pandas, glob, and Tensorflow are used in the implementation. Pandas and glob are used for reading the dataset files whereas the Tensorflow library is used for modeling the data.

5.1 Transformer-CNN hybrid

Transformer-CNN architecture, is a novel approach in the field of deep learning, for tasks involving image classification. The best characteristics of Convolutional Neural Networks (CNNs) and Transformer models are combined in this architecture. CNNs are known for quickly processing image data by using convolutional filters to extract features. Transformers, on the other hand, were originally made for processing natural language, and their attention mechanisms make them great at capturing long-range dependencies and contextual relationships Vaswani et al. (2017) (Vaswani et al., 2017).

In the case of image classification, the Transformer-CNN architecture aims to use CNNs' ability to extract local features and Transformers' ability to understand global context. This mixed method works really well for jobs that need to understand both small details and big-picture relationships, like medical image analysis.

The given code sets up the Transformer-CNN architecture with encoder and decoder parts that are separate from each other. The encoder starts with an input layer that is set up by Input(shape=input shape) and is shaped to fit the images. The input is then sent to a dense layer (Dense(3, activation="relu")) and then to an attention layer (Attention(score mode="dot")), which uses dot-product attention to fix attention to certain parts of the input. There is also a connection in this part that adds the attention layer's output to the original input using Add (). The next step is layer normalisation (LayerNormalization(epsilon=1e-6)) to make sure the learning process stays stable.

The encoder's feed-forward network has two dense layers that are activated by linear functions and ReLU functions, respectively. It also has a residual connection that takes the output of the previous layer and adds the output of these dense layers to it. The output of the layers is then normalised.

A separate target input is added to the decoder part of the architecture. A scaling factor is added to the target input to use positional encoding. This tells the model where the pixels are. The decoder has residual connections and layer normalisation, which are also parts of the encoder. It also has a dense layer and an attention layer.

After the decoder's output is flattened, it is sent through a final dense layer that is activated by ReLU. The model is defined by the function Model(inputs=input, outputs=x), which also shows how data moves from the input layer to the output layer. The Adam optimizer and the categorical cross-entropy loss function are then used to put the model together. It can be used for tasks involving multiclass classification.

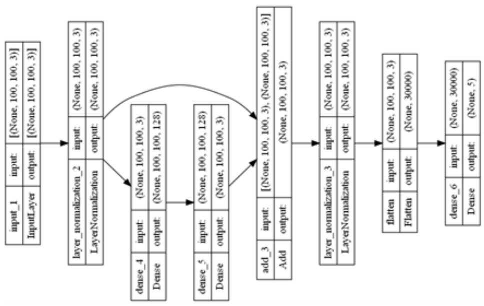


Figure 10: Architectural summary of Transformer-CNN hybrid model

5.2 Efficient Net

EfficientNetV2L is a type of the EfficientNet family of deep learning models, which is famous for its efficient performance. This model, distinguished by the "V2L" indicating its second version, enhances the original EfficientNet's design through optimised scaling and convolution techniques with more layers and parameters that make it better at learning and more accurate, but it also needs more computing power. Since EfficientNetV2L was trained on the large ImageNet dataset, it is very good at extracting features. This makes it very flexible and good for many image classification tasks Tan and Le (2019) (Tan and Le, 2019). Its balanced approach to scaling network dimensions (depth, width, resolution) allows it to achieve high performance with relatively lower computational costs, positioning it as a popular choice in both research and practical applications.

The model is built on top of EfficientNetV2L, that has been modified for this study. The model is trained on the ImageNet dataset and is used in the study through a transfer learning approach. The top layers of the base model are omitted (include top=False), which lets custom layers to be added. The model's output goes through a GlobalAveragePooling2D layer, which turns the spatial data into a flat vector. This makes the feature map easier to understand while still keeping important data. To prevent overfitting, a Dropout layer with a 5% drop rate is added. During training iterations, this layer randomly removes some of the input units. After that, a dense layer with 8 units and sigmoid activation is added, which combines non-linear features. The architecture ends with a dense layer that has 5 units and a softmax activation, for the multi-class classification problem. The Adam optimizer is used to compile the model. This optimizer is used because it works quickly and doesn't use much memory. The loss function is categorical crossentropy, which works well for multi-class classification.

5.3 ResNet50

ResNet50 is a well-known deep learning model from the Residual Network (ResNet) family. It is known for being innovative in classifying images. The main thing that makes ResNet models unique is that they use residual connections, which let data skip over some network layers. This architecture makes the vanishing gradient problem a lot easier to deal with, which makes it possible to train deeper neural networks successfully. ResNet50 is a 50-layer network that finds a good balance between model depth and computational efficiency. This model is widely used in image classification tasks because it can learn complicated patterns without requiring a lot of computing power. ResNet50 has been pre-trained on the ImageNet dataset and has powerful feature extraction that let it be used for a wide range of image classification tasks He et al. (2016) (He et al., 2016).

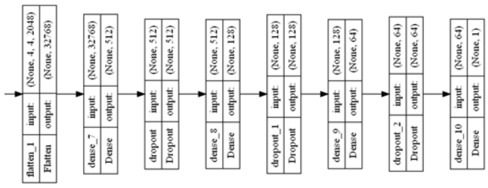


Figure 11: Custom Layers of the Model

The ResNet50V2 model is modified for the classification task by getting rid of the default top layers and using ImageNet weights. After the base model, custom layers are added to handle the output of the model. After going through a Flatten() layer, the output is changed by a set of dense layers with tanh activations. The first dense layer has 512 units, and then there is a dropout layer with a rate of 0.5. After these dense layers with 128 and 64 units, dropout layers with rates of 0.01 and 0.05 are added. The last layer is dense layer and has a single unit and a sigmoid activation function.

5.4 Inception Net

InceptionNet, especially its InceptionV3 version, is a well-known deep convolutional neural network efficient at classifying images. InceptionNet is able to capture a wide range of feature sizes thanks to its Inception modules, which use convolutional filters of different sizes in the same layer. The architecture of this system lowers the computational load and parameters, which speeds up training and lowers the chance of overfitting. The latest version, InceptionV3, improves these features even more by factorising convolutions and making computations more efficient Szegedy et al. (2016) Szegedy et al. (2016). InceptionV3 is widely used in many computer vision tasks because it is fast, accurate, and well-trained on the ImageNet dataset.

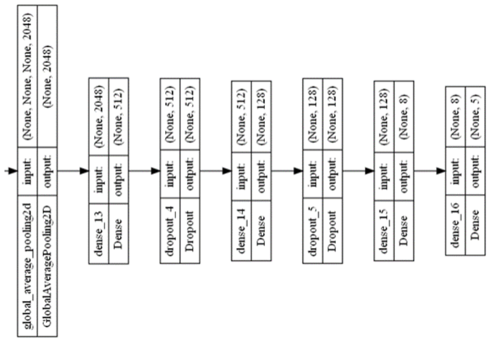


Figure 12: Custom layers added to the Inception Net

Similar to other models in the study, InceptionV3 architecture of the Inception Net pre-trained on the ImageNet dataset is used through transfer learning. The top layers of the models are omitted to add custom layers to model for classifying the images in 4 classes against 1024 in ImageNet. After the base the model's output custom layers are added. These layers include a set of dense layers with different units and activation functions. A GlobalAveragePooling2D layer is followed by multiple dense layers. First of them is a dense layer with 512 units and a 'relu' activation followed by a dropout layer with rate of 0.15. The dropout layer is followed by another dense layer of 128 and 8 units respectively are added. They all have a sigmoid activation function and are connected to dropout layers with a rate of 0.05.

The assembled model, combining the InceptionV3 foundation with the custom top layers, undergoes compilation with the Adam optimizer. It employs 'categorical_crossentropy' as the loss function, aligning with its multi-class classification nature, and uses accuracy as the performance evaluation metric.

6 Evaluation

The study aimed to leverage advanced deep learning models such as Transformer CNN, EfficientNet, ResNet50, and InceptionNet. The goal was to accurately classify images to assist in the early detection of this diabetic retinopathy, which is crucial for effective treatment and management. The models are evaluated based on the accuracy that they obtained in classifying the images pertaining to diabetic retinopathy in 5 distinct classes. The models are first trained on the training data after which they are evaluated on the test dataset to obtain the accuracy.

Table 3: Performance evaluation of the implemented models

Serial Number	**Model**	**Accuracy**
1	Transformer CNN	73.778307
2	ResNet50	79.999959
3	InceptionNet	73.778307
4	EfficientNet	73.778307

With an accuracy rate of almost 80%, ResNet50 stood out as having the best performance. This means that its structure might be better suited to the specifics of retinal imaging. This could be because of its depth and residual learning framework. However, because medical diagnostics are so important, even this model that works the best could be better.

The accuracy levels of the Transformer CNN, EfficientNet, and InceptionNet models were all about 73.7%. This level of performance is good, but it might not meet the clinical standards for diagnostic tools. It shows that these architectures might not be able to handle the complexity of medical images, which need to find small, complicated patterns all the time. The novel models of Transformer CNN presented has hence performed at par with the other advanced deep learning models.

However, the results provoke critical considerations:

The different model performances bring up an important point about AI in healthcare that the models need to be both technically advanced and able to work with the unique properties of medical data. It's not always true that more complicated models are better for clinical use. For example, the Transformer CNN's complicated mechanisms, while interesting in theory, did not perform much better than simpler architectures like EfficientNet in this case however it was not as bad.

Because medical diagnoses are so important, the interpretability of these models are necessary. The wide range of accuracy seen makes people wonder about the dependability and consistency of medical diagnoses made by deep learning models.

The Trasnformer-CNN model combines the global contextual understanding of transformers with the localised feature extraction skills of CNNs in a way that works well together. This kind of hybrid approach is fairly new in the field of ophthalmology and is a big step forward in the field of medical image analysis.

Further improvement of the model can yield better results comparable to the other state-of-the-art in the field of Diabetic Retinopathy Detection from retinal scans.

7 Conclusion and Future Work

The study into how advanced deep learning algorithms can be used to diagnose diabetic retinopathy is a big step forward in the area where artificial intelligence and medical imaging meet. The research is based on creating a Transformer-CNN hybrid model, which is a new way to make the process of diagnosing diabetic retinopathy better. The study did a good job of comparing how well this new model worked with well-known deep learning architectures like ResNet50, InceptionNet, and EfficientNet. The current comparison study has not only shown that these models work well in a medical setting, but it has also given us a better understanding of how well they can do certain medical imaging tasks.

The remarkable performance of the ResNet50 model demonstrates that it is suitable for processing complex image data, including retinal scans. On the other hand, the comparable accuracy levels shown in Transformer-CNN, EfficientNet, and InceptionNet suggest that these models may find use in the diagnostics industry.

The utilisation of advanced deep learning algorithms for the identification of diabetic retinopathy signifies a noteworthy progression in the fields of artificial intelligence and medical imaging. The creation of a Transformer-CNN hybrid model that uses a novel method to help with the early diagnosis of diabetic retinopathy is essential to this study. This study successfully evaluated this new model's performance against popular deep learning architectures such as ResNet50, InceptionNet, and EfficientNet. In addition to validating these models' performance in a medical aspect, this comparison analysis has improved the understanding of their possible uses for certain medical imaging tasks.

ResNet50 model works remarkably well on complicated medical image data such as retinal scans. In the meanwhile, the accuracy scores of Transformer-CNN, EfficientNet, and InceptionNet are comparable, indicating that they could be helpful in medical diagnosis.

Notwithstanding its successes, the study has a number of drawbacks. The study's dependence may impact the results' generalizability on a small dataset. Increasing the size of the dataset to include a wider range of retinal pictures may improve the models' resilience. The lack of comprehensive real-world testing limits the direct use of these models in clinical settings, where variables like patient circumstances and equipment variations are important considerations. Because the models are sophisticated, they need a large amount of processing power, which might restrict their use and scalability in contexts with limited resources.

The primary research question of the study was: "How does the performance of a hybrid Transformer-CNN model compare to that of established deep learning architectures, namely InceptionNet, EfficientNet, and ResNet50, in diabetic retinopathy detection?"

The research has successfully answered this question, which demonstrates that while the Transformer-CNN model has some promise, it does not considerably outperform the ResNet50 model. This is an important study because it shows that while combining CNN and transformer features is novel, in certain medical imaging scenarios it may not necessarily perform better than more conventional deep learning architectures.

Future Work

Future work in the field of detecting diabetic retinopathy can make use of more diverse and extensive dataset that can be used for training and validation of the presented Transformer-CNN architecture of deep learning. This shall help in identifying the generalizability of the model.

If this model can be tested in a real-world medical setting, its ability to detect retinopathy in real time could be explored helping the healthcare professional in faster and early diagnosis of the disease.

The presented transformer-CNN model can be tested on different medical applications to identify its reliability and use cases.

References

Abràmoff, M. D., Reinhardt, J. M., Russell, S. R., Folk, J. C., Mahajan, V. B., Niemeijer, M. and Quellec, G. (2010). Automated early detection of diabetic retinopathy, *Ophthalmology* **117**(6): 1147–1154.

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**: 100377.

Bora, A., Balasubramanian, S., Babenko, B., Virmani, S., Venugopalan, S., Mitani, A., de Oliveira Marinho, G., Cuadros, J., Ruamviboonsuk, P., Corrado, G. S. et al. (2021). Predicting the risk of developing diabetic retinopathy using deep learning, *The Lancet Digital Health* **3**(1): e10–e19.

Brik, Y., Attallah, B., Aiche, I., Lahmar, H. and Zohra, Z. (2022). Deep learning-based framework for automatic diabetic retinopathy detection, *2022 32nd International Conference on Computer Theory and Applications (ICCTA)*, IEEE, pp. 129–134.

Chetoui, M., Akhloufi, M. A. and Kardouchi, M. (2018). Diabetic retinopathy detection using machine learning and texture features, *2018 IEEE Canadian conference on electrical & computer engineering (CCECE)*, IEEE, pp. 1–4.

Doshi, D., Shenoy, A., Sidhpura, D. and Gharpure, P. (2016). Diabetic retinopathy detection using deep convolutional neural networks, *2016 international conference on computing, analytics and security trends (CAST)*, IEEE, pp. 261–266.

Gao, Z., Li, J., Guo, J., Chen, Y., Yi, Z. and Zhong, J. (2018). Diagnosis of diabetic retinopathy using deep neural networks, *IEEE Access* **7**: 3360–3370.

Gargeya, R. and Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning, *Ophthalmology* **124**(7): 962–969.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *jama* **316**(22): 2402–2410.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Kalyani, G., Janakiramaiah, B., Karuna, A. and Prasad, L. N. (2023). Diabetic retinopathy detection and classification using capsule networks, *Complex & Intelligent Systems* **9**(3): 2651–2664.

Kandel, I. and Castelli, M. (2020). Transfer learning with convolutional neural networks for diabetic retinopathy image classification. a review, *Applied Sciences* **10**(6): 2021.

Li, X., Shen, L., Shen, M., Tan, F. and Qiu, C. S. (2019). Deep learning based early stage diabetic retinopathy detection using optical coherence tomography, *Neurocomputing* **369**: 134–144.

Nguyen, Q. H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., Nguyen, B. P. and Chua, M. (2020). Diabetic retinopathy detection using deep learning, *Proceedings of the 4th international conference on machine learning and soft computing*, pp. 103–107.

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**: 93–106.

Priya, R. and Aruna, P. (2013). Diagnosis of diabetic retinopathy using machine learning techniques, *ICTACT Journal on soft computing* **3**(4): 563–575.

Qiao, L., Zhu, Y. and Zhou, H. (2020). Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms, *IEEE Access* **8**: 104292–104302.

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**: 150530–150539.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning, *Journal of big data* **6**(1): 1–48.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016). Rethinking the inception architecture for computer vision, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks, *International conference on machine learning*, PMLR, pp. 6105–6114.

Tymchenko, B., Marchenko, P. and Spodarets, D. (2020). Deep learning approach to diabetic retinopathy detection, *arXiv preprint arXiv:2003.02261* .

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need, *Advances in neural information processing systems* **30**.