Classification of Different Stages of Glaucoma Using Deep Learning Approaches

MSc Research Project

Data Analytics

Manoj Kumar Shukla

Student ID: x18195881

School of Computing

National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Manoj Kumar Shukla		
Student ID:	X18195881		
Program:	MSc In Data Analytics	**Year:**	2019/2020
Module:	Research in Computing		
Supervisor:	……		
Submission Due Date:	……		
Project Title:	Classification of Different Stages of Glaucoma Using Deep Learning Approaches		
Word Count:	7963 ……………………………	**Page Count**	21 …………………………………………………

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: Manoj Kumar Shukla

Date: 17-August-2020

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):	

Classification of Different Stages of Glaucoma Using Deep Learning Approaches

Manoj Shukla

X18195881

Abstract

Classification of glaucoma with high accuracy is most critical in slowing down glaucoma at an early stage. For the detection of glaucoma using fundus images, an analysis of an expert is required. The purpose of this research is to classify glaucoma in the Mild and severe category. A MESSIDOR dataset consists of 3200 fundus TIFF images were used. These images were converted in jpg before using for further analysis. In this study, six classification deep learning models were used. VGG16, VGG19, InceptionV3, InceptionResNetV2, ResNet50 and DenseNet16 9 to classify glaucoma in Mild and Severe categories. In this study accuracy of all the model are examined. From all the deep learning models used in this research, DenseNet169 was able to produce an accuracy of 85.19%. This strategy was successful in classifying the glaucoma stage using fundus images with highest accuracy achieved based on known previous researches.

1 Introduction

Glaucoma is an ocular eye disease, which is also called a group of eye disease. Glaucoma is an irreversible eye disease (Thakoor et al., 2019) Millions of people are suffering because of this disease. Every year numerous people around the globe lose their vision due to glaucoma. In one of the studies, it was found that 60.6 million people were affected with primary open-angle glaucoma (POAG) by 2010. furthermore, studies show that this number will go high as 80 million by the end of 2020 and 111.8 million by 2040 (Qin et al., 2019).

1.1 Research Motivation and Background

Glaucoma causes due to high intraocular pressure inside the eye. This ends up damaging the optic nerve, and damaged optic nerve form a blind spot Fig 1, damage optic nerve fibbers are hard to detect in early stages. It has been seen that by the time it is identified, the optic nerves are significantly harmed. The work of the optic nerve is to send the signal to the brain once the optic nerve is failing to do so, this started causing blackspot and partial blindness. To differentiate the normal vision and glaucoma vision can be done by more than one way, IOP calculation and cup to disk ratio value. A normal person has an IOP of 14-16 millimetres of mercury, till 20 millimetre's is considered safe, more than 20 millimetres' is falling into the glaucoma category. Stages of Glaucoma can be determined by the calculation OD and cup ratio. If the ratio goes up to 0.3 it is considered as a healthy eye, between 0.3-0.5 is called mild glaucoma, the ratio between 0.5-0.8 Severe level of glaucoma and more than 0.8 is considered as a critical level glaucoma. Glaucoma cannot be fully treated but it can be slowed down. If it can be identified at an early stage.

The convolutional neural network has shown tremendous results in the field of image classification with the help of transfer learning and various other techniques can help easily cheaply and quickly identify then at an early stage. this will help clinic trials and can work as a second opinion to any practitioner professionals.

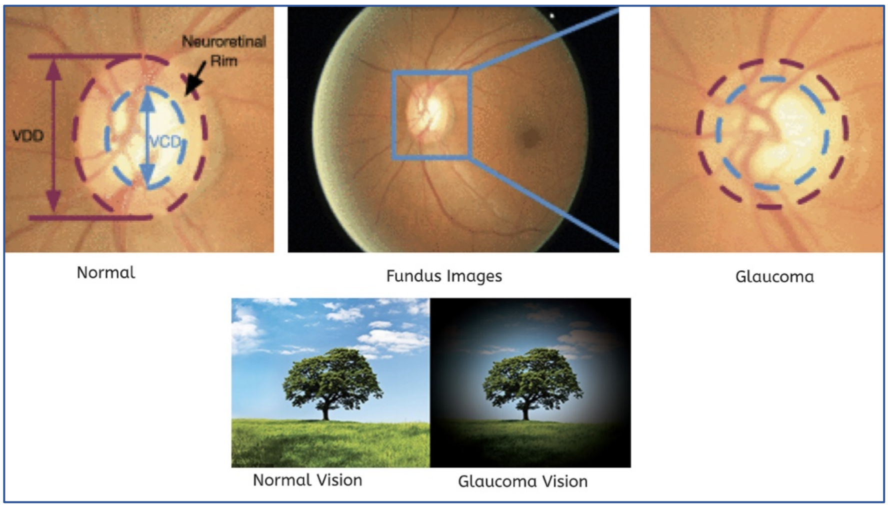


Figure 1: Fundus Images with Normal and Glaucoma vision

1.3 Research Question

This research question engraves, the necessity to perceive, interpret, and execute different classification techniques using deep learning on fundus images, in order to classify glaucoma various stages. why glaucoma classification and why deep learning? glaucoma is a chronic eye disease and the leading causes of blindness (Gaddipati et al., 2019). Classification of glaucoma in early stage helps slow it down. CNN has shown great results in image classification in recent years, it is vital to understand to what extent CNN and transfer learning can classify different stages of glaucoma using fundus images.

RQ: "How well can classification using deep learning, transfer learning and segmentation improve identification of glaucoma (mild/severe)?"

1.2 Research Objectives and Contribution

The subsequent research purpose in Table 1 will be reviewed and performed in detail in order so, that the research proposal can be acknowledged.

Table 1 Research Description and Objectives

obj	Description	Evolution Metrics
1	A critical analysis of the existing research on Glaucoma segmentation, and classification (2018-2020)	
2	Data preparation Segmentation, CDR calculation, cropping, feature extraction.	
3 (3.1) (3.2) (3.3) (3.4) (3.5) (3.6)	Implement CNN Models to classify Glaucoma stages. Implement VGG16 Implement VGG19 Implement ResNet50 InceptionV3 Implement InceptionResNetV2 Implement DenseNet169	Accuracy Loss Precision Recall f1 Score
4	A comparison of developed models (Objective 3)	
5	Comparison of implemented models with (Objective 4) vs existing models.	

2 Related Work

Glaucoma classification with fundus images is a tidy task to accomplish. Various steps would be required to work together to complete this task. An automated system would solve this issue, which can accept a fundus image and predict the glaucoma stages. Due to the recent success of the machine learning approach in the field of image-based classification, a lot of research has taken place. In this study, multiple modelling strategies will be analysed. The study will help understand, get quality results.

2.1 Cup and Disc Segmentation from Fundus Images

Optic disc (OD) and cup segmentation are an important step towards glaucoma detection. The author (Joshua et al., 2019) has utilized the advanced U-net model to segment OD value from fundus images. The data set used for this work is DRISHTI-GS. The author has cropped fundus images to the OD location. then the contrast of the images is further improved by stretching to help understand optic cup (OD) boundaries, then histogram equalization performed on the images. this technique performed much better than sliding- window models and achieve an accuracy of 99%. The author has done great work to achieve better accuracy, but dataset used in this paper has only few hundred images, accuracy can go down if we use images with a large dataset which is available in different shape size and different resolutions, whereas (Sun et al., 2018) has used a new pipeline technique, to segment OD retinal fundus images with the help of deep object detection networks. OD boundaries are measured by prognosticated bounding box into a vertical and non-rotated oval shape, ORIGA data set is used with R-CNN as an object detector because of the flexible nature and achieved an accuracy of 93.1%.

In an, another study is done on the Drishti dataset by (Manabe et al., 2019) to segment automatic optic disc and cup segmentation, using a fully connected neural network (FCN), due to this approach does not require cropping of OD region for cup prediction. this approach performs better and produces a better F1 score in comparison with the state-of-the-art solution, although the author has not mentioned any other details of this study make this study little open to discussion furthermore.

In research (Cheng et al., 2013) a technique is proposed to Segment using super pixel classification, this approach author has used a histogram and centre-surround statistics technique is used to distinguish disc or non-disc area. To evaluate the quality of the OD segmentation self-assessment reliability score is used for this study 650 images were used result produced by this study shows the area under the curve of 0.800.

The Author (Manabe et al., 2019) proposed a segmentation technique to overcome an issue of laparoscope control for cholecystectomy. In this technique, the author introduced a common transposed convolution, the flip-based subpixel reconstruction. The throughput achieved using this approach is claimed in the paper will be sufficient for a clinical trials of laparoscope robots using Nvidia GTX 1080 GPU. The author has mentioned due to insufficient learning dataset max accuracy achieved was 61.2%, results could have been better by applying some other deep learning techniques such as transfer learning or utilizing u-net modal for segmentation.

In a research (Fu et al., 2018) used multilabel deep network and polar transformation network is also called as M-Net, to segmentation optic disc (OD) and optic cup (OC). the author has used a multiscale input layer with a convolutional neural network (CNN) of U- shape. This approach forms an input of multiscale which helps achieve multiple levels receptive this helps to boost the performance of the model. for this research ORIGA, a dataset with the help of multilabel loss function segmentation of OD and OC is done together. the author has done really good work because with segmentation overfitting is always an issue having a multi-label loss function make sure overfitting will not take place.

In a research (Aloudat et al., 2018) has proposed a technique in which examining front eye images intraocular pressure (IOP) pressure can be calculated, In this research 400 frontal eye images were used and using decision tree on MATLAB six other features were extracted, and this model performed well and achieved an accuracy of 95.5%. feature extracted from frontal eye images were Pupil/Iris ratio, red area percentage, and in mean redness level of the sclera, and three novel features from the sclera contour (angle area, and distance). The author has done a great job, this technique can help clinical trials produce results in a very quick and precise way.

In a research (Patel, 2018) proposed to segment optic disc and cup ratio threshold approach was taken by the author, apart from this a binary image segmentation is also performed again on both red and green channels individually for both cup and disc values. this work is done using python and open cv. an interesting finding in this research was a combination of the G-R color channel has a big difference in the cup to disk ratio (CDR) between healthy funds images of an eye and glaucoma fundus images of an eye.

(Qin et al., 2019) aimed a general procedure for the automatic segmentation of optic disc and cup. this technique is based on a fully connected layer (FCN). To localization the disc a pre-processing method is created. optic disc and cup segmentation most of the state-of-the-art techniques are using single region segmentation but the proposed method can segment both methods simultaneously. To accomplish this task Hough circle detection is used. This approach is able to identify the region of OD and OC very well the author has done a great job utilizing a single segmentation technique with he helps of image preprocessing and Hough principal.

In the research of (Issac et al., 2015) has proposed a method to segment optic disc (OD) and optic cup (OC). In this research adaptive thresholding technique is used to obtain the feature from images. standard deviation is used to remove unwanted information from fundus images red and green channels. OD is segmented from the green channel and OC is segmented from the red channel. The proposed approach has achieved segmentation with an accuracy of 92.06% in a very quick time. This approach can be beneficial for large images data set and a real-time application as well.

2.2 Convolutional Neural Network with Transfer Learning

In the study of (Aloudat et al., 2018) convolutional neural network (CNN) is used to distinguish between glaucoma and non-glaucoma images. A fully connected CNN is used to achieve classification. A total of 477 fundus images used in this research, the author has not used any additional feature extraction. Still, the score achieved by a fully connected network (FCN) is quite good. Accuracy, specificity, and sensitivity of 88.2%, 90.8%, and 85%, This results explain that CNN is a useful technique in the field of glaucoma classification, there is some limitation, this research dataset size is not large, transfer learning and augmentation is not used to increase the scores, whereas in another research (Norouzifard et al., 2019) has used transfer learning on 277 ONH photographs, a small size dataset causes overfitting issues, to overcome this problem transfer learning used, two pre-trained models InceptionResNet-V2 and VGG19. The result shows InceptionResNet-V2 accuracy is better than VGG19. The specificity and sensitivity of the InceptionResNet-V2 model recorded at 100 % and 90.1%. The use of this technique raises some questions? If the dataset size is not big enough then, why not use augmentation, regularizes, and dropouts.

Data Augmentation is an important technique in convolutional neural networks (CNN).In this research (Z. Wang et al., 2019) to diagnosis of glaucoma using active learning and adversarial data augmentation is used. Two major issues of glaucoma classification are discussed in this study. It is difficult and expensive to label a large amount of data.
Severe data imbalance makes the classifier easily over-fitting. To address the first issue an autoencoder is trained on labelled and unlabelled data to extract the features and with the help of active learning annotate data from the unlabelled dataset and to deal with overfitting issue deep convolutional generative adversarial network (DCGAN) is used. Overall this is a very good approach to deal with unlabelled data this can help get better accuracy in the convolutional neural network. Also, the score achieved 96% by this technique is better than other state of the art techniques, such as No augment and Augment by designed noise, furthermore an another research carried out using DCGAN (J. Wang et al., 2019) proposed a technique, that will help label images with the help of auxiliary domain images. To accomplish this task an adversarial transfer learning method is used. Under the condition on label information to match the distributions of source and target domains. It is a unique approach to deal with distribution in comparison with other available techniques where marginal distribution matching only. To carry three different datasets is chosen. Recall, F1, and G-mean scores of the ORIGA dataset achieved 0.8191, 0.6968, and 0.8007]. The author has done a great job of introducing a new way to label the training data by utilizing auxiliary domain images.

In the study of (Liao et al., 2020) proposed a network called EAMNet this will help in the diagnosis of glaucoma and also on highlighting the sharp regions. This is a novel scheme proposed, in this study where aggregating features from different scales to promote the performance of glaucoma diagnosis. This is also referred to as the M-LAP technique, it generates glaucoma activation, which helps fill the gap between location and global semantical diagnosis. AUC curve value achieved (0.88) in this approach beats the state-of-the-art solution. This technique is unique and useful this will also solve the shortage of interpretability of convolutional methods. The author has done, a great job, this approach will help weakly-supervised optic disc segmentation in clinical trials.

In this research of (Serener & Serte, 2019) a transfer learning approach is proposed to detect glaucoma using Resnet-50 and Google Lenet. In this approach Google Lenet model outperform Resnet-50. Here UC achieve in Google Net and Resnet is 0.91 and 0.84 respectively. This study has shown good results but a lot more image net model is available which, a fair comparison with those models would be required to understand and say that google Net is the best model for glaucoma classification and producing state of the art solution. Whereas in research of (Phasuk et al., 2019) an ensemble method is proposed to detect glaucoma in this technique multiple model is used with artificial neural network (ANN). This technique is shown results that beat the result of a deep neural network DENet using a public dataset. This approach can be useful in some cases where CNN is not producing better results, but CNN has to offer a lot of things and having GPU and TPU enabled cloud platform and with the help of transfer learning, CNN still can be the best option available in glaucoma classification.

In this research (M. Li et al., 2019) with other researchers conducted research and proposed a machine learning procedure, A large dataset is an essential thing to have in case of CNN with fewer data results cannot show great results with deep learning in application it is required to have a big dataset. In this study, researchers are attempting to solve this issue. At first, the model is trained on a nonmedical labeled dataset and after that fine-tuned and trained on a fundus image dataset. A self-trained strategy is applied to predict the labels for the unlabeled dataset, this approach can help in terms of low data and despite having low data results can be produced on high accuracy.

In this research, (Ferreira et al., 2020) has proposed a method for the segmentation of the anterior area by analysing optical coherence tomography images to classify angle-closure glaucoma, using transfer learning. The proposed method for angle-closure detection is distributed in four different steps at first acquisition and preparation of images use of AS-OCT to extract the feature, and at last, a multilevel network is created by the combination of CNN. this approach works well and produces a result of an AUC of 0.972. This approach can be useful even for small scale datasets.

2.3 Capsule Network and U-Net Model

In a research, carried (Gaddipati et al., 2019) is proposed to use a capsule network with Optical coherence tomographic (OCT) images. OCT images provide valuable information. In the proposed solution. The proposed model is trained on the resized input just to have a fair comparison. The suggested technique for the evaluation of glaucoma directly from the OCT volumes. Capsule network can learn highly discriminative features from OCT images. This network has achieved a 0.97 area under the curve value. this is a useful study that can help increase accuracy and extract features.

In this cloud-based U-Net model Fig 2 (Civit-Masot et al., 2019) has done a study, the author has proposed the use of the U-Net model see in Fig 2 based on cloud application google cloud with TPU setup. with fundus images. to segment optic disk (OD) and optic curve (OC). A U-net is a deep learning model which have shown better results in the medical filed with segmentation. With the recent growth in the tensor processing unit. the author has used a multiple dataset to segment OD and OC with a 97% accuracy this proves a TPU enabled machine can be very helpful when we deal with images that have lots of features and take a lot of to process.

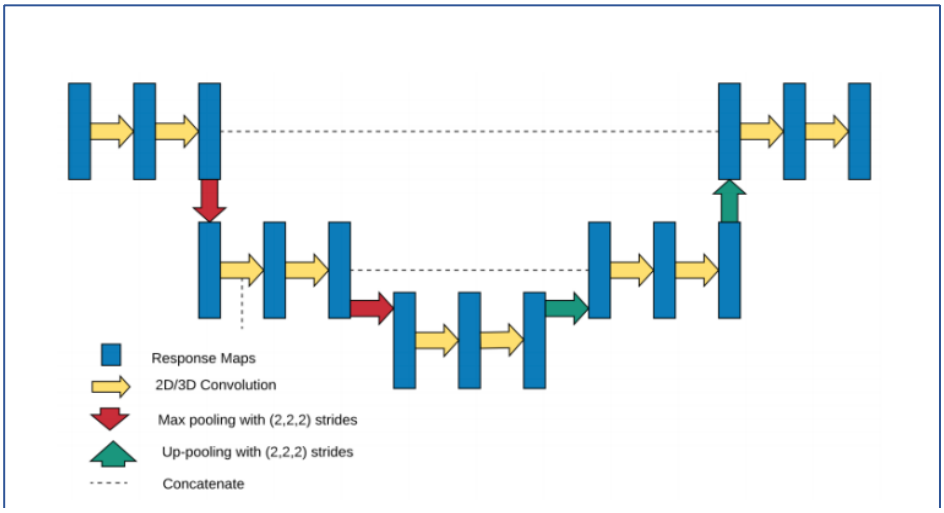


Figure 2: U-Net Model Architecture.

Analysing all the related works, techniques mentioned above. CNN can be a very useful technique to classify glaucoma. Through the right implementations, maximum accuracy can be achieved in minimal time, and with a smaller dataset. Data Augmentation, Data pre-process, Optics disc (OD), and optic cup (OC) segmentation and calculation of cup to disc ratio and use of transfer learning can produce the state-of-the-art results, below few of the implementation is used in a modified way to produce better results.

3 Research Methodology

3.1 Introduction

This research is based on Glaucoma stage classification, with the use of convolutional neural networks (CNN). After thoroughly scrutinizing the necessity of the research, the CRISP-DM strategy seems best befitted for the research. A few of the major reasons for picking CRISP-DM are flexibility, reliability, cost, and efficiency. This approach empowers us to modify approach as per the research requirement. A new approach based on CRISP-DM is prepared and explained below Fig 3. Multiple dataset was used for this research, in section 4.3 will discuss the design decision involvement and justify the architecture. One of dataset was used to verify segmentation accuracy by comparing expert result with model prediction. Other dataset is unlabelled data, this data is used for research directly.

3.2 Glaucoma Classification Methodology

Glaucoma classification Methodology Figure 3 was motivated by the knowledge discovery and data (KDD) strategy. It consists of the following stages. 1) Data Conversion: Convert (.tiff) images to JPG images. 2) Feature Selection: Feature selection using data Segmentation. 3) Data Cleaning, Optic cup (OD), and Optic disk (OD) calculation.4) Transform: Pre-processing: Data is preparation for the next step by calculating the cup to disk ratio. 5) Data Mining: Model architecture and their implementation and execution using Keras and TensorFlow. 6) Evaluation: All the results assessed and interpreted.

3.3 Research Understanding

The initial phase of the glaucoma classification methodology is the research objectives. Collected information is converted into a machine learning problem. Here, overall planning for the research conducted, and the goal of the research is well presented. The main objective of this research is to create a system that can classify the glaucoma stages using a fundus image, which can utilize minimum resources and provide maximum accuracy. The outcome of the system would be to classify the glaucoma stages, which will help take measure precautionary action on disease.

3.4 Data Understanding

Data gathering is the initial step of any research. Due to this data availability is become an essential factor in the research. In this research two image dataset is being used, DRISHTI dataset and MESSIDOR dataset. DRISHTI dataset has 100 images in the test and 100 images in train folders with additional details of the cup to disc ratio manually marked by 4 different experts. MESSIDOR dataset has an unlabelled 3220 images into the TIFF format.

3.5 Feature Selection

In research, it is vital to understand the data which will help answer the problems research is trying to solve. Here, fundus images are available in data sources and by analysing them glaucoma classification will be done. To classify glaucoma, two main features of fundus images need to be extracted. Cup value and disk value from fundus images and the ratio of these two will help classify images in different categories. for this segmentation, the technique was used which is explained in the next step.

3.6 Segmentation

An image is made as a group of several pixels. In segmentation grouping of pixels which has a similar attribute is done. In simple work extracting, the required details from an image can be classified as segmentation. In this research, segmentation was used to extract the value of cup Fig 4(b) and disk Fig 4(a), so that ratio of the cup to disk CDR was calculated and fundus images can be classified in different categories, to work with Convolutional neural networks in next step.

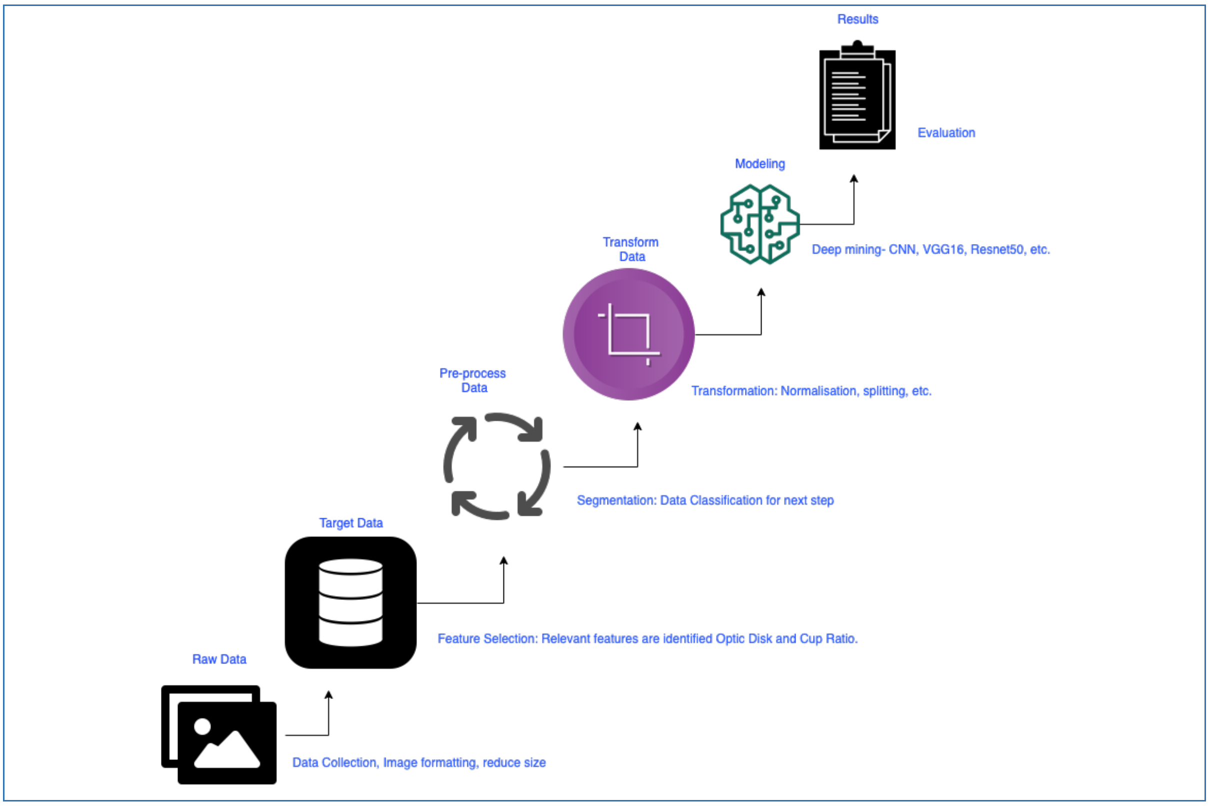


Figure 3: Glaucoma classification methodology

.

3.7 Data Transformation

As data were classified into multiple categories, it needs to be divided into training, test, and validation set. To Complete this task python is used in which a split of 60%, 20% 20% was done for training, test, and validation respectively. The initial format of data was in TIFF which is also converted into the Joint Photographic Experts Group (JPEG) format, before segmentation to reduce the weight of an image. The actual image dataset source has 23 gigabytes (GB) data after pre-processing without losing the resolution data size becomes 3.2 (GB).

Fig 4(a) **Fig 4(b)**

Fig 4: Optic disc 4(a) and Optic cup 4(b) after segmentation

3.8 Modelling

Pre-processed data can be used as an input into a model that is ready to use. In this research six classification modeling approach has been used, for further implementation. In this research, feature extraction is done using segmentation which helps classify the data in two Mild and sever Categories. further different modeling technique is used on given dataset to train models. In this research, six different imageNet models were used to find out how well a CNN model can classify the different categories of glaucoma. In this research, six proposed models are ResNet50, VGG16, VGG19, DenseNet169, InceptionV3, InceptionResNetV2. These models are simple to use and provide high accuracy because these are pre-trained models on millions of images and every modal is a state of the art solution provider.

3.8.1 ResNet50

(OVREIU et al., 2020) explains that residual networks are very effective to classify early-stage glaucoma. Early-stage detection produces great results with ResNet50 and achieve an of 96.95%. Through ResNet, it is possible to train network till 150+ network, earlier it was difficult to train deeper networks due to vanishing gradients. ResNet50 model has shown great success, this can help classify the glaucoma stage efficiently.

3.8.2 VGG16

(F. Li et al., 2018) use VGG16 to glaucoma and non-glaucoma visual field identification. VGG16 lift the accuracy score highest in compare with SVM and KNN 71% and 72 % respectively to 87%. This model outperforms Alexnet and achieve an accuracy of 92.7 % on ImageNet dataset. It will be interesting to see how well it perform when we camper with ResNet50.

3.8.3 VGG19

(De Moura Lima et al., 2018) used VGG19 in glaucoma diagnosis and achieved an accuracy of 77% on ORGIA dataset, implementing VGG19 on MESDIOR dataset can provide an insight how well this model perform with medical dataset and different changes in layers can produced better results than 77%.

3.8.4 DenseNet169

(Zhen1 et al., 2018) is carried a research to find the severity of glaucoma using DenseNet and achieve an accuracy of 75%. Using this model current dataset in this research, it will be Interesting to compare how well DenseNet perform and whether this research model can outperform and cross the accuracy of 75 %.

3.8.5 InceptionV3

(De Moura Lima et al., 2018) used InceptionV3 and get and accuracy of 79%. it has 48 deep layers, result produced by InceptionV3 can be compared with model which goes double deep while training, this model will be used to train and predict and compare with others which can few understanding regrading training deep layers from 48and more than 48 how much difference it can create in final accuracy.

3.8.6 InceptionResNetv2

(Norouzifard et al., 2019) has used InceptionResNetV2 and an outstanding accuracy of 100% of this showcase the strength of this model. This model can be very useful in this research and also comparing with different modal will give us insight how much better it preforms then the other models on same dataset.

3.9 Performance Evaluation

To evaluate the performance of the model in this research, the technique was used such as accuracy value, precision, and recall rate, and finally F1 Score with model execution time. This evaluation is done on six modals RssNet50, VGG16, VGG19, DenseNet169, Inceptionv3, InceptionResNetv2.

3.10 Conclusion

A customized methodology – Glaucoma Classification Methodology was designed for this research project. The three-tier architecture was used for this project, in which data flows from the client layer to the persistent layer to the business layer and returning to the client tier with results. Table 1 presents a comparison of the different models to date.

Table 2 Accuracy in previous research

Dataset	Classifier	Accuracy	Author
Custom Private	VGG16	87%	(F. Li et al., 2018)
ORGIA	VGG19	77%	(De Moura Lima et al., 2018)
proprietary dataset	ResNet50	96.5%	(OVREIU et al., 2020)
Princess Basma Hospital	InceptionResNetV2	100%	(Aloudat et al., 2018)
Custom Data Prepared (Zhongshan Ophthalmic Center)	DenseNet169	75%	(Zhen1 et al., 2018)
ORGIA	InceptionV3	79%	(De Moura Lima et al., 2018)

4 Design Specification

The Project structure design is illustrated in Fig 5 for Glaucoma classification is a three-tier framework, a client layer, a business logic layer, and the data persistent layer. To develop a method that can classify different stages of glaucoma their austerity an architecture is designed. Here, tools and techniques are described which is used to complete this research. In data persistent layer describe the source of data, and the business layer explains the pre-processing of the data, segmentation, training, and evaluation of the models, and result interpretation. The segmentation was performed by open CV and modeling was performed using TensorFlow and Keras API with the help of python in the google collab platform with a pro version scheme, which allows us to use GPU and a RAM of 25 GB size.

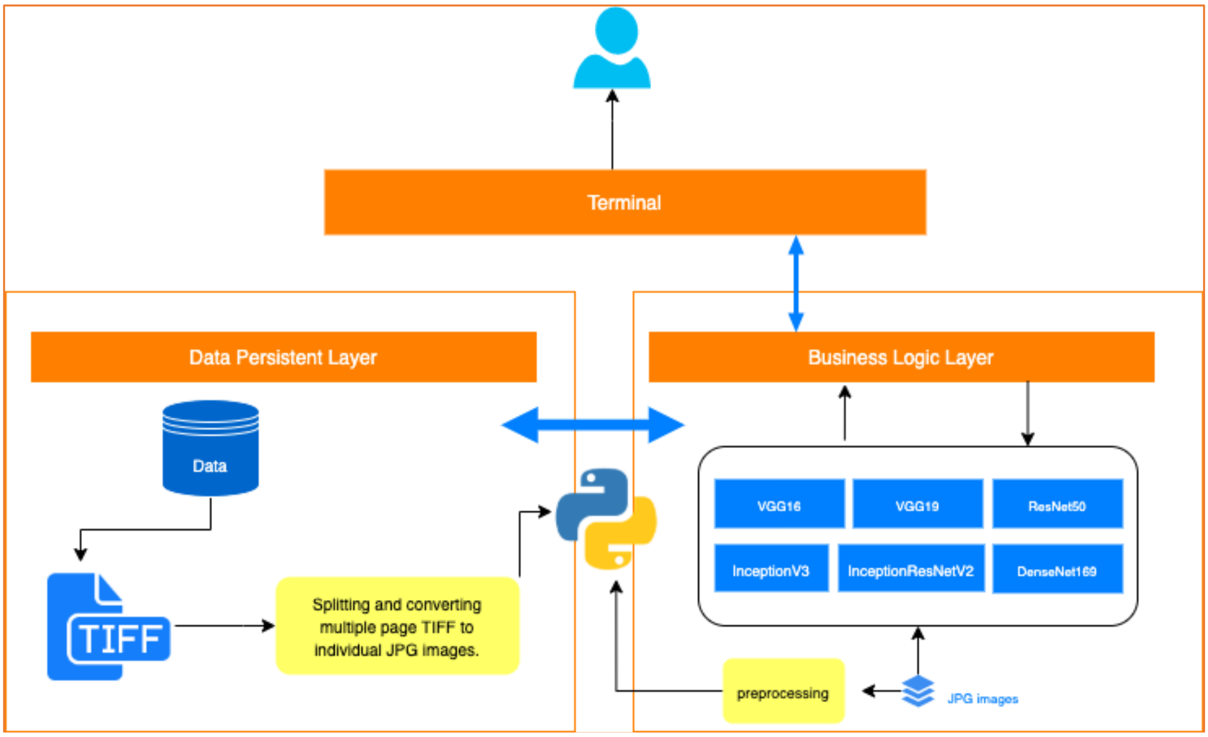


Fig 5: Three Tier Design Specification for Glaucoma Classification

5 Implementation

In this section, the overall implementation of the glaucoma classification addressed. Further, all the actions carried out during the research will explain. The first step of the process was environment setup, and details about the tools used in this research. Further data segmentation, data splitting, and image processing take place, further detailed analysis of model design and working is explained.

5.1 Environment Setup

In this research, a 16Gb RAM macOS is used. Two IDE is used for this research spyder for local use and googles colabPro version. Programming language Python 3.75 is used and for data storage unit AWS S3 bucket is used and google drive is used.

5.2 Data Pre-processing and Transformation

In this research two dataset was used named as DRISTI and MESSIDOR. DRISHTI dataset has 100 images with four experts predicted cup to disc (CDR) ratio available in Kaggle. In MESSIDOR dataset has 3220 TIFF images. MESSIDOR dataset images TIFF image converted using python in same dimension as it was in TIFF format.

5.2.1 Segmentation

In this step, A python model with the help of open CV has developed. In this approach cup to disk, value is calculated from the DRISHTI dataset. To validate the results of the model compared to Table 2 with the given CDR values by four different glaucoma experts. this model shows an accuracy of 85%. This gave them the confidence to use this model on the MESDIOR dataset to calculate the CDR. After calculating CDR of 3220 images. It was overserved that out of 3220 images only 1000 images fell into a mild and severe condition of glaucoma. CDR values more than 0.5 and above fell in Mild and severe categories. So, Images were split into mild and severe categories for further modeling.

5.2.2 Image Pre-processing & Splitting

In this step image pre-processing took place and circular images were extracted from actual image data and the unwanted black background was removed from the dataset, further divided into three different categories for machine learning modeling.

5.3 Classification Techniques

In this section, an explanation of multiple convolutional neural networks explanation will take place by which is used in this research to classify glaucoma. In this research, all the models that were used are ha given state of the art results and trained on millions of image dataset imagenet.

5.3.1 VGG16

VGG16 significantly exceed the former model in ILSVRC-2012 and ILSVRC-2013 competitions. Three fully connected (FC) layers accompany a stack of convolutional layers in the first two layers have 4096 channels all, it uses the final layer as SoftMax. This model has shown great results in the field of Image classification with normal and medical images, (Fei Li et al, 2018) has performed research and achieved an accuracy of 87%, Results are discussed on details on in evaluation section.

5.3.2 VGG19

VGG19 has 19 convolutional neural networks trained on image net dataset and it has 19 layers. VGG 19 is the extended version of VGG16 and in research of (Alan Carlos de Moura Lima at el, 2018) it can be seen how well VGG19 performed and produce an accuracy of 77% so this model was implanted and results was disused in details in evaluation section.

5.3.3 RestNet50

ResNet50 has 50 layers, and pre-trained on the image-net dataset and produce and accuracy which is state of the art. Residual architecture is very useful in biomedical images research carried by (Silvia OVREIU et al, 2020) performed well and achieve an accuracy of 96.5%, ResNet50 model is implemented over MESSIDOR dataset in this research and results are discussed in details in evaluation section.

5.3.4 InceptionV3

Inception V3 is a widely used image classification model it has achieved more than 78.1% accuracy on the ImageNet dataset. (Alan Carlos de Moura Lima at el, 2018) carried research and achieved an accuracy of 79%. The model was implemented on the MESSIDOR dataset and results are discussed in detail in the evaluation section.

5.3.5 InceptionResNetV2

Inception is a convolutional neural network and achieve a state-of-the-art accuracy in ImageNet dataset in ILSVRC image classification benchmark. A research carried out by (Aloudat et al., 2018) and achieved an accuracy of 100%. InceptionResnet model was used to classify glaucoma in this research and results are discussed in detail in evaluation section.

5.3.6 DenseNet169

DenseNet169 is a convolutional neural network, can be substantially deeper, more accurate, and gives the state-of-the-art result on the ImageNet dataset. Research carried out by (Yi Zhen at el, 2018) and achieved an accuracy of 75%. ResNet169 is implemented in this research and results are discussed in the evaluation section of the report.

6 Evaluation and Results

In this evaluation section, a thorough and complete review of the evaluations and results achieved by implementing different convolutional neural networks (CNN). The best performing models and parameters were chosen. In this research six models were selected which was used for classification of glaucoma VGG16, VGG19, InceptionV2, InceptionResNetV2, ResNet50, DenseNet169.

6.1 Evaluation

To evaluate the implemented models following metrics were used and briefly explained.

Accuracy: A model accuracy depends on the number of accurate predictions over the overall number of predictions.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

Precision: Precision can be calculated as total number of correct predictions over total number of correct predictions plus total number of false positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall can be calculated as total number of correct predictions over total number of correct predictions plus total number of false negative

$$Recall = \frac{TP}{TP + FN}$$

F1 Score: Mean of precision and recall is called as consider as F1 score.

6.2 Results

The results demonstrated below are collected in the experiment form so, the best classification model for glaucoma can be identified. These experiments are conducted based on criteria of Accuracy, Precision, Recall and F1 Score received during evaluation.

6.2.1 Experiment 1: VGG16

In this experiment, a fine-tuned model used. In this process, different optimizers were used with different learning rates. Fine-tuned model and Stochastic Gradient Descent (**SGD**) optimizer with a learning rate of 0.005, loss function categorical crossentropy with 50 epochs an accuracy of 63% were noted. This fine-tuned model was able to produce precision, recall, and f1 score 60%, 90%, and 72% respectively. Training with more data can increase the accuracy. Training accuracy against validation accuracy graph Fig: 6 (a) and training loss against validation loss graph Fig: 6(b).

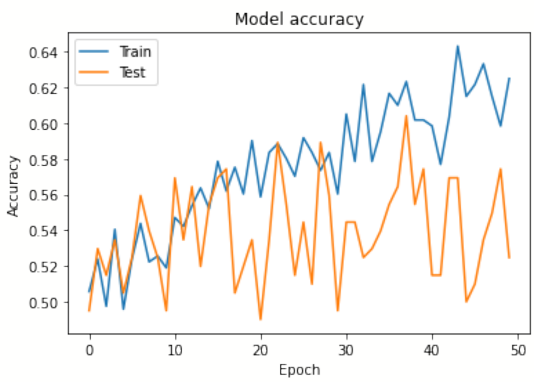


Figure 6 (a) Training Accuracy

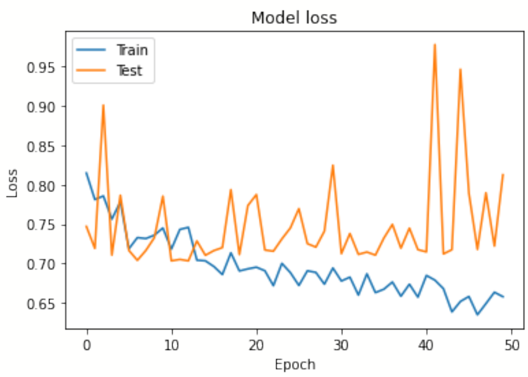
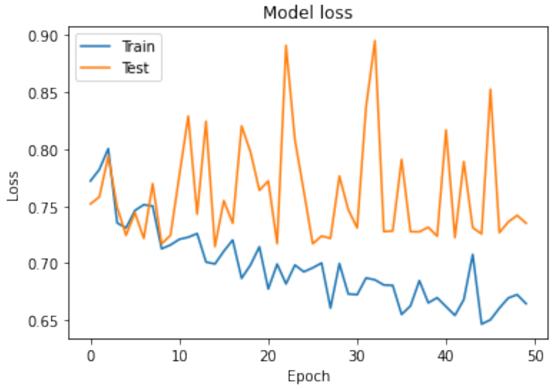


Figure 6(b) Validation Accuracy

6.2.2 Experiment 2: VGG19

In this experiment, multiple combinations of the optimizers were tried (i.e. Adam, SGD) with different learning rates. A fine-tuned model VGG19 with Stochastic Gradient Descent (**SGD**) optimizer with a learning rate of 0.001, categorical cross-entropy loss function used with 50 epochs an accuracy of 61% achieved. This fine-tuned model was able to produce precision, recall, and f1 score 75%, 71%, and 71%, respectively. Training with more data can increase the accuracy. Training accuracy against validation accuracy graph Fig: 7(a). Training loss against validation loss graph Fig: 7(b).



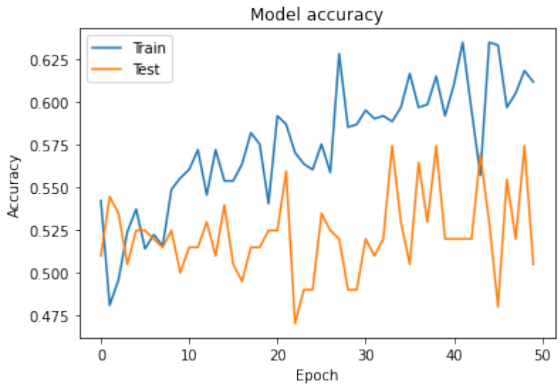
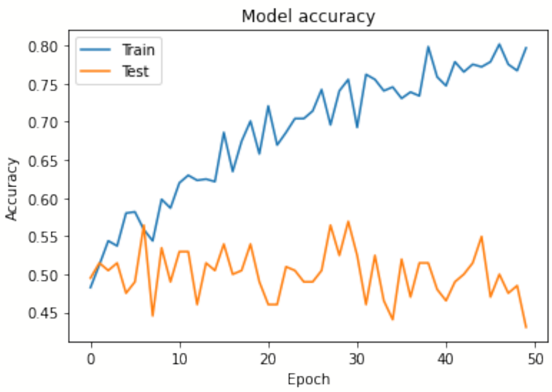


Figure 7 (a) Training Accuracy **Figure 7(b) Validation Accuracy**

6.2.3 Experiment 3: Inceptionv3

In this experiment multiple combinations of the optimizer (i.e. Adam, SGD) with different learning rate is tried. Using a fine-tuned model and SGD optimizer with a learning rate of 0.005, loss function categorical_crossentropy with 50 epochs an accuracy of 80% was noted. This fine-tuned model was able to produce precision, recall, and f1 score 75%, 71%, and 71% respectively. Training with more data can increase the accuracy. Training accuracy against validation accuracy graph Fig: 8(b).Training loss against validation loss graph Fig: 8(a).



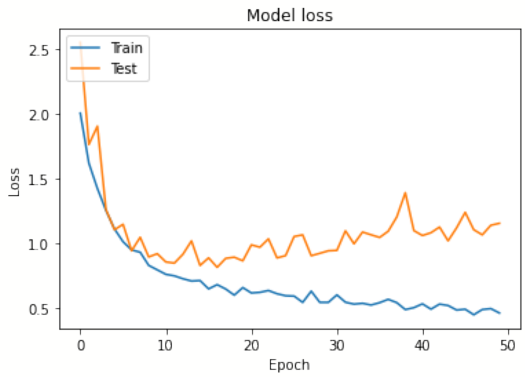


Figure 8 (a) Training Accuracy **Figure 8(b) Validation Accuracy**

6.2.4 Experiment 4: InceptionResnetv2

In this experiment multiple combinations of the optimizer (i.e. Adam, SGD) with different learning rate is tried. Using a fine-tuned model and SGD optimizer with a learning rate of 0.005, loss function categorical crossentropy with 50 epochs an accuracy of 61% was noted. This fine-tuned model was able to produce precision, recall, and f1 score 75%, 71%, and 71% respectively. Training with more data can increase the accuracy. Training accuracy against validation accuracy graph Fig: 9(a). Training loss against validation loss graph can be seen and Fig: 9(b).

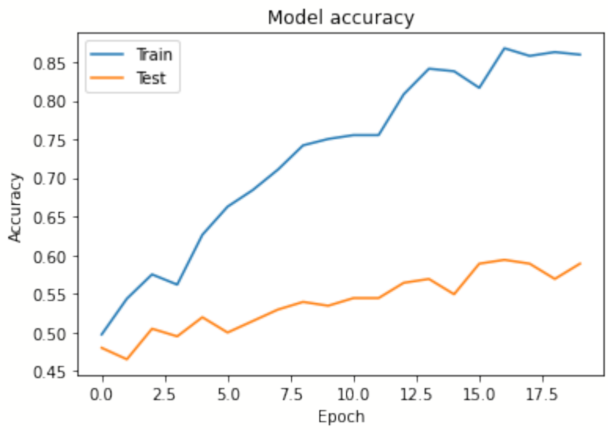


Figure 9 (a) Training Accuracy

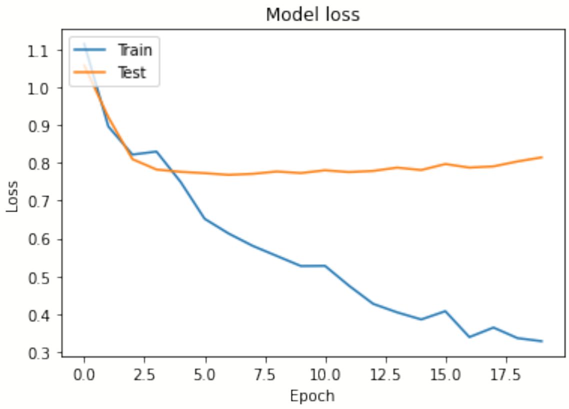


Figure 9(b) Validation Accuracy

6.2.5 Experiment 5: ResNet50

In this experiment, three custom layers were created and trained by stoping all the other layers and multiple combinations of the optimizer (i.e. Adam, SGD) with different learning rate is tried. A model with adam optimizer performed well, after that all the layers were trained, using binary cross-entropy and Adam optimizer with a learning rate of 0.005 on 10 epochs and the model was able to get an accuracy of 56% was noted. This fine-tuned model was able to produce precision, recall, and f1 score 64%, 70%, and 67% respectively. Training with more data can increase the accuracy. Training accuracy against validation accuracy graph Fig: 10(a). Training loss against the validation loss graph Fig: 10(b).

Figure 9 (a): Model Accuracy **Figure 9(b): Model Loss**

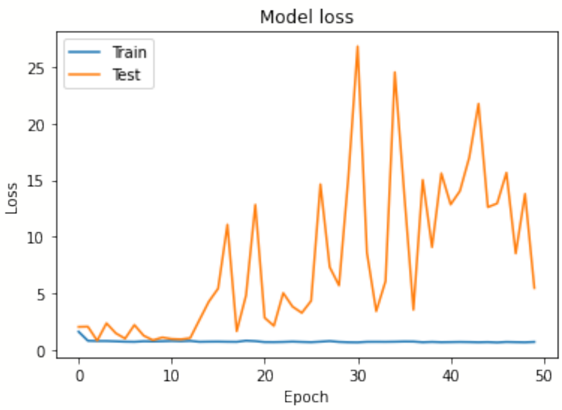


Figure 10 (a) Training Accuracy

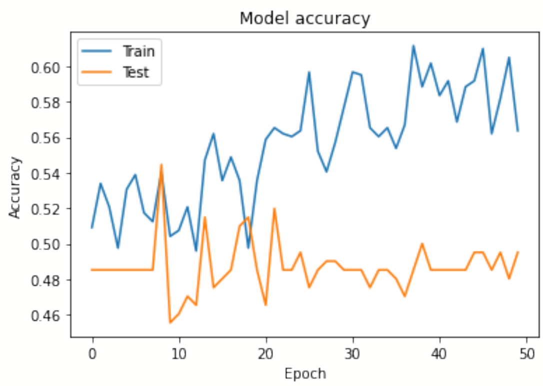


Figure 10(b) Validation Accuracy

6.2.6 Experiment 6: DenseNet169

In this experiment, multiple combinations of the optimizer tried but Adam performed well with different learning rates is tried. Three custom layers were created and trained by freezing all the other layers. Loss function categorical cross-entropy was used and the model has trained on 50 epochs. The trained model was able to achieve an accuracy of 85.95%. The model was able to produce precision, recall, and f1 score 75%, 71%, and 71% respectively. Training with more data can increase the accuracy. Training accuracy against the validation accuracy graph Fig: 11(a) and training loss against the validation loss graph Fig: 11(b).

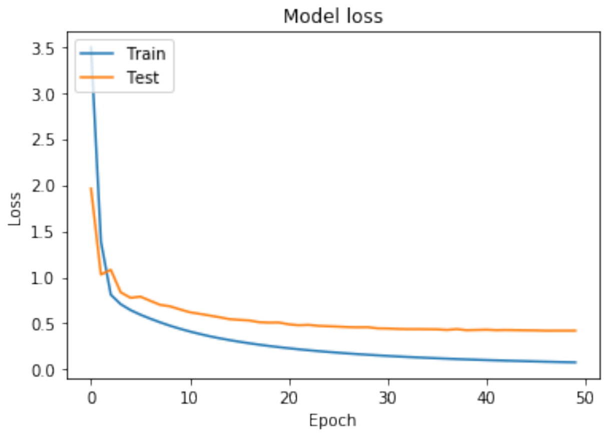


Figure 11 (a) Training Accuracy

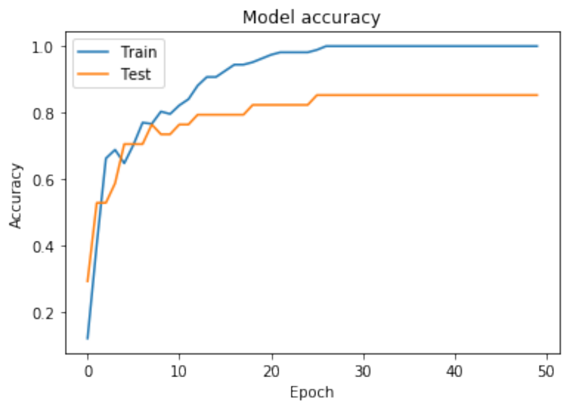


Figure 11(b) Validation Accuracy

7 Discussion and Comparison of Results

7.1.1 Experiment 7: Comparative Study of Classification Models in Terms of Accuracy

In this section comparison between the accuracy of the different models implemented in this research is done. In Fig 12 it can be observed Densenet169 was performing well and acquiring accuracy of 85.19%.

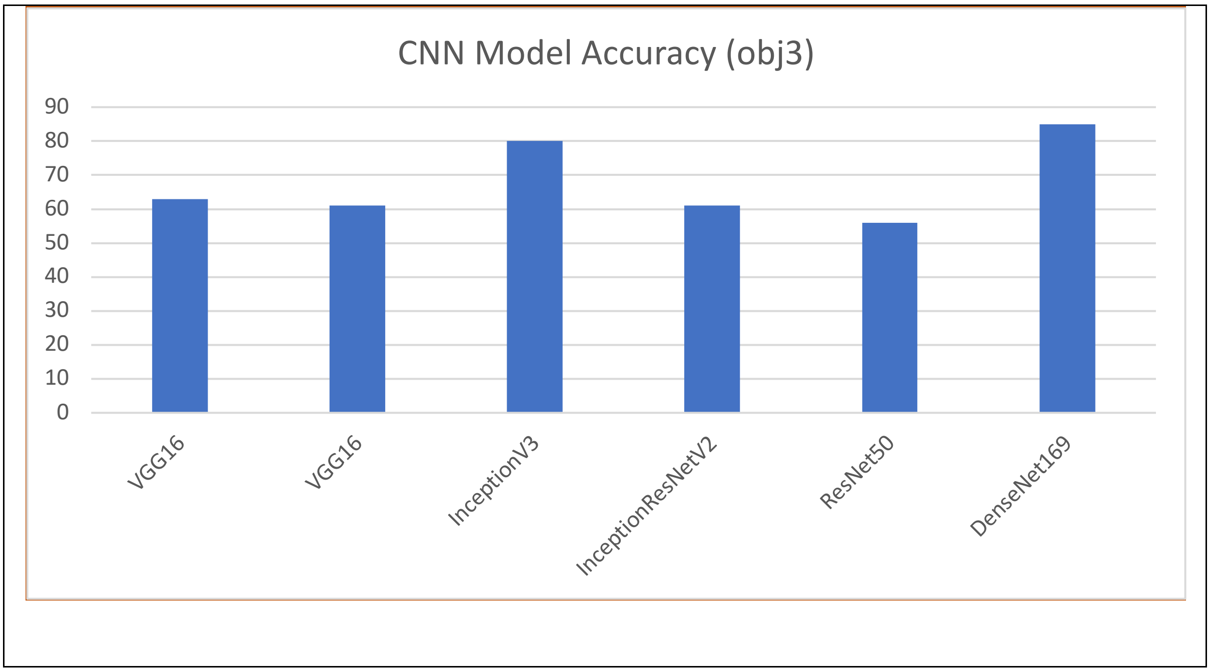


Figure 12: CNN Mode Accuracy comparison (obj3)

7.1.2 Experiment 8: Accuracy Comparison Between Current and Previous Researches

In this section, an accuracy comparison Table 3 between previous research and current research was done. objective 4 is completed of this research. Section 1subsection 1.3 is completed. DenseNet169 was able to achieve an accuracy of 85.19 which is better than previous research.

Table 3 Model Accuracy comparison (Objective 4).

Dataset	Classifier	Accuracy	Author
Custom Private	VGG16	87%	(F. Li et al., 2018)
ORGIA	VGG19	77%	(De Moura Lima et al., 2018)
proprietary dataset	ResNet50	96.5%	(OVREIU et al., 2020)
Princess Basma Hospital	InceptionResNetV2	100%	(Aloudat et al., 2018)
Custom Data Prepared (Zhongshan Ophthalmic Center)	DenseNet169	75%	(Zhen1 et al., 2018)
ORGIA	InceptionV3	79%	(De Moura Lima et al., 2018)
MESSIDOR	DenseNet169	85.95%	Current Research

In this research, deep learning models 6pretrained models VGG16, VGG19, InceptionV3, InceptionResNetV2, and ResNet50 and DenseNet169 were used. for feature extraction segmentation were used with the help of open cv and python programming language. These methods were trained in 661 images up to 50 epochs with various combinations of pre-trained models and custom layers, SGD and Adam optimizer used to optimize the performance, to handle loss, binary cross entropy, categorical cross entropy used. to manage overflow data augmentation is used. DenseNet169 performed well and achieve the highest accuracy. Validation accuracy is not up to the mark that further needs to be improved. Although the research goal was met it can further be improved by cropping images very close to CDR values, for segmentation U-Net modal can be beneficial. The limitation in this research was accomplishing more sets of fundus pictures that could help train model well and accuracy will increase further. The field of study can be expanded to other related fields such as, Diabetic retinopathy classification system.

Conclude the research question (chapter1, sub-section 1.3) has been solved and all the objectives (chapter 1, Table 1) have been implemented Table 3 and (chapter 1, subsection 1.2) objective is also achieved in Fig 12.

8 Conclusion and Future Work

To conclude all the findings of this research. Glaucoma classification into a mild and severe category is done, the system was able to classify fundus images into mild and severe categories with the highest accuracy. In the end, all the implemented models were compared to assess the resulting quality. DenseNet169 was able to obtain the highest accuracy of 85.19%. InceptionResNet was also able to achieve 80% accuracy, VGG16, and VGG19 achieved an accuracy of 63%, 61% respectively.

In further research on glaucoma classification, can be done by utilizing caps net, in a few of the research it has been observed that caps-net perform well especially with medical images. Due to limited time and resource availability, further hyperparameter tuning and testing were not performed in this research. In the future, more images in mild and severe categories should be added in the dataset for better results and validation scores. This study was based on reinforcement learning because the DRISRI dataset was used to build a model for segmentation which we used later to predict CDR value of the MESSIDOR dataset. In the future, the U-Net model can be used to do segmentation. A supervised technique can benefit from known features and already categorized images.

9 Acknowledgement

I would like to express my gratitude towards my supervisor and guide Dr. Catherine Mulwa, for overseeing and keeping me motivated throughout the research and also answering all my queries by discussing in-depth and showing a lot of good and bad examples. Her, feedback helped me write a better report and because of that, I was able to achieve a state-of-the-art solution project. I would also like to thank my family for giving me consistent support during my project long working hours and in the end. I want to my say thanks to friends who always come up with healthy discussions which helped me build better projects.

References

Aloudat, M., Faezipour, M., & El-Sayed, A. (2018). High Intraocular Pressure Detection from Frontal Eye Images: A Machine Learning Based Approach. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2018*, 5406–5409. https://doi.org/10.1109/EMBC.2018.8513645

Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D. W. K., Tan, N. M., Tao, D., Cheng, C. Y., Aung, T., & Wong, T. Y. (2013). Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging*, *32*(6), 1019–1032. https://doi.org/10.1109/TMI.2013.2247770

Civit-Masot, J., Luna-Perejon, F., Vicente-Diaz, S., Rodriguez Corral, J. M., & Civit, A. (2019). TPU cloud-based generalized U-Net for eye fundus image segmentation. *IEEE Access*, *7*, 142379–142387. https://doi.org/10.1109/ACCESS.2019.2944692

De Moura Lima, A. C., Maia, L. B., Pinheiro Pereira, R. M., Junior, G. B., Dallyson Sousa De Almeida, J., & De Paiva, A. C. (2018). Glaucoma Diagnosis over Eye Fundus Image through Deep Features. *International Conference on Systems, Signals, and Image Processing*, *2018-June*. https://doi.org/10.1109/IWSSIP.2018.8439477

Ferreira, M. M., Esteve, G. P., Junior, G. B., de Almeida, J. D. S., de Paiva, A. C., & Veras, R. (2020). *Multilevel CNN for Angle Closure Glaucoma Detection using AS-OCT Images*. 105–110. https://doi.org/10.1109/iwssip48289.2020.9145110

Fu, H., Cheng, J., Xu, Y., Wong, D. W. K., Liu, J., & Cao, X. (2018). Joint Optic Disc and Cup Segmentation Based on Multi-Label Deep Network and Polar Transformation. *IEEE Transactions on Medical Imaging*, *37*(7), 1597–1605. https://doi.org/10.1109/TMI.2018.2791488

Gaddipati, D. J., Desai, A., Sivaswamy, J., & Vermeer, K. A. (2019). Glaucoma Assessment from OCT images using Capsule Network. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 5581–5584. https://doi.org/10.1109/EMBC.2019.8857493

Issac, A., Parthasarthi, M., & Dutta, M. K. (2015). An adaptive threshold based algorithm for optic disc and cup segmentation in fundus images. *2nd International Conference on Signal Processing and Integrated Networks, SPIN 2015*, 143–147. https://doi.org/10.1109/SPIN.2015.7095384

Joshua, A. O., Nelwamondo, F. V., & Mabuza-Hocquet, G. (2019). Segmentation of Optic Cup and Disc for Diagnosis of Glaucoma on Retinal Fundus Images. *Proceedings - 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa, SAUPEC/RobMech/PRASA 2019*, 183–187. https://doi.org/10.1109/RoboMech.2019.8704727

Li, F., Wang, Z., Qu, G., Song, D., Yuan, Y., Xu, Y., Gao, K., Luo, G., Xiao, Z., Lam, D. S. C., Zhong, H., Qiao, Y., & Zhang, X. (2018). Automatic differentiation of Glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network. *BMC Medical Imaging*, *18*(1), 1–7. https://doi.org/10.1186/s12880-018-0273-5

Li, M., Mohamed, M. A., & Shousha, A. (2019). *SEMI-SUPERVISED TRANSFER LEARNING FOR CONVOLUTIONAL NEURAL NETWORKS FOR GLAUCOMA DETECTION Umm Al-Qura University , Department of Computer Science , Saudi Arabia Harbin Institute of Technology , Department of Computer Science , China University of Miami ,*. 3812–3816.

Liao, W., Zou, B., Zhao, R., Chen, Y., He, Z., & Zhou, M. (2020). Clinical Interpretable Deep Learning Model for Glaucoma Diagnosis. *IEEE Journal of Biomedical and Health Informatics*, *24*(5), 1405–1412. https://doi.org/10.1109/JBHI.2019.2949075

Manabe, T., Tomonaga, K., & Shibata, Y. (2019). CNN Architecture for Surgical Image Segmentation Systems with Recursive Network Structure to Mitigate Overfitting. *Proceedings - 2019 7th International Symposium on Computing and Networking, CANDAR 2019*. https://doi.org/10.1109/CANDAR.2019.00029

Norouzifard, M., Nemati, A., Gholamhosseini, H., Klette, R., Nouri-Mahdavi, K., & Yousefi, S. (2019). Automated Glaucoma Diagnosis Using Deep and Transfer Learning: Proposal of a System for Clinical Testing. *International Conference Image and Vision Computing New Zealand*, *2018-November*, 1–6. https://doi.org/10.1109/IVCNZ.2018.8634671

OVREIU, S., CRISTESCU, I., BALTA, F., SULTANA, A., & OVREIU, E. (2020). *Early Detection of Glaucoma Using Residual Networks*. 161–164. https://doi.org/10.1109/comm48946.2020.9141990

Patel, M. I. (2018). *Detection*. *Icoei*, 2018–2021.

Phasuk, S., Poopresert, P., Yaemsuk, A., Suvannachart, P., Itthipanichpong, R., Chansangpetch, S., Manassakorn, A., Tantisevi, V., Rojanapongpun, P., & Tantibundhit, C. (2019). Automated Glaucoma Screening from Retinal Fundus Image Using Deep Learning. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 904–907. https://doi.org/10.1109/EMBC.2019.8857136

Qin, P., Wang, L., & Lv, H. (2019). Optic disc and cup segmentation based on deep learning. *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, *Itnec*, 1835–1840. https://doi.org/10.1109/ITNEC.2019.8729455

Serener, A., & Serte, S. (2019). Transfer learning for early and advanced glaucoma detection with convolutional neural networks. *TIPTEKNO 2019 - Tip Teknolojileri Kongresi*.

https://doi.org/10.1109/TIPTEKNO.2019.8894965
Sun, X., Xu, Y., Zhao, W., You, T., & Liu, J. (2018). Optic Disc Segmentation from Retinal Fundus Images via Deep Object Detection Networks. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2018*(IEEE), 5954–5957. https://doi.org/10.1109/EMBC.2018.8513592
Thakoor, K. A., Li, X., Tsamis, E., Sajda, P., & Hood, D. C. (2019). Enhancing the Accuracy of Glaucoma Detection from OCT Probability Maps using Convolutional Neural Networks. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2036–2040. https://doi.org/10.1109/EMBC.2019.8856899
Wang, J., Yan, Y., Xu, Y., Zhao, W., Min, H., Tan, M., & Liu, J. (2019). Conditional Adversarial Transfer for Glaucoma Diagnosis. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2032–2035. https://doi.org/10.1109/EMBC.2019.8857308
Wang, Z., Wang, Z., Qu, G., Li, F., Yuan, Y., Lam, D. S. C., Zhang, X., Zhang, Y., & Qiao, Y. (2019). INTELLIGENT GLAUCOMA DIAGNOSIS VIA ACTIVE LEARNING AND ADVERSARIAL DATA AUGMENTATION Graduate School at Shenzhen , Tsinghua University , China 2 Sensetime Group Limited , Beijing , China Shenzhen Institutes of Advanced Technology , Chinese Academy of Scie. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, *Isbi*, 1234–1237.
Zhen1, Y., Wang2, L., Liu2, H., & Zhang3, J. (2018). *Performance assessment of the deep learning technologies in grading glaucoma severity*. *arxiv*.