Optic Disc and Cup Segmentation in Glaucoma Screening using Mask RCNN

MSc Research Project
Data Analytics

Nandita Sharma
Student ID: x18185100

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Nandita Sharma
Student ID:	x18185100
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	17/08/2020
Project Title:	Optic Disc and Cup Segmentation in Glaucoma Screening using Mask RCNN
Word Count:	6171
Page Count:	19

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:	Nandita sharma
Date:	28th September 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):	

Optic Disc and Cup Segmentation in Glaucoma Screening using Mask RCNN

Nandita Sharma
x18185100

Abstract

Glaucoma is a prevalent ocular disease, which leads to irreversible loss in vision because the optic nerves, that are connected directly to the brain gets damaged. Compared to the healthy fundus image, enlargement of an optic cup could be observed by covering a portion of the optic disc in the fundus image of glaucoma. Ophthalmologists believe that it can be treated to some extent if early detection is possible. Several studies have been done so far in this field. However, the detection and segmentation of the optic cup and disc is a challenging task. Therefore, in this paper, a different deep learning approach is adopted to detect and segment the prominent location of optic disc and cup from the fundus image using Mask RCNN. The promising result by Mask-RCNN could be seen in other state-of-art in detecting and segmenting salient objects from images. This work is formulated on high resolution public available RIGA dataset of fundus images comprises of ophthalmologists labeled data. Various performance metrics such as F1 score, Precision, Recall, Accuracy has been analysed in this study.

1 Introduction

Glaucoma is a chronic eye disease. It is a major cause of blindness. It usually comes at early stages with no obvious symptoms. When there is a blockage in the eye canal from which aqueous humor flows, glaucoma sets in. This blockage causes the interocular pressure to gradually increase and thus the size of the optic cup to rise. The increased size of the optic cup causes a gradual loss of optic nerve fiber and is observed in its victim as a gradual loss of vision. Patients do not have symptoms at an early stage, but as the disease progresses, the visual field usually narrows from the peripheral point of view. This is a destructive act of eye ailment that is irreversible and later makes a person completely blind. Hence, there is a requirement of technology that could make early detection possible and can save a person's vision from getting lost. Optic cup and disc segmentation are usually performed manually by trained professionals. The workout is a repetitive one affected by exhaustion and mental instability. They calculate cup to disc ratio (CDR), ration less than 0.5 is considered as non- glaucomatous, and greater than 0.5 is considered as glaucomatous. Several approaches have been applied so far to measure the CDR. Technology in machine learning and deep learning has also played a major contribution in the medical field and given a modern turn to the physician in their work by reducing the time consuming manual work with minimal cost.

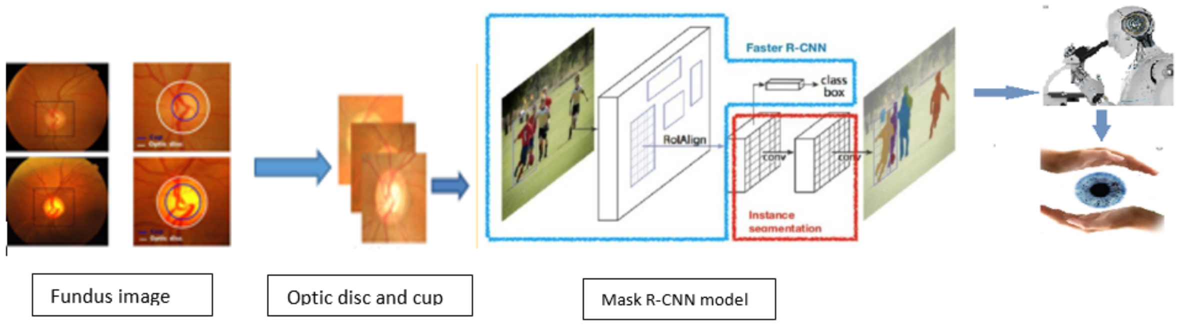

Figure 1: Approach to Glaucoma with Mask-RCNN

There is various approach presented by a researcher in the past to segment optic disc and cup from eye fundus image to identify glaucoma. Method of thresholding has been applied to gauge optics disc candidate areas. The blood vessels and features derived from candidate areas that were used to estimate the optic disc. Some studies adapted the ellipse fitting method to evaluate the optic disc area. Recently, different deep learning tailored to image detection and image segmentation from the provided image. The promising results were observed by previous research using FCN, FCN2, U-Net, etc. Nevertheless, accurate detection and segmentation is still a challenging task.

Therefore, this paper is implementing a new approach to automatic detection and segmentation using Mask R-CNN. It is an advanced version of faster RCNN. The proposed technique used ResNet 101 architecture and generates Region of interests (ROI) in pixel to pixel manner from Region proposal Network (RPN). The segmented mask, image detection with the label could be possible with this method.

1.1 Research Question

The proposed research addresses the problem faced in achieving precise segmentation of optic cup and disc and would help pathologists for better decision making.

Research Question: How well Mask R-CNN model can segment the optic disc and optic cup in screening the glaucoma?

To answer the research question. The rest of this research work is structured according to the following. Section 2 explains the related work. Section 3 explains the method and proposed technique in detail. In Section 4 we discuss the design specification, Section 5 discuss the implementation, experimental techniques, and Section 6 ends with a conclusion and future work.

2 Related Work

2.1 Machine Learning

The brain tumor and edema segmented automatically in Schmidt et al. (2005), by combining Ab features with textural which has given a substantial increase in performance

level.
Without prior knowledge, the non-linear distribution of image could be learned by one-class support vector machine (SVM), through the automated process of training SVM parameters and implicit learning. Then segmentation task is performed after the learning process in (Zhou et al.; 2005)

The Markerless gating and tracking are algorithms are based on machine learning formulated in research Lin et al. (2008). In this study, PCA and ANN were used to resolve the problem of two classifications of gating technic. Furthermore, the problem of tracking was a regression method, which uses the correlation between the location of the tumor and nearby features like surrogate anatomic presented in an image. The experiment tested on 4 regression methods such as ANN, 1-degree, and 2-degree linear regression and SVM. The applied methods demonstrated the superb performance for 10 fluoroscopic images that were sequences of 9 patients.

Murthi and Madheswaran (2012) followed the conversion of the grayscale and equalization of the histogram in their research in the selected area of interest. A threshold value of 0.3 was used to differentiate between the two groups after localization of the optic cup and optic disc. The 20 fundus image was tested for the analysis of their research and concluded that the technique of CDR in detecting glaucoma is successful.

Brain tumor segmentation is a difficult task. Gupta et al. (2015) proposed a method to segment tumor using T2 weighted images from MRI images. The segmentation approach followed in this work takes into account the symmetry properties of the human brain to classify different regions with a likelihood of tumor existence accompanied by area-widening technique to classify the same tumor region. The result was evaluated in both the term quantitatively and qualitative concerning ground truth which was given by radiologists. The study showed higher Dice Similarity Coefficient (DSC) scores compared to another state of the art for tumor segmentation.

Unlike unsupervised learning, supervised learning approaches require classification and then segmentation of all of the new instances. However, Soltaninejad et al. (2016) taken an approach of combining the random forest classifier with each superpixel detection and segmentation for the FLAIR MRI (Magnetic resonance imaging) images.

Using machine learning, the Diagnostic tool for Diabetic retinopathy is designed and implemented by Santhakumar R et al. (2016). The screening and diagnosis are possible through these tools. The Image and patch level prediction are two segments of screening tools. The algorithm such as support vector machine (SVM) is used in patch classification for the prediction of Hemorrhage and Hard Exudates in the fundus image. The result obtained using the SVM model is 96% and 94% in accuracy and sensitivity respectively for Hardexudates. However, accuracy and sensitivity are 85% and 77% respectively for haemorrhage.

Also, Thangaraj and Natarajan (2017) have used classification techniques using a supervised learning method in their researcher for glaucoma detection with an accuracy of 93%. The presence or absence of glaucoma is classified by calculating the CDR ratio in image and PCA used as feature extraction. The DRIONS DB, RIM ONE, and STARK PROJECT database brought to test normality and abnormality in the image.

The new method proposed by Bonte et al. (2018) , which can delimit the tumor tissues from a minimum amount of data. Hence, random forest classification is combined with abnormal characteristics and voxel-wise texture on a FLAIR MRI and T1.

Abbas et al. (2019) proposed the LIPC methodology for brain tumor segmentation and classification using machine learning as segmentation is a very important task to detect tumors at the initial stage. As noisy data can create a problem of inaccuracy, hence image processing techniques were used to remove noise. Image enhancement is applied to enhance the dark images for clear visibility. The PCA was used for feature reduction with the LIPC model. The tumor is segmented using LIPC approach using the assumption that it could be reconstructed from different patches of different tumor classes by predicting the tumor onto certain patches.

Table 1 Summary of Related work in Machine Learning

Author(s) and Title	Study Aims & Objectives	Research Design	Sample	Source	Data collection methods	Findings relevant to the review.
(Schmidt et al., 2005) and "Segmenting brain tumors using alignment-based features"	To Achieve efficient segmentation	Combined AB features with textural feature	25 normal image, 10 with tumor with 4 types of tumor.	IEEE	DNN database	New approach by combining Ab feature given increase in performance level
(Lin et al., 2008). And "Tumor Targeting for Lung Cancer Radiotherapy Using Machine Learning Techniques"	Targeting tumor in lungs using machine learning	New method of Gating and tracking algorithm is used. Along with PCA and ANN	600 frames	Seventh International Conference	Private data	Excellent result is achieved on 10 fluoroscopic.
(Murthi and Madheswaran, 2012) and "Enhancement of Optic Cup to Disc Ratio Detection In Glaucoma Diagnosis"	Aim is to estimated more accurately neuro-retinal optic cup and CDR detection.	Method used here ARGALI (Automatic cup-to-disc Ratio measurement system for Glaucoma detection and Analysis)		(ICCCI) conference		This technique achieved better CDR ration.
(Gupta et al., 2016) and "Brain Tumor Segmentation by Integrating Symmetric Property with Region Growing Approach"	Aim to segment tumor by integrating symmetric property	Various tumor region is identified using symmetry property.		IEEE INDICON	MICCAI 2012	Achieved good Dice score.
(Soltaninejad et al., 2017) and "Automated brain tumour detection and segmentation using super pixel-based extremely randomized trees in FLAIR MRi"	Aim to segment and detect abnormal tissues present in brain which can cause tumor	Designed architecture with random forest classifer + each super pixel	30 image 20 high and 10 low grade	International journal of computer assisted Radiology and surgery	19 MRI FLAIR images	Average detection sensitivity =89.48%, balanced error rate=6% and the Dice overlap=0.91for the segmented tumour against the ground truth.
(Soltaninejad et al., 2017) and "Machine learning based brain tumour segmentation on limited data using local texture and abnormality."	To segment tumor in brain with the help of texture and abnormality	combined random forest with abnormal characteristics and voxel-wise texture		Conference research	BraTS 2017	Result shown different Dice score for low and high grade glioma
(Santhakumar et al., 2017) and "Machine learning algorithm for retinal image analysis"	Aim is to perform analysis on retinal image of Diabetic retinopathy.	SVM is used to diagnostic tool for diabetic retinopathy	Exudate Patches=319(train),96(test) Hemorrhage Patches= 448(train),225(test)	IEEE		The accuracy and sensitivity for Hardexudates is better than haemorrhage with 96% and 94% respect.
(Thangaraj and Natarajan, 2017) and "Glaucoma diagnosis using support vector machine"	Aim is to detect glaucoma.	The support vector algorithm and PCA for feature selection	40 fundus image, 30 normal,10 affected	ICICCS conference	DRIONS DB, RIM ONE, and STARK PROJECT	SVM performed better in accuracy and computational efficiency
(Abbas et al., 2019) and "Automatic Brain Tumor Detection in Medical Imaging using Machine Learning"	Aim is to segment and classify Brain tumour	LIPC model with PCA is used.	30 patient image	ICTC conference	MICCAI 2013	This result have given satisfactory accuracy with Dice score of 0.95.

2.2 Deep Learning

The modified version of the U-net convolution neural network (CNN) was used in this research paper. The applied architecture has less number of the parameter as compared to the general U-net. The experiment was performed on two datasets DRISHTI-GS and the RIM-ONE v.3. The model is trained for 250 epochs, used 159 and 50 fundus images from both dataset RIM-ONE v.3 and DRISHTI-GS respectively. The fundus image in RGB format was cropped from optic disc location, then spline interpolation, image resizing of 128x128, and histogram equalization were performed before feeding it into a neural network in (Joshua et al.; 2019a) proposed study. The IOU score and dice score for RIM-ONE v.3 dataset performed better than DRISHTI-GS database.

The new approach to automatically segment optic disc and cup in characterizing the presence of glaucoma from Region of interest (ROI). Two steps was followed in this study to estimate ratio: one is ROI from a fundus image then followed by the segmentation process. The researcher applied fully convolutional networks (FCN) along with U-Net architectures for their study to segment optic cup and disc. The binary and multi-class fully convolution networks are implemented by the researcher along with two different Regions of interest as an input, one is the original image (Original ROI) and another is the masked image (masked ROI) to get a result in the segmentation process. Binary class FCNs show better performance compared to multi-class FCNs. Also, the full convolution network for optic cup and disc segmentation performed better in ACC, f-measure, and Jaccard index than existing other algorithms in performance. The applied FCNs technique showed a promising result (Kim et al.; 2019).

Interpretable Glaucoma detector (InterGD) is a multitask model that was implemented by Mojab et al. (2019). This InterGB consists of two-component. One component is segmentation modules which used to segment the optic disc and cup from fundus image and another component is prediction modules employed in this study to predict the likelihood of glaucoma in the image. Both components are integrated with a multi-task framework to perform end to end training. The study is performed on three datasets, among which two are public datasets, DRISHTI-GS and RIGA and one is private dataset I-ODA. Mojab et al. (2019) used five models as a baseline. As a part of data preprocessing, resizing and channel-wise normalization techniques were introduced to all images. Data were divided into training and testing with 80% and 20% respectively. The following evaluation technique was used to evaluate both the model such as recall, f1-score, and precision for evaluating prediction model, Disc similarity coefficient (DSC) for the segmentation model. Through this study, the capabilities of InterGD tries to resolve the existing challenges by gaining interpretability at a minimal cost.

Qian et al. (2019) The researcher has used Mask RCNN model for object detection and observed this algorithm outperformed in detecting objects as compared to other start-of-art models with an accuracy of 96.8 percent that to achieved in 116ms. The study is performed on SAR images to locate ships in multitemporal. Both multiple and single masks are used in the experiment. The denoising and data augmentation was used as pre-processing. As a result, it was noticed that multi-mask proven to be more effective than a single- mask in detecting SAR ship. The study showed Mask RCNN can detect an object sufficiently within less time consumption for SAR image in ship detection.

The detailed comparison between FCNs and Mask RCNNs was implemented in the context of salient object detection (SOD) by Krinski et al. (2019). The architecture of both the FCNs and mask R-CNN is explored in this study to detect the salient area from the image. The study aimed to demonstrate effectively Mask RCNNs dominance over FCNs in the SOD images and proved Mask RCNN achieved better performance level over FCNs on F-measures gain of up to 47%. The experiment was performed over 8 publically available datasets and four metrics were used for evaluation such as precision, f-measure, recall, and MAE (Mean Absolute Error). The Mask RCNN has an RPN which is attached before the segmentation module. This decreases the false-negative error that results in improving segmentation. The researcher endorses the Mask RCNN is a promising solution for SOD problem-solving.

There are different segmentation approaches and algorithms. Each has its advantage and limitation. Madhupriya et al. (2019) has proposed the U-net model, which is based on a deep convolution network. BRATS 2017 datasets were used, which contain real images of High-Grade Glioma (HGG) and Low Grade Glioma (LGG) patients. The model showed an efficient and robust result by achieving maximum DSC(Dice Similarity Coefficient) metric of 0.82 and 0.81 for the BRATS dataset.

Another, study empirical on medical is nuclei segmentation. The U-net and Mask-RCNN are two popular segmentation framework was introduced by previous research. An ensemble model, by combining both models is proposed in (Vuola et al.; 2019). This presented the improvement in results by merging the predictive power of U-net and mask RCNN.

Kassani et al. (2019) approached a novel CNN model with Xception architecture in it for classification of DR severity. The Xception is aggregated with a layer of deep CNN. In pre-processing min-pooling is employed on input image for color contrast. This study performed no data augmentation for training model. The overfitting issues are resolved using L1 and L2 regularization. This Xception deep extractor showed better performance in comparison with original Xception architecture on APTOS dataset. This has given promising results on the limited number of training data through this approach of research with accuracy, sensitivity, a specificity of 83.09%, 88.24%, 87% respectively.

Glaucoma detection using AG-CNN (Attention-based CNN model) was implemented in Li et al. (2020). The LAG Large-scaled labelled dataset was used for the research. Their concept was to bridge the gap between pathological area and human attention, hence the study proposed to improve the projected focus maps by adding a feature analysis framework for the identification of glaucoma. It was learned in a weakly supervised manner. The back-propagation method is employed to identify the pathological area which was based on the predicted attention map. As a result, the focus maps could be optimized and then used to illustrate the area which was most important in detecting glaucoma. The proposed attention-based mechanism could find salient parts of the feature in an image. The study achieved an accuracy of 96.7%.

The quality of the fundus image is assessed in this paper Raj et al. (2020). The multivariate regression based CNN model is used on the FIQuA dataset. A total of 1500 images are prepared along with seven subjective inputs that are given by ophthalmologists. Among seven, six are quality parameters and left one is class of quality as good, fair, and poor. The model peculiarity is that using the subjective feedback given by ophthalmologists, it extracts the optimized features for classification. It has two blocks, Block1 manage the optimized functionality from 4 pre-trained CNN models. There models as ResNet-151, Xception Inception-V3, and DenseNet-121 trained against the six subjective scores given by the ophthalmologists through transfer learning. Alternatively, these streamlined features are integrated and forwarded Block-2 classifies the background images into three classes (good, fair, bad). The block2 technique achieved 95.66% accuracy.

Recently, the application of deep learning Mask RCNN expounds in the dental area of medical. Zhu et al. (2020) performed tooth detection and segmentation in their study that would automatically detect and segment the tooth in the mouth. Since no publically dataset was available, therefore 100 images from the private hospital were considered for the experiment, among training is carried out on 80 images, 10 for validation, and the remaining 10 used for testing. The model is trained for 50 epochs and observed a total loss of 0.3093. In terms of accuracy, Pixel accuracy showed a result of 97.4%. It can be stated as in future applied technique is merged with other more networks to predict the presence of diseases like orthodontics.

Table 2: Summary of Related work in Deep Learning

Author(s) and Title	Study Aims & Objectives	Research Design	Sample	Source	Data collection methods	Findings relevant to the review.
(Kassani et al., 2019 and "Diabetic Retinopathy Classification Using a Modified Xception Architecture (Deep learning)"	Objective is to classify Diabetic retinopathy.	novel CNN model with Xception architectures used in this study, L1 and L2 resolve the issue of overfitting	2657 images	IEEE paper	APTOS Dataset	This Xception performed well and better result for APTOS dataset
(Kim et al., 2019) and "Optic Disc and Cup Segmentation for Glaucoma Characterization Using Deep Learning"	Aim is to perform automatic segmentation of optic and cup.	Technique FCNs + U-net is taken into consideration for experiment	699 image used analysis	IEEE international paper	Experiment is carried out on publicly available :RIGA dataset	Result showed promising result.
(Joshua, Nelwamondo and Mabuza-Hocquet, 2019) and "Segmentation of Optic Cup and Disc for Diagnosis of Glaucoma on Retinal Fundus Images"	Goal is to perform segmentation using U-netCNN	Improved U-net CNN is used in this study	50 and 159 fundus image	PRASA Conference	DRISHTI-GS and the RIM-ONE v.3	IOU and dice score of RIM-ONE v3 performed better than DRISHTI
Afolabi O. Joshua, Fulufhelo V. Nelwamondo(Segmentation of Optic Cup and Disc for Diagnosis of Glaucoma on Retinal Fundus Images)	Aim is optic and disc segmentation to detect glaucoma.	The improved architecture of U-net is used. RGB fundus image is cropped from OD location, applied histogram equalization and feed into Neural network.		PRASA Conference paper	A famous publicly available dataset is used. RIM-ONE v3	The result showed improved performance level in segmentation of optic cup and disc. This paper suggest there is need to improve IOU score in future.
(Cherguif et al., 2019) and "Brain Tumor Segmentation Based on Deep Learning"	Aim is to segment tumor from brain image.	Proposed different approach using U-net		IEEE paper	BRATS 2017 datasets	Achieved good score in dice score of 0.81
(Raj et al., 2020). And "Multivariate Regression-Based Convolutional Neural Network Model for Fundus Image Quality Assessment. "	Aim is to evaluate the quality of fundus data.	CNN based on multivariate regression based. Model have resnet151, Xception, DenseNet-121	1500 macula	IEEE paper	FIQuA dataset	Result achieved accuracy of 95.66% for block 2 classifer.
(Li et al., 2020). And "A Large-Scale Database and a CNN Model for Attention-Based Glaucoma Detection"	To determine automatic glaucoma detection and pathological area localization upon fundus images	AG-CNN (Attention-based CNN model) is designed by researcher.	11 760 fundus images	IEEE TRANSACTIONS ON MEDICAL IMAGING	LAG Large dataset	Observed that predicted attention map enhanced the detection of glaucoma.

3 Methodology

The proposed methodology automatic segment and detect the optic cup and optic disc from fundus images. This research implemented an advanced technique based on previous research Qin et al. (2019); Joshua et al. (2019a). The proposed research process segments the optic disc and cup simultaneously from the input image. Both cups and discs at the same could be detected in this research method. The Mask RCNN deep neural network architecture is being utilized in this research. The CRISP-DM methodology is used in this study.

It is a cross-industry process involves data and business understanding, pre-processing part, Data modelling, and evaluation techniques (Martínez-Plumed et al.; 2019). The flowdiagram of CRISP-DM is shown in figure 2.

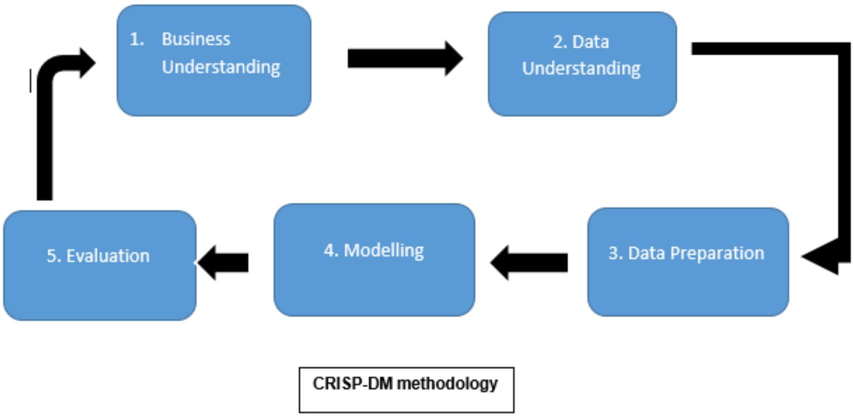

Figure 2: CRISP-DM methodology

3.1 Business Understanding

Glaucoma can be prevented from getting complete loss of vision if the early stage of it is detected. It requires a precise method to segment cup and disc. Doctor and pathologist evaluate fundus images of person after performing pre-processing of images. The pathologist finds estimating fundus image under the microscope is time-consuming and it is highly affected by doctor experience in this field. Therefore, as discussed Mask RCNN model is used to reduce the time and manual human efforts required for evaluating the fundus image of Glaucoma. The segmentation is a tough and challenging task to get precise segmentation of optic cup and disc. Thus, the proposed method is applied to overcome this problem.

3.2 Data Understanding and Collection

This research is performed on RIGA-dataset, which is extracted from open source "Deep Blue Data" repository. The link for RIGA dataset is `https://deepblue.lib.umich.edu/data/concern/data_sets/3b591905z`. The dataset consists of fundus images of the eye, provided by the University of Michigan. This data has three sets such as Magrabi, Bin Rushed, and MESSIDOR. The data is tested by high professional ophthalmologists. All fundus image consists of the optic cup and optic disc is marked and annotated by six experienced ophthalmologists. The data consists of high-quality TIFF and JPG format total 4500 fundus image. The size of a fundus image is 2160x1440.

3.3 Data Preparation

This is a prominent step to perform on images before training the model. In the dataset, the images are pathologically labelled. In the data cleaning, unlabelled fundus images are removed and 1400 labelled images are kept. As different resolution of images may cause a model to train inaccurately, thus high-resolution images are sized to 512 x 512 shape to bring all images in the dataset at the same level. The Edge enhancement is applied in order to enhance the salient portion from an image. The effect of edge enhancement can be seen in figure 3.

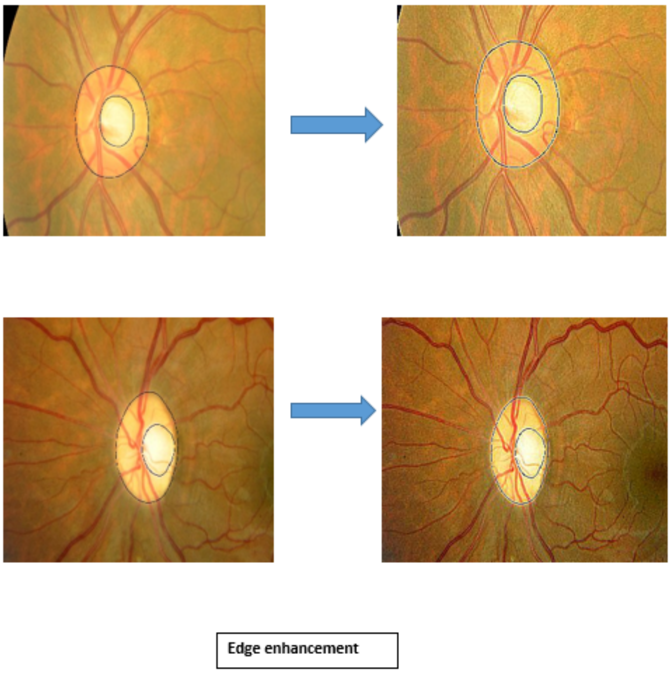

Figure 3: Edge Enhancement of fundus image

3.4 Modelling

Mask RCNN model: Various models such as Full conventional network (FCN), CNN, FCN with U-net have been used in the previous research to segment optic disc and cup accurately. But performing segmentation of the optic disc and cup simultaneously requires further improvement. Thus, this study presents an advanced deep neural network architecture model ”Mask RCNN”. It is an advanced technique built over faster RCNN. While Faster RCNN mainly used for object detection and Mask RCNN widely used in computer vision for detection and instance segmentation from the input image or a video. The architecture of Mask RCNN is well designed which result in three output at a single phase. It gives classes, bounding boxes, and masks of an object. Qin et al. (2019) experiment used two-layer architecture. This study uses three layers of architecture. The figure represents the architecture of Mask RCNN. The proposal regions are generated in the first stage where an object might be present. In second, object class, bounding boxes, and generated masks are predicted based on the output of the first stage. These stages are connected to the FPN backbone, which is a lateral connection, bottom-up, and top-bottom pathway. Here, used ResNet 101 as a bottom pathway that extracts features from the input image. The same size as the bottom pathway, the pyramid feature map is generated by top pathways. Convolution and introduction of operations between two related stages of the two pathways are lateral connection. The propose-regions and FPN top-bottom pathways are scanned by a lightweight RPN network. Anchors are used to scan a feature map in efficient ways. The anchors are of different scales. It contains various levels of feature maps. RPN uses these anchors to determine where an object can be accessed from the feature map and the size of its bounding box. Here, the convolving, upsampling, and down sampling would retain the features in the same relative position as the artifacts in the original picture, and would not mess them around. The procedure in the second stage looks very similar to the first stage, as this stage scans received feature map from the first stages using ROIAlign and it generates a class of object, bounding box, and mask. There is no Anchors are involved in the second stage.

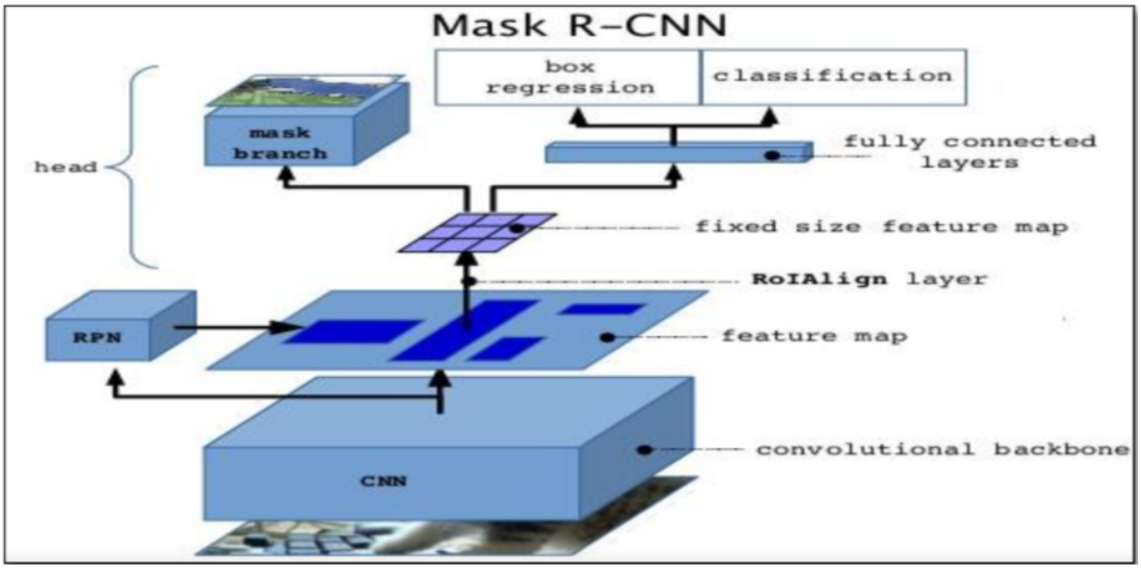

Figure 4: Architecture of Mask RCNN

The proposed method and applied Mask RCNN model in this research work performed better in accuracy, F1 score, precision, and recall than previous research work by (Qin et al.; 2019; Joshua et al.; 2019b).

4 Design Specification

- **Residual Networks (ResNet) 101:** The Mask R-CNN model supports both ResNet 50 and ResNet 101. In this proposed research, the model is trained on ResNet 101 architecture. Its is Residual network with 101 layers, resolved the problem of degradation. Thus the study obtained significant accuracy gain compared to ResNet 50 for RIGA- Dataset.

- **Region proposal network (RPN) :** The RPN is used in this proposed method. It is a region proposal network which detects the object from an input image and can be integrated into any object detection network for allowing useful end to end training of the model. Like CNN's learn to distinguish from feature-map, similarly, RPN draws proposals from the feature map. It has three steps involved. The Anchor boxes are generated from anchor points, which are provided by the feature map. Scales and aspect ratios are two parameters that are used to make candidate boxes. Then in the anchor target process, these targets are generated in comparison to ground-truth boxes with anchor boxes. It calculates the IOU score which returns confidence scores of predicted boxes over GT boxes. This RPN technique has an advantage as requires less computation cost compared to other stats of art. Figure 5. shows the output of RPN.

- **Region of Interest (ROI) Detection:** The Region of interest (ROI) employs a single feature map for all RPN created proposals in a single pass. The problem of

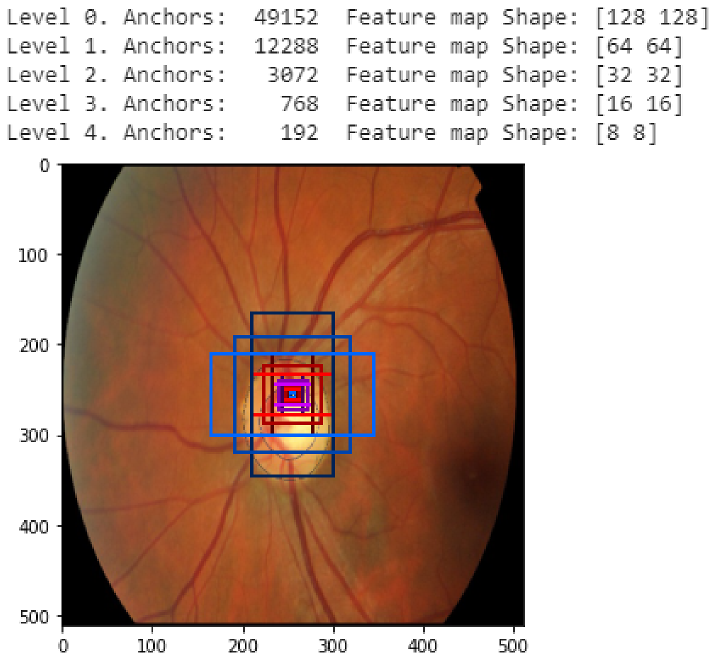

Figure 5: RPN

fixed image size is resolved using ROI pooling techniques as it generates a fixed-size feature map by performing max-pooling limit on input image from non-uniform inputs. Similarly, the research uses ROI Align which brings non-uniform target cells to the same size. Figure 6. shows the output of ROI Align:

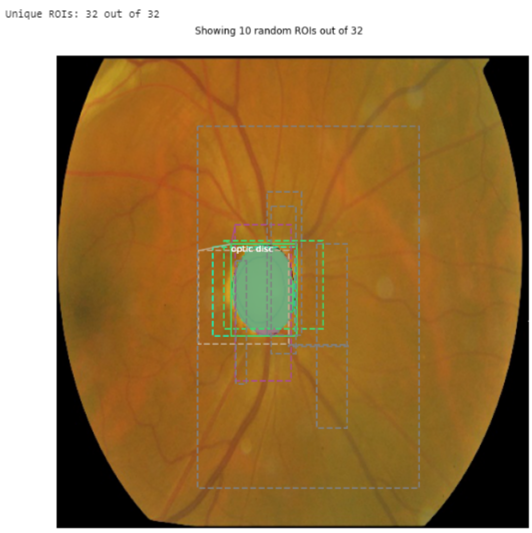

Figure 6: ROI

- **Optimizer:** Stochastic gradient descent optimizer is used here with a learning rate of 0.001. Figure 7 shows resultant output of ROI mask for optic cup and disc.

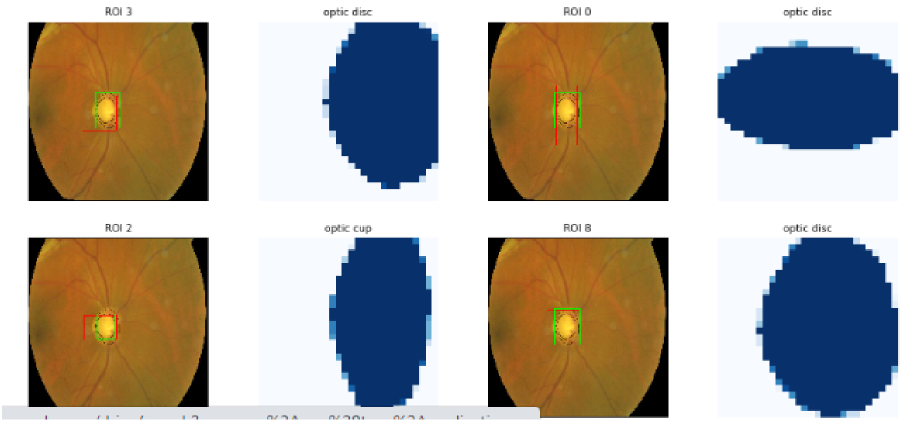

Figure 7: ROI mask for optic cup and disc

5 Implementation

Software Requirement: The model is trained on 48.97 GB run-time GPU and 13 GB RAM which is provided free by Gmail google cloud. Programming language Python version 3 with Tensorflow version 1.14 with Keras 2.1.1 is used for Mask R-CNN architecture. The dataset with folder train, validation, and test are uploaded on google drive. Then data is accessed by mounting drive with google colaboratory notebook using python API.

The implementation, evaluation, and result of the model is discussed in this section. The Mask RCNN model with transfer learning ResNet 50 and ResNet 101 is used in this research work. These transfer learning have MSCOCO pre-trained model weight from ResNet 101 and ResNet 50 with 101 layers and 50 layers respectively. This experiment detects and segments multiple objects (like Cup and Disc) of Glaucoma, Thus, it is considered as a multiclass problem of deep learning. As explained above, following things in the configuration file is set up for implementation of this research work. The model uses backbone of 101 layer that is also called ResNet 101 with a backbone strides [4,8,16,32,64]. To reduce the losses, the attributes such as weights and learning rate of the neural network are changed, which is done by optimizer. Kim et al. (2019) used adam optimizer. This study applied Stochastic Gradient Descent (SGD) with momentum as it accelerates gradient vectors towards the right direction and which leads to faster converging. This method works better and faster than applied optimizer in a previous paper Qin et al. (2019). The value of Learning momentum 0.9 and batch size 1 is used. The study involves a multiclass problem, thus number of classes is 3. One is a background, second optic cup, and third optic disc. The image Max dimension and Min dimension is set to be 512 and 512 respectively as input image is of size 512 x 512. The applied method uses a high minimum Detection confidence level of 0.7, which means the likelihood of a model for genuinely detecting an object in the box. The model trains one image per GPU, therefore image per GPU set to be 1. This model is trained in batches of samples i.e step per epoch is 200 for eight epoch (epoch=8). The proposed study prevents the exploding gradients problem by using a gradient clipping norm of 5.0 as it keeps things stable. Another parameter is Weight decay is used. It prevents unnecessary weight gain. For the training model, datasets is split into 80 and 20 percent for train and validation set.

The Keras call-back function and mean precision function employed to calculate accuracy after each 4th epoch.

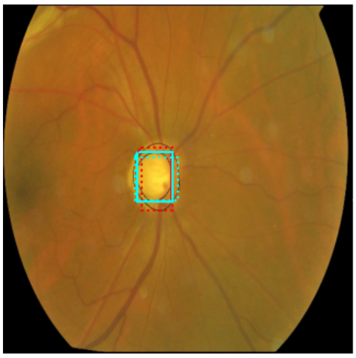

Figure 8: Anchor-box on optic cup and disc of fundus image

The above figure 8 shown Anchor box of the optic cup and disc of fundus image and the detected image of optic cup and disc can be observed through figure 9.

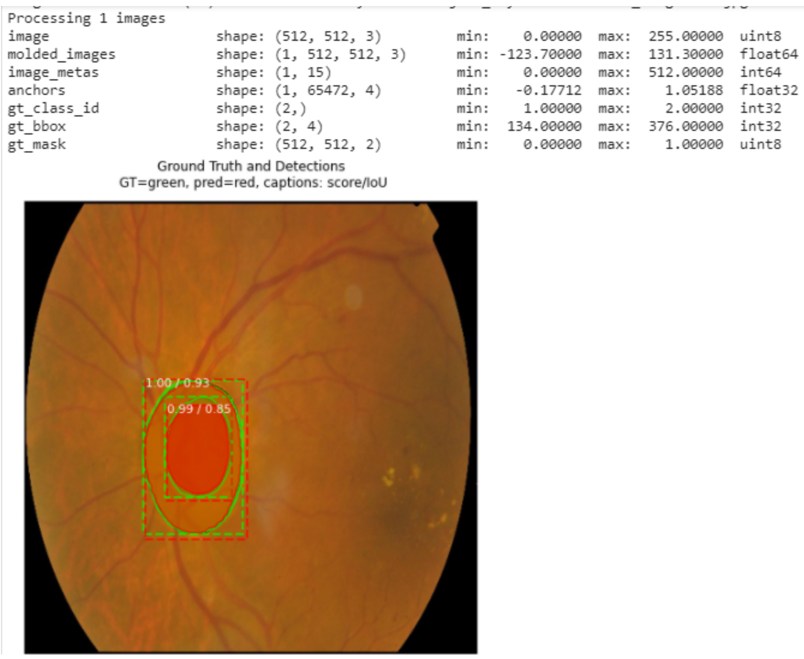

Figure 9: Detected optic cup and disc

6 Evaluation and Result

The comprehensive analysis of the applied techniques is discussed in this section. The various parameter is tuned to get the best performance result. For this research paper, the Mask RCNN model is utilized on the fundus image dataset from the publicly available repository. For evaluation process, four metrics are used to evaluate the performance of model. The loss and accuracy of training and validation data for the assessment process are calculated per epoch. The accuracy and various loss such as bounding box loss, mask loss, class loss, rpn box are plotted for training and validation data. The proposed model in this research work outstripped the result in accuracy with 99% for training and 98% for validation data. Figure 10 shows the segmented optics cup and disc from fundus image. Below figure are of loss graph: Figure 11 draws training and validation loss per epoch. The graph reveals both losses is continuously decreasing for the 7th epoch.

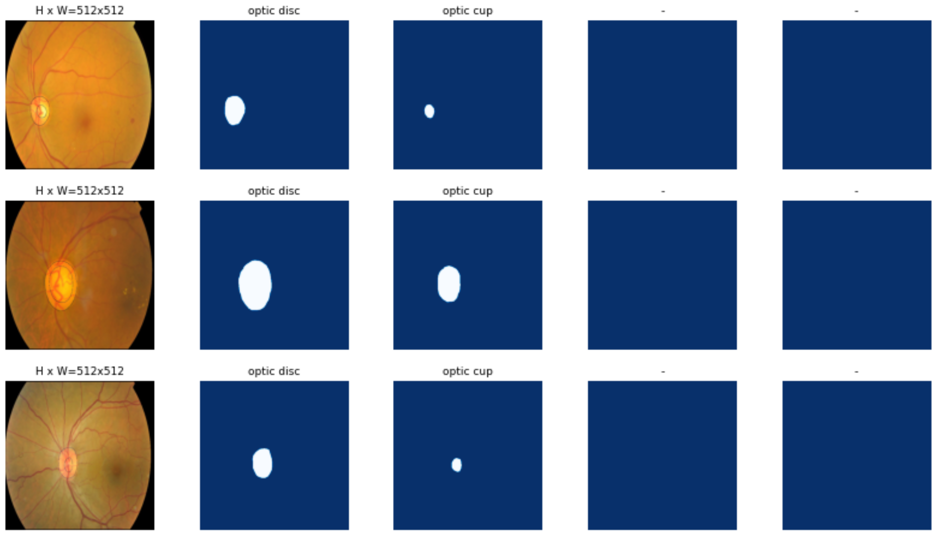

Figure 10: Segmented Optic cup and disc

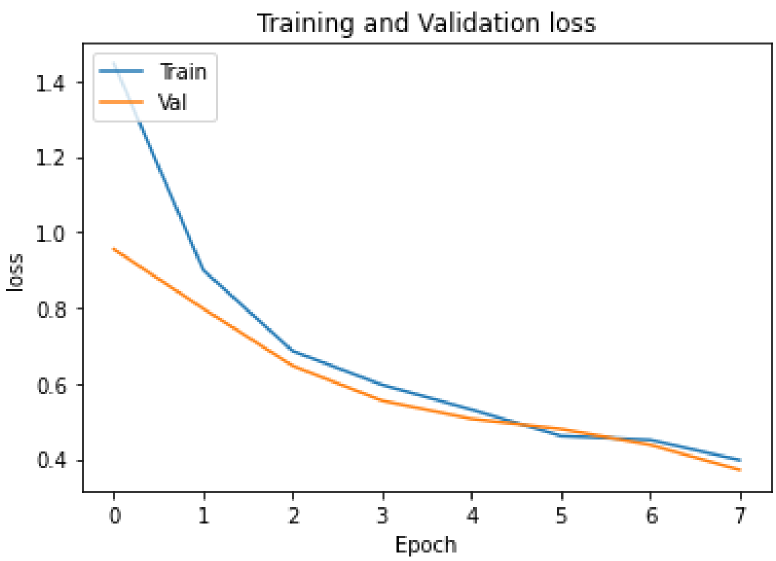

Figure 11: Training and val loss

Figure 12 and 13, shows losses for Masks, RPN box, bounding box.

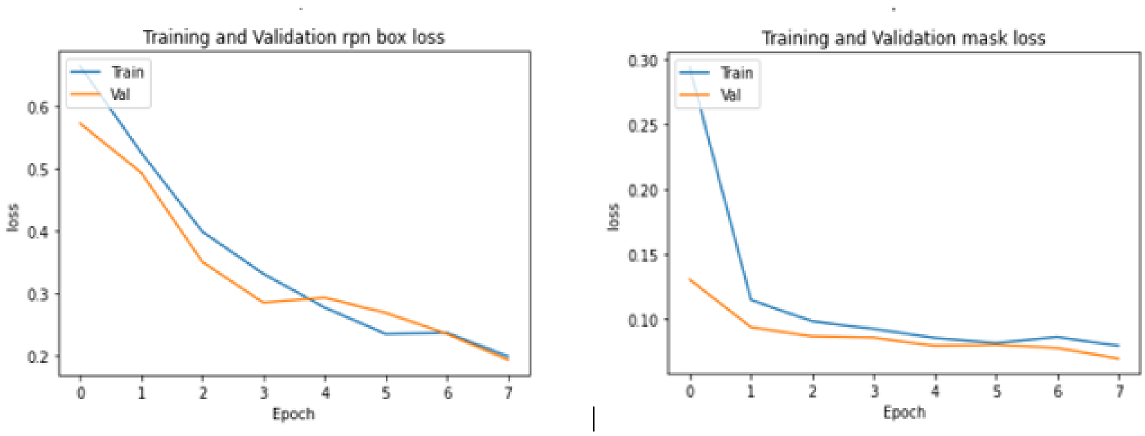

Figure 12: (left) RPN box loss and (right) Masks loss

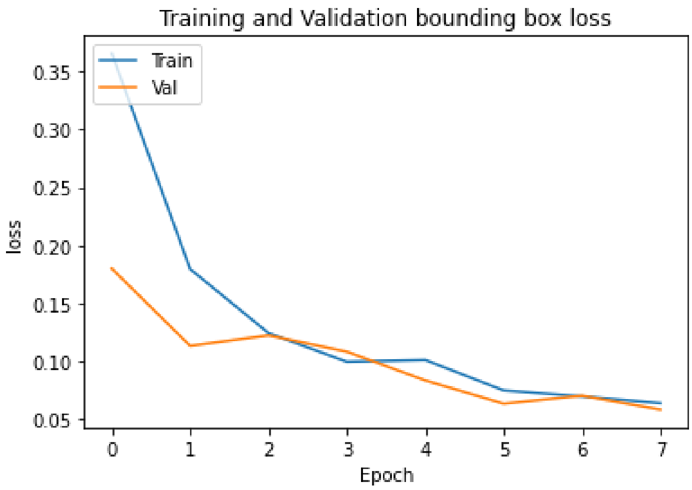

Figure 13: Training and Val Bounding box loss

False-positive and false negative is identified by precision and recall respectively. Below figure 14 shows precision and recall of mathematical formula. For detailed analysis, the combined confusion metrics for multiple classes are calculated to get precision, recall, and F1 score. The confusion matrix helps to calculate count for correct and incorrect predictions on a classification problem.

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

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

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

TP = True positive

TN = True negative

FP = False positive

FN = False negative

Figure 14: Formula of Precision, Recall, F1 score

Confusion matrix can be seen in Figure 15. Diagonal represent correctly classified cases. Here class 1 blank as its background. Class 2 depicts the optic cup and class 3 represents optic disc. The result shows 0.96 precision and 0.95 recall for optic cup and 0.99 precision and 0.94 recall for the optic disc.

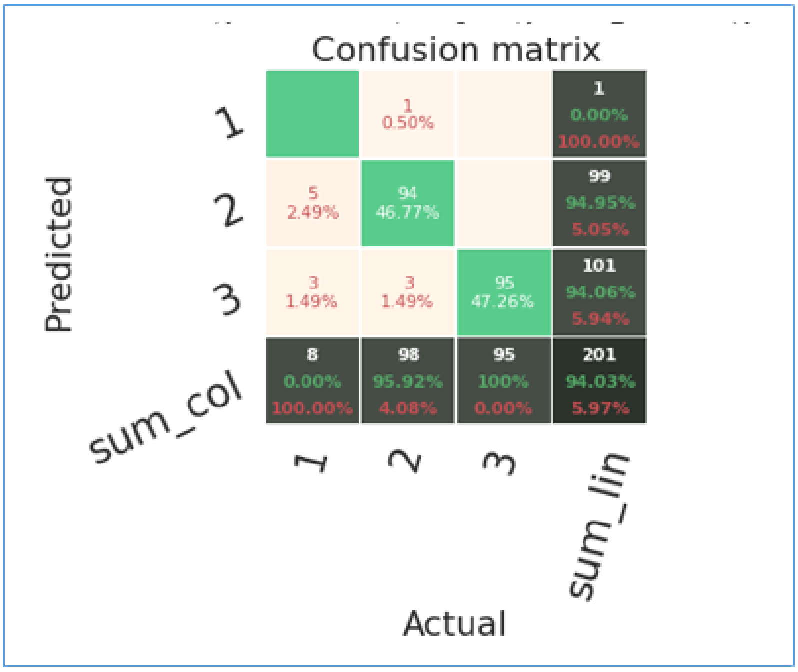

Figure 15: Confusion Matrix

For the balance between precision and recall, the F1 score is also calculated here. This experiment outperformed for optic cup segmentation in f1 score with 0.95 in comparison with a previous study (Kim et al.; 2019).

Table 3: Table summary of evaluation

Metrics	optic cup (in %)	optic disc (in %)
F1 score	0.95	0.97
Recall	0.95	0.94
Precision	0.96	0.99
Accuracy	Training=0.99	Validation=0.98

However, the f1 score for optic disc surpass performance with 0.97 for the used data set. Table 3 is Table summary of evaluation performance.

7 Conclusion and Future Work

To address the answer to the research question, the study proposes the automatic segmentation of the optic cup and optic disc from fundus image. The study also detects the region of cup and disc along with segmentation at the same time, which wasn't the case in any of the previous research. The advance approach is applied over previous research Qin et al. (2019); Joshua et al. (2019b) to overcome the segmentation problem. The applied model "Mask RCNN" showed better evaluation results in term accuracy with 99%. The model architecture is well designed in such a way, it plays a significant role in reducing computation cost. The model has three prominent layers such as ResNet 101, RPN, ROIAlign is carried out in this research work. Other evaluation metrics obtained high F1 score for optic disc achieved 97% and optic cup with 95.43%. As future work, a more complex model can be designed such as GAN with different datasets and approaches.

References

Abbas, K., Khan, P. W., Ahmed, K. T. and Song, W. (2019). Automatic brain tumor detection in medical imaging using machine learning, *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 531–536.

Bonte, S., Goethals, I. and Van Holen, R. (2018). Machine learning based brain tumour segmentation on limited data using local texture and abnormality, *Computers in Biology and Medicine* **98**.

Gupta, M., Gayatri, K. S., Harika, K., Prabhakar Rao, B. V. V. S. N., Rajagopalan, V., Das, A. and Kesavadas, C. (2015). Brain tumor segmentation by integrating symmetric property with region growing approach, *2015 Annual IEEE India Conference (INDICON)*, pp. 1–5.

Joshua, A. O., Nelwamondo, F. V. and Mabuza-Hocquet, G. (2019a). Segmentation of optic cup and disc for diagnosis of glaucoma on retinal fundus images, *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, pp. 183–187.

Joshua, A. O., Nelwamondo, F. V. and Mabuza-Hocquet, G. (2019b). Segmentation of optic cup and disc for diagnosis of glaucoma on retinal fundus images, *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa*, pp. 183–187.

Kassani, S. H., Kassani, P. H., Khazaeinezhad, R., Wesolowski, M. J., Schneider, K. A. and Deters, R. (2019). Diabetic retinopathy classification using a modified xception architecture, *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–6.

Kim, J., Tran, L., Chew, E. Y. and Antani, S. (2019). Optic disc and cup segmentation for glaucoma characterization using deep learning, *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 489–494.

Krinski, B. A., Ruiz, D. V., Machado, G. Z. and Todt, E. (2019). Masking salient object detection, a mask region-based convolutional neural network analysis for segmentation of salient objects, *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, pp. 55–60.

Li, L., Xu, M., Liu, H., Li, Y., Wang, X., Jiang, L., Wang, Z., Fan, X. and Wang, N. (2020). A large-scale database and a cnn model for attention-based glaucoma detection, *IEEE Transactions on Medical Imaging* **39**(2): 413–424.

Lin, T., Cervino, L., Tang, X., Vasconcelos, N. and Jiang, S. B. (2008). Tumor targeting for lung cancer radiotherapy using machine learning techniques, *2008 Seventh International Conference on Machine Learning and Applications*, pp. 533–538.

Madhupriya, G., Guru, N. M., Praveen, S. and Nivetha, B. (2019). Brain tumor segmentation with deep learning technique, *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 758–763.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández Orallo, J., Kull, M., Lachiche, N., Ramírez Quintana, M. J. and Flach, P. A. (2019). Crisp-dm twenty years later: From data mining processes to data science trajectories, *IEEE Transactions on Knowledge and Data Engineering* pp. 1–1.

Mojab, N., Noroozi, V., Yu, P. S. and Hallak, J. A. (2019). Deep multi-task learning for interpretable glaucoma detection, *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 167–174.

Murthi, A. and Madheswaran, M. (2012). Enhancement of optic cup to disc ratio detection in glaucoma diagnosis, *2012 International Conference on Computer Communication and Informatics*, pp. 1–5.

Qian, Y., Liu, Q., Zhu, H., Fan, H., Du, B. and Liu, S. (2019). Mask r-cnn for object detection in multitemporal sar images, *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pp. 1–4.

Qin, P., Wang, L. and Lv, H. (2019). Optic disc and cup segmentation based on deep learning, *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 1835–1840.

Raj, A., Shah, N. A., Tiwari, A. K. and Martini, M. G. (2020). Multivariate regression-based convolutional neural network model for fundus image quality assessment, *IEEE Access* **8**: 57810–57821.

Santhakumar R, Tandur, M., Rajkumar, E. R., Geetha K S, Haritz, G. and Rajamani, K. T. (2016). Machine learning algorithm for retinal image analysis, *2016 IEEE Region 10 Conference (TENCON)*, pp. 1236–1240.

Schmidt, M., Levner, I., Greiner, R., Murtha, A. and Bistritz, A. (2005). Segmenting brain tumors using alignment-based features, *Fourth International Conference on Machine Learning and Applications (ICMLA'05)*, pp. 6 pp.–.

Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T., Barrick, T., Howe, F. and Ye, X. (2016). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in flair mri, *International Journal of Computer Assisted Radiology and Surgery* **12**.

Thangaraj, V. and Natarajan, V. (2017). Glaucoma diagnosis using support vector machine, *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 394–399.

Vuola, A. O., Akram, S. U. and Kannala, J. (2019). Mask-rcnn and u-net ensembled for nuclei segmentation, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 208–212.

Zhou, J., Chan, K. L., Chong, V. F. H. and Krishnan, S. M. (2005). Extraction of brain tumor from mr images using one-class support vector machine, *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 6411–6414.

Zhu, G., Piao, Z. and Kim, S. C. (2020). Tooth detection and segmentation with mask r-cnn, *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 070–072.