

AI Based Ocular Multi-Disease Segmentation and Classification Using U-Net and Efficient-Net

Mudavarthi Kavya¹, Dr.R.V.S. Satyanarayana²

¹M. Tech Student, Department of Electronics & Communication Engineering, S.V. University college of Engineering, Tirupati, A.P.India.

²Professor, Department of Electronics & Communication Engineering, S.V. University college of Engineering, Tirupati, A.P.India.

Abstract—Automated analysis of retinal fundus images can support efficient screening of ocular abnormalities by combining image segmentation and disease classification. This paper proposes an AI-based framework for analysis of Diabetic Retinopathy (DR), Glaucoma, and Age-related Macular Degeneration (AMD) with Normal images as the reference class. The method proposed in this work is based on fundus image preprocessing, U-Net based segmentation, region-of-interest extraction, EfficientNet feature learning and disease classification. The spatial segmentation masks are created by U-Net and the discriminative deep features are extracted by EfficientNet for classification. A 789 sample four class confusion matrix was used for the evaluation of the experimental results. Of these, 728 samples were correctly classified with an overall accuracy of 92.27%. The macro F1-Score achieved was 86.99% and the weighted F1-Score achieved was 92.58%. The correspondence of input images, reference masks, predicted masks and overlays are shown with qualitative segmentation results.

Index Terms—Fundus images, Diabetic Retinopathy, Glaucoma, Age-related Macular Degeneration, U-Net, Efficient-Net, image segmentation, deep learning, multi-label classification

I. INTRODUCTION

The image of the retina contains important information regarding disease related abnormalities, blood vessels, the optic disc and the macula. Retinal image analysis is a time consuming process and needs special ophthalmic skill to analyze large amounts of images. Machine learning techniques can help automate analysis of retinal images and disease screening. This study introduces a hybrid model comprising U-Net and EfficientNet for the

analysis of three vision tasks: Diabetic Retinopathy (DR), Glaucoma, and Age-Related Macular Degeneration (AMD). U-Net is an encoder-decoder network with skip connections that uses pixel-wise segmentation to retain salient spatial information during segmentation [2]. To extract deeper features and classify diseases, the deep feature extraction and disease classification are performed by compound scaling of network depth, width and input resolution using efficientNet [3].

The ODIR-5K dataset consists of 5000 patient records containing images of both the left and right eye fundus, and the disease labels such as normal, diabetes, glaucoma, cataract, AMD, hypertension, myopia, and others are labeled [1]. The proposed pipeline includes image preprocessing, U-Net segmentation, ROI extraction, EfficientNet feature extraction, and disease classification. The Dice coefficient and the IoU are computed for evaluating the segmentation performance; the accuracy, precision, recall, F1 score and AUC are computed for evaluating the classification performance. The integrated approach offers both spatial information and disease-level predictions for automated retinal image analysis.

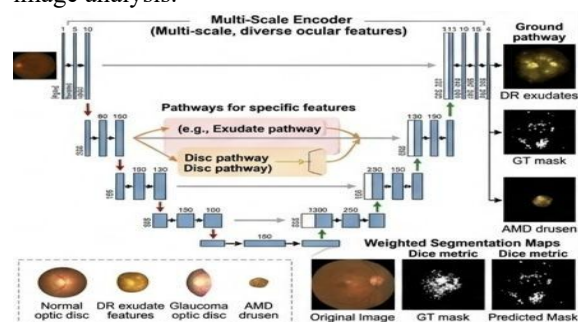


Fig 1. Multi-Scale U-Net Segmentation.

II. LITERATURE REVIEW

U-Net was proposed by Ronneberger et al., in which skip connections were used to maintain spatial relationships and an expanding path was used to produce precise localization while a contracting path was used to establish context. [1] In EfficientNet and their compound scaling principle, Tan and Le proposed to scale the depth, width and input resolution of a convolutional network jointly to enhance the accuracy-efficiency trade-off. [2]

The IDRiD dataset and challenge was created by Porwal et al. for diabetic retinopathy segmentation and grading. The dataset is annotated at the pixel level with important DR lesions and at the image level with DR disease-related grading which is useful for the development and evaluation of automated DR analysis systems [3].

The REFUGE challenge was created by Orlando et al. to establish a standardized framework for the glaucoma assessment based on the fundus photographs. The data can be used for optic-disc/optic-cup segmentation as well as for classification of glaucoma, which enables the study of structural retinal features related to glaucoma [4].

In this paper, Yan et al. introduced a multi-task deep-learning model to classify Age-related Macular Degeneration (AMD). The approach jointly learns clinically meaningful retinal characteristics, including drusen and pigmentary abnormalities, before determining the severity of AMD. The study also used a deep learning model to learn disease-related features and classification simultaneously, which showed that deep learning could provide a promising means for automated assessment of AMD from retinal fundus images. This work is applicable to the current study because pre-classification of the disease by identifying AMD specific retinal characteristics could be useful for reliable classification. [5]

The motivation for this is provided by these studies, which suggest a structure for relating prediction of disease to spatial analysis. The current work thus explores a U-Net plus EfficientNet pipeline for the three target diseases

- DR, Glaucoma and AMD.

III. PROPOSED METHODOLOGY

The six major steps of the proposed framework are (1) acquisition of fundus images, (2) preprocessing, (3) U-Net segmentation, (4) ROI extraction, (5) efficientNet feature extraction, and (6) multi-disease classification. The segmentation and classification stage are related however should be assessed independently due to the different target information used.

3.1 MATERIALS AND METHOD

3.1.1. IDRiD Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRiD) consists of retinal fundus images for analysis of diabetic retinopathy with annotations. The segmentation component delivers pixel-level segmentation for retinal lesions such as: Microaneurysms, Hemorrhages, Hard exudates, Soft exudates.

3.1.2. REFUGE Dataset

The Retinal Fundus Glaucoma Challenge (REFUGE) dataset is used for the purpose of retinal image analysis related to glaucoma. It contains fundus images and annotations for various structures like optic disc, optic cup and information about glaucoma classification.

3.1.3. ODIR-5K Dataset

Disease-level classification is done with the help of the ODIR-5K. There are 5,000 patients in the official dataset that have bilateral fundus photographs and diagnostic keywords. In this study, the diseases that are targeted for the labels are DR, Glaucoma and AMD. ODIR-5K is a multi-label dataset, meaning an image can have more than one label for a disease. The final label encoding should thus be based on the diagnosed information about the eye position and not on the assumption that all images are of exclusively one disease.

Table 1. Performace Evaluation of the segmentation Classification

Stage	Operation	Purpose
1	Image acquisition	Collect retinal fundus photographs and labels.
2	Preprocessing	Resize, normalize and improve image quality.
3	U-Net	Generate pixel-level segmentation mask.

4	ROI extraction	Focus on the segmented retinal region.
6	Classification	Predict Normal/DR/Glaucoma/AMD.
7	Evaluation	Measure classification and segmentation performance.

The standard ODIR-5K data set description does not include a fixed set of lesion masks for the DR, Glaucoma or AMD cases, but instead includes labels and diagnostic keywords for the disease. In the final paper, therefore, the masks in the figures submitted should be described by the actual origin (e.g., external experts annotations, algorithmically generated reference masks).

3.2 MATERIALS AND METHODS

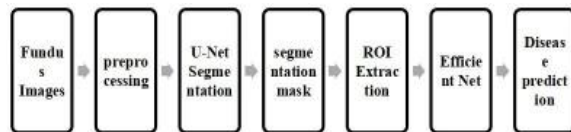


Fig 2. Flow chart of proposed methodology

3.2.1. Dataset Preparation

The classification experiment is conducted on the ODIR-5K dataset or an organization built from the ODIR-5K which includes fundus photographs and disease labels. ODIR-5K comprises 5,000 patients and pictures of both eyes [1]. Since the source annotations may be combined into a compound statement, the label-generating procedure must be clearly expressed in a final publication. The present reported confusion matrix contains four mutually exclusive output categories which are assessed: Normal, Diabetic Retinopathy, Glaucoma and AMD.

3.2.2. Image Preprocessing

Images of the fundus are resized to a common network input dimension, and normalized. Image enhancement can consist of noise and contrast reduction, before training-time augmentation. A form of min-max normalization is:

$$I_{\text{norm}} = \frac{I - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}}$$

where I is intensity of the input and I_{norm} is the normalized value of the input. It is crucial to have consistent pre-processing as ODIR-5K is a set of images acquired with non homogenous conditions.

3.2.3. U-Net Segmentation

The preprocessed fundus image is fed to the UNet segmentation network. U-Net is a combination of encoder, bottleneck and decoder with skip connections. The encoder is responsible for extracting hierarchical features while the decoder gradually recovers the spatial resolution and reconstructs a segmentation mask for each pixel. The network can recognize DR lesions or anatomical structures of interest, like the optic disc and optic cup, depending on the data set. U-Net is a popular convolutional neural network (CNN) architecture for pixel-level biomedical image segmentation. In the suggested ocular disease analysis framework, the U-Net is utilized to recognize and localize the regions of the retina that are relevant to the disease in fundus photographs before classification by EfficientNet-B0. One of the key benefits of U-Net is its encoder-decoder architecture with skip connections that allows it to capture high-level semantic information while retaining fine spatial details. This property can be very helpful in retinal imaging, where many of the abnormalities observed clinically only cover a small part of the image, for example, the small lesions of DR and the structures of the optic disc. Fundus Image.



Fig 3. Structure of U-Net The model generates a mask M

$$M = U(I)$$

Segementation Evaluation:

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

The segmentation mask is then used to segment out the area of interest (ROI) from the retinal image.

The ROI is obtained as:

$$\text{ROI} = I \odot M$$

where $\odot$ denotes element-wise multiplication. This step puts emphasis on the particular area of the retina and minimizes the effect of the background pixels that are not relevant.

3.2.5. EfficientNet Classification

The extracted ROI is then passed on to EfficientNet-B0 for discriminative learning of retinal features in

convolutional layers and compound scaling of network depth, width and input resolution [3]. In this 4-class experiment, this is the reason that a Softmax output layer is used: Each sample belongs to only one mutually exclusive class: Normal, DR, Glaucoma or AMD. The class which is predicted is

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

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

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

$$\text{F1-Score} = \frac{\text{Precision/Recall}}{\text{Precision+Recall}}$$

IV. RESULTS AND DISCUSSIONS

There are two kinds of output provided: qualitative segmentation visualizations and a quantitative confusion matrix. The first visualization includes four disease categories (Normal, DR, Glaucoma, AMD) and an original fundus image, reference mask, predicted mask, and prediction overlay. The second figure shows the numerically presented confusion matrix of the classification analysis.

3.2.4. ROI Extraction

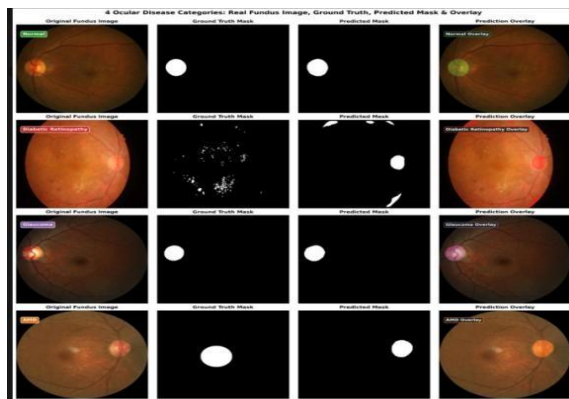


Fig 4. Ocular Disease classification of normal,DR, Glaucoma and AMD

The results of segmentation and localization are shown qualitatively in figure 1. The regions related to compact discs are shown in the Normal and Glaucoma examples, and the DR example has several lesion-like structures in the reference mask. An example of a central abnormality pattern is seen in the AMD example. The observations are qualitative,

and the figure does not represent a numerical Dice or IoU value.

4.1 Confusion Matrix

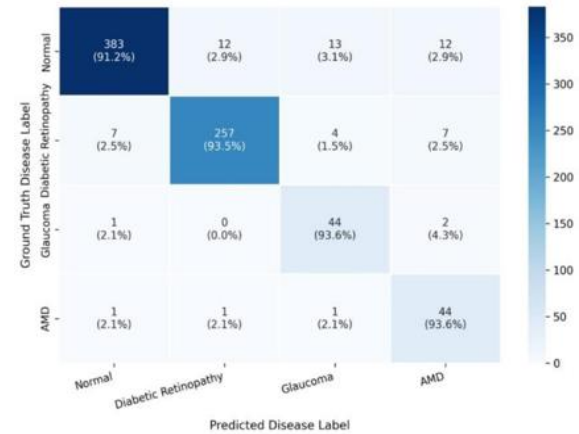


Fig 4. Confusion matrix Of four Classes

There are 789 samples that have been evaluated in the confusion matrix. The diagonal entries are 383 Normal, 257 DR, 44 Glaucoma and 44 AMD predictions. The sum of the diagonal is 728 and 61 are misclassified. So, the overall accuracy of the classification is:

$$\text{Accuracy} = 728 / 789 \times 100 = 92.27\%$$

Normal makes 383 correct predictions of 420 samples. DR has 257 correct predictions out of 275 samples. Glaucoma has 44 correct predictions out of 47 samples, and AMD has 44 correct predictions out of 47 samples. Percentages listed within the provided matrix are row-wise percentages which indicate the percentage of each ground truth class in each prediction class.

Table 2. Class-wise classification performance

Class	Support	Precision (%)	Recall (%)	F1-score (%)
Normal	420	97.70	91.19	94.33
Diabetic Retinopathy	275	95.19	93.45	94.31
Glaucoma	47	70.97	93.62	80.73
AMD	47	67.69	93.62	78.57
Macro Average	789	82.89	92.97	86.99
Weighted Average	789	93.45	92.27	92.58

The model achieves an 82.89% precision, 92.97% recall, and 86.99% F1-score at the macro level. Weighted precision, recall and F1-score are 93.45%, 92.27% and 92.58%, respectively. The difference between macro and weighted measures is due to a difference in class supports, where there are more samples from Normal and DR than there are from Glaucoma and AMD.

V.CONCLUSION

This paper presented an AI-based ocular multi-disease segmentation and classification framework using U-Net and EfficientNet for the analysis of Diabetic Retinopathy, Glaucoma and AMD. The methodology combines preprocessing, U-Net-based segmentation, ROI extraction and EfficientNet-based classification. The supplied qualitative outputs demonstrate a four-panel representation of retinal input, reference mask, predicted mask and overlay, while the supplied confusion matrix provides quantitative evidence for four-class classification.

From 789 samples, 728 were correctly classified, resulting in an overall accuracy of 92.27%. The calculated macro F1-score is 86.99%, while the weighted F1-score is 92.58%. DR produced an F1-score of 94.31%, Normal 94.33%, Glaucoma 80.73% and AMD 78.57%.

These values should be reported together with the exact test protocol and class-definition procedure.

The proposed framework provides a foundation for combining disease recognition with spatial retinal analysis. Future work should include verified disease-specific pixel-level annotations, external validation, class-imbalance handling, calibration of prediction thresholds, and comparison with additional lightweight and transformer-based architectures. The system should be regarded as a research prototype rather than a clinically validated diagnostic device.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[2] M. Tan and Q. V. Le, "EfficientNet:

Rethinking Model Scaling for Convolutional Neural Networks," in Proc. 36th International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.

[3] P. Porwal et al., "IDRiD: Diabetic Retinopathy—Segmentation and Grading Challenge," Medical Image Analysis, vol. 59, Art. no. 101561, 2020, doi: 10.1016/j.media.2019.101561.

[4] J. I. Orlando et al., "REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment From sFundus Photographs," Medical Image Analysis, vol. 59, Art. no. 101570, 2020, doi: 10.1016/j.media.2019.101570.

[5] X. Liu et al., "Optic Disc Segmentation in Fundus Images Using Adversarial Training," IET Image Processing, 2019, doi: 10.1049/iet-ipr.2018.5922.

[6] J. Yan et al., "A Multi-Task Deep Learning Model for the Classification of Age-Related Macular Degeneration," Scientific Reports, 2019.