Detection of Optic disc (OD) & Optic cup (OC)

Jad Kutbi
Student number: 6827426

Cem Ilgn
Student number: 6832781

Ahmed Abubakar
Student number: 6731105

Manish Panda
Student number: 6614056

***Abstract*—With the use of a convolutional neural network (CNN), more precisely the UNet model, this study investigates the segmentation of the optic disc (OD) and optic cup (OC) in retinal pictures using the "Glaucoma Fundus Imaging Datasets." The project entails preprocessing to standardize image forms, then partitioning the dataset into sets for testing, training, and validation. Model robustness is improved by transformations. The Sørensen–Dice coefficient is used in the evaluation process. To address issues like image quality, the results will be viewed and analyzed. This work presents machine learning algorithms applied to retinal image processing, with potential applications in automated detection of glaucoma.**

***Index Terms*—optic disc, optic cup, glaucoma fundus.**

I. Introduction

Millions of people worldwide are impacted by glaucoma, which is one of the main causes of permanent blindness. The disease's progression must be slowed down and serious visual impairment must be avoided by early detection and intervention [1]. One of the most important aspects of diagnosing glaucoma is evaluating the optic nerve head, namely the optic disc (OD) and optic cup (OC). Glaucoma diagnosis relies heavily on the Cup-to-Disc Ratio (CDR), which is the ratio of the diameters of the optic cup to the optic disc. Conventional approaches to this analysis mostly depend on the manual evaluation of these structures by knowledgeable medical professionals, which can be laborious, subjective, and prone to error. By precisely identifying the optic disc and optic cup in retinal fundus pictures, automated techniques that can help with the early detection of glaucoma have been made possible by developments in digital imaging and machine learning. Convolutional neural networks (CNNs), one of several machine learning techniques, have become a very useful tool for medical image analysis because of its capacity to derive features and hierarchical representations straight from the data. With the UNet architecture in particular, which is well known for its effectiveness in medical image segmentation tasks, this study seeks to use the potential of CNNs. [2] The aim is to create a robust model that can reliably estimate the CDR by autonomously identifying and segmenting the optic disc and optic cup from retinal pictures. The project's goal is to improve glaucoma screening and diagnosis by automating this procedure, which will ultimately help at-risk persons maintain their eyesight.

II. Literature Review

The development of image processing and machine learning techniques has had a significant impact on recent advances in the automatic recognition of the optic disc (OD) and optic cup (OC) in retinal pictures. At first, approaches depended on traditional image processing techniques like morphological operations and thresholding. Later, machine learning techniques using hand-crafted features and models like support vector machines (SVM) and random forests replaced these techniques [3]. Nevertheless, these methods frequently encountered difficulties due to anatomical variations and inconsistent image quality. Performance has greatly improved since the advent of deep learning, specifically convolutional neural networks (CNNs). CNNs automate feature extraction and adjust to the intricate variability found in medical images—a critical development made possible by architectures such as U-Net, which are now the industry standard for segmenting medical images. [4]

For unsupervised hierarchical feature extraction from images without user specification, the U-Net architecture—which is frequently used in medical image analysis—performs exceptionally well [5]. An expanding path for accurate localization—which is critical for medical imaging—and a contracting path for context collection are features of this CNN. Skip connections provide multi-level feature integration and improve fine-grained segmentation, which is a key component of U-Net's efficacy. Furthermore, given the dearth of annotated medical images, data augmentation increases the size of training datasets, enhancing model generalization in a variety of scenarios [6].

Even with these tremendous advances, there are still many obstacles to overcome, especially when it comes to computing efficiency—which is critical for real-time applications—and generalizing across incredibly diverse datasets [7]. The optimization of deep learning models for increased performance and reduced resource usage is still being investigated in current research. Furthermore, ongoing research is being done on how these models respond to low-quality images that are frequently seen in clinical settings [8]. Future studies can close these gaps and improve CNNs' usefulness in clinical settings by making them more reliable and approachable for everyday diagnostic use. These initiatives are in line with the overarching objective of enhancing early detection and treatment of diseases like glaucoma by utilizing more precise and effective medical imaging technologies.

III. Methodology

A. Glaucoma Fundus Imaging Datasets

The optic disc (OD) and optic cup (OC) in retinal fundus images—which are essential for the diagnosis of glaucoma-related conditions—are separated using a machine learning model trained on the Glaucoma Fundus Imaging Datasets, a composite of three distinct datasets. Preprocessing involves scaling or cropping each image to a standard size so that the aggregated datasets are consistent. In order to avoid overfitting and improve the resilience of the model, data augmentation techniques including random rotations, and horizontal and vertical flips are used. To guarantee equal data distribution and facilitate more effective and efficient model training, normalize pixel values using specified mean and standard deviation parameters.

B. Model selection

CNNs are the best at identifying spatial hierarchies in pictures, which makes them perfect for segmenting intricate structures in fundus images, such as the optic disc and cup. The architecture of UNet, which consists of a symmetric expanding path for precise localization and a contracting path to collect context, makes it especially well-suited for medical image segmentation. Because of its effectiveness in learning from a small number of images—which is frequently the case in medical imaging studies—UNet is suggested due to its strong performance on similar tasks.

C. Implementation

DataLoader in PyTorch is used to load data in batches, which allows for efficient memory management. Resizing, tensor conversion, normalization, and random augmentations are examples of data preprocessing techniques. The UNet model uses pooling layers to minimize spatial dimensions while maintaining important characteristics. It consists of many convolutional layers with ReLU activations and batch normalization. For training, a learning rate scheduler that reacts to validation loss and three different optimizer's are used namely, Adam, SGD and RMSpropsay. Throughout training sessions, the model reduces the loss of binary cross-entropy. 10% of the dataset is set aside for validation in order to adjust the model's parameters and avoid overfitting, and the remaining 10% is set aside for final testing in order to assess the model's performance on fresh, untested data. The remaining 80% of the dataset is used for training, which focuses on adjusting the model's weights for accurate segmentation tasks.

D. Evaluation Metrics: Sørensen–Dice Coefficient

The model's performance is assessed using the Sørensen–Dice coefficient, sometimes referred to as the F1 Score or Dice Similarity Coefficient. It assigns a score ranging from 0 (no overlap) to 1 (perfect overlap) based on how closely the anticipated segmentation and the ground truth overlap. It is computed by dividing the total number of pixels in both the true and predicted masks by twice the area of overlap between them, plus a smoothing term to prevent division by zero.

IV. Experimental Results

A. OD & OC as a single class

Consistency in batch processing during neural network training requires preprocessing retinal pictures and masks to a standard size of 256x256 pixels. To maintain binary values—which are essential for precise segmentation—masks use nearest neighbor interpolation. By normalizing pixel intensity scales, conventional ImageNet parameters—[0.485, 0.456, 0.406] for the mean and [0.229, 0.224, 0.225] for the standard—are applied to the images in order to stabilize learning. Model robustness against changing imaging settings is improved by optional augmentations such flips and rotations, which increase the variability of the dataset. To train accurate segmentation models, mask binarization is necessary to guarantee that pixel values are either strictly 0 or 1. This creates distinct segmentation boundaries for the optic disc and cup.

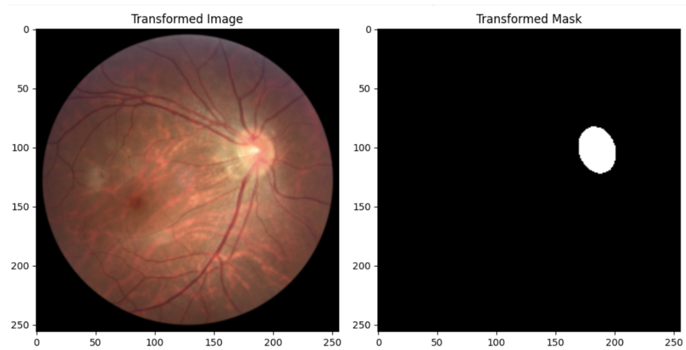

Fig. 1. Preprocessing of the Optic Disc (OD) and Optic Cup (OC) as a single class

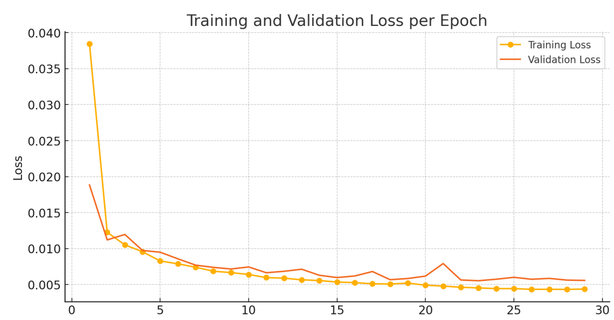

Fig. 2. Training & Validation loss per epoch

Effective learning is demonstrated by the training and validation losses graph, where high model performance and generalization are indicated by both metrics declining over epochs. Since there is no gain in the validation Dice score, the training ends early at epoch 29, indicating that more training could result in overfitting. Variations in validation loss signify phases in which the model adapted to characteristics that were not as prevalent in the training set. The model's rising accuracy in segmenting medical pictures is demonstrated by the convergence of both losses and sporadic improvements in Dice score. To improve robustness and level out validation loss spikes, training strategy adjustments could be made to further optimize the process.

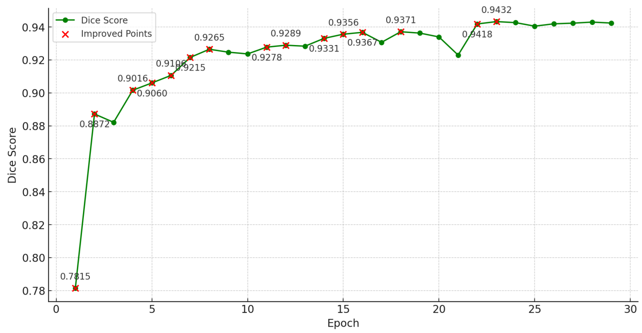

Fig. 3. Dice score per epoch

The model's learning trajectory and efficacy in segmenting medical images are depicted in the Dice Score graph, which prominently displays notable advancements in the early epochs. This quick learning suggests strong initial adaption to the properties of the dataset. Up until about epoch 18, the performance increases gradually before plateauing, indicating that the model is getting close to its learning capacity. In order to prevent overfitting and guarantee that the model maintains its capacity for generalization, the early stopping mechanism essentially ends training at epoch 29. The graph's noticeable gains demonstrate how adaptable the model is to training modifications. Given the essential significance of medical picture segmentation tasks, the modest oscillations in later epochs show possibility for further optimization through different learning rates, augmentation methodologies, or alternate architectures.

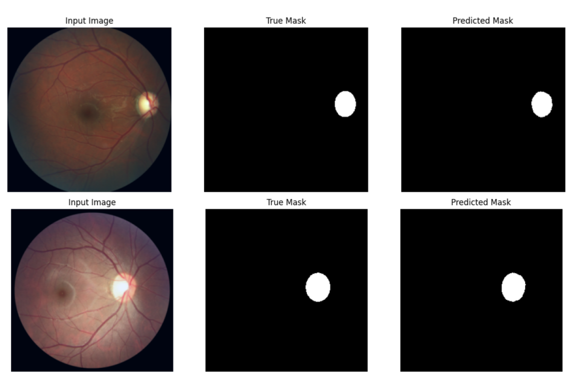

Fig. 4. Prediction with True Mask

With Dice coefficients average approximately 0.9425, the images demonstrate the performance of a U-Net model trained for segmenting the optic disc in retinal images. The outcomes show how well the model consistently and accurately detects the optic disc in a wide range of images, even in the presence of occlusions, focus changes, and lighting differences. The robust model performance even under less than perfect settings is indicated by the tight match between the true and predicted masks. These results show how adaptable the model is to many imaging circumstances, which makes it a useful tool for automated medical diagnostics when accurate and consistent image segmentation is essential.

B. Separate models to detect OD & OC

Preprocessing techniques encompass scaling, normalization, and data augmentation methods such as rotation, flipping both horizontally and vertically. By adding diversity and strengthening the model's resistance to varying orientations and lighting conditions in the images, these techniques are essential for efficiently training the neural network. The graphs' lowering loss and stabilizing validation metrics show how the modifications improve the model's ability to generalize to new data. In addition to improving model training, this preprocessing probably played a role in the high performance measures, including the Dice score increases shown over the course of the epochs.

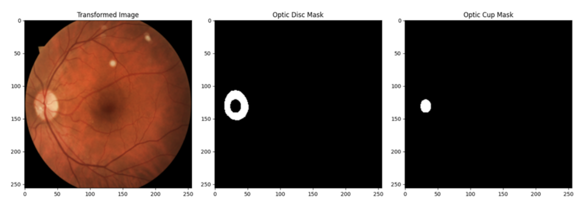

Fig. 5. Preprocessing of the Optic Disc (OD) and Optic Cup (OC) as separate classes

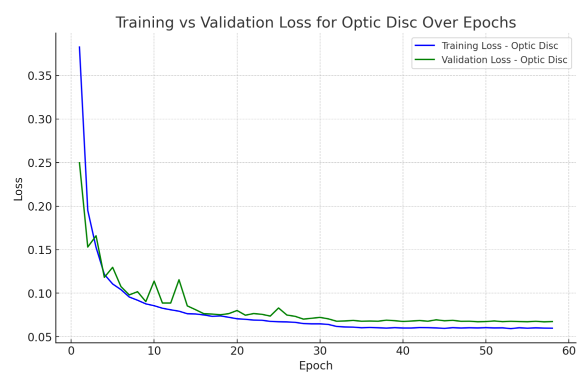

Fig. 6. Training versus validation loss for optic disc over epochs

Both graphs for the Optic Disc and Optic Cup show decreasing training and validation losses over 58 epochs, indicating successful learning and generalization. The convergence of these losses suggests good model fit without significant overfitting, especially in later epochs, hinting at the potential benefit of implementing early stopping. When training the Optic Disc (OD) and Optic Cup (OC) as a single class, the average Dice Coefficient of 0.9425 shows good model performance, indicating a high degree of agreement between the true and predicted segmentation masks. This is in

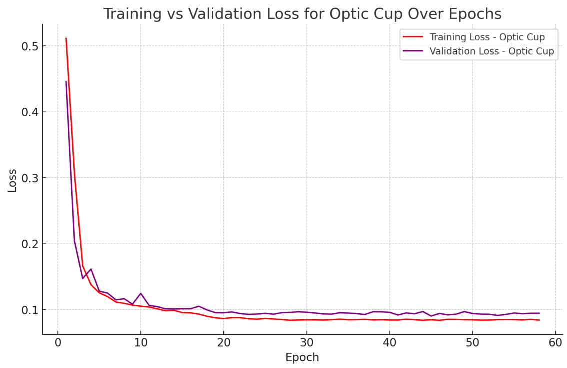

Fig. 7. Training versus validation loss for optic cup over epochs

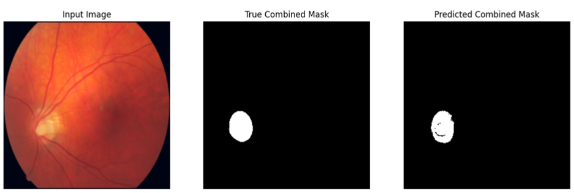

Fig. 8. Prediction with true mask

TABLE I
COMPARISON OF DICE SCORES FOR DIFFERENT TRAINING METHODS

Training method	**OD and OC trained as a single class**	**OD and OC as two separate classes**
Dice Score	**0.9425**	0.9286

contrast to the case when a Dice score of 0.9286 is obtained by training OD and OC as distinct classes. While each method yields excellent results, training both methods as a single class slightly beats training them separately, suggesting that merging the two structures into a single class could improve the model's segmentation accuracy for these specific anatomical traits.

C. Effect of the Adam optimizer on the OD & OC as a single class

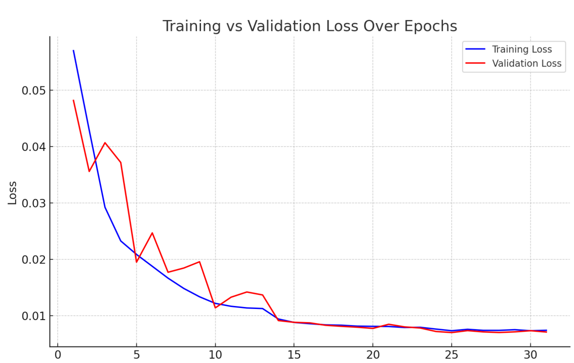

Fig. 9. Training versus validation loss over epochs

The model yielded a high average Dice Coefficient of 0.9193 using the Adam optimizer. This implies that there is a high degree of agreement between the projected segmentation masks and the ground truth masks, which is necessary for precise medical imaging analysis. Adam's successful application, early halting strategies, and efficient scheduling of learning rates all demonstrate the model's ability to produce accurate and reliable predictions. This performance level is particularly important in medical applications as segmentation accuracy directly affects diagnosis and treatment decisions. The results validate the effectiveness of the chosen optimizer and training technique and show the potential of the model in clinical settings. Nevertheless, the model obtained an even higher Dice Coefficient of 0.9425 using the SGD optimizer, suggesting that SGD was much better than the Adam optimizer.

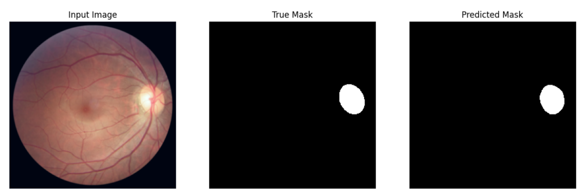

Fig. 10. Prediction with true mask

D. Effect of RMSprop optimizer

Fig. 11. Training versus validation loss over epochs for RMSprop optimizer

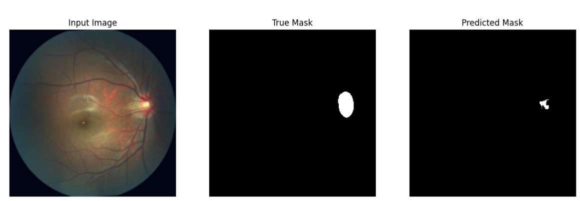

Fig. 12. Effect of focal loss

By normalizing gradients using a moving average of squared gradients, the RMSprop optimizer was able to deal with non-stationary objectives with an average Dice Coefficient of 0.6567 in the U-Net model for image segmentation.

TABLE II
COMPARISON OF OPTIMIZER PERFORMANCE USING DICE SCORE

Optimizers	**SGD**	**Adam**	**RMSprop**
Dice score	**0.9425**	0.9193	0.6567

TABLE III
COMPARISON OF LOSS FUNCTIONS USING DICE SCORE

Losses	**Cross-entropy**	**Combined Cross-entropy and Dice score**	**Focal loss on CE**
Dice score	0.9425	**0.9426**	0.9221

RMSprop's performance seems subpar for this particular task, though, when contrasted with the Adam optimizer, which obtained a higher Dice Coefficient of 0.9193. Adam has an advantage in that its adaptive learning rate mechanism can potentially result in more stable and effective convergence in complicated neural network designs by adjusting rates based on the first and second moments of gradients. This implies that Adam is more appropriate for jobs requiring many parameters and big datasets, while RMSprop might need additional parameter adjustment to meet Adam's efficiency in picture segmentation applications.

E. Effect of dice score loss with Cross-entropy

By employing the SGD optimizer in combination with a Dice score loss function and Cross-Entropy, the model managed to attain a remarkable average Dice Coefficient of 0.9426. This illustrates how well these loss measurements were integrated, striking a balance between segmentation overlap and pixel-level accuracy. Comparatively, Adam produced a Dice Coefficient of 0.9193 and an RMSprop score of 0.6567, indicating that although Adam converges quickly, in certain segmentation tasks SGD paired with a loss technique can perform better than expected. This configuration emphasizes how important it is to match the optimizer and loss strategy to the needs of the model in order to optimize learning dynamics and segmentation quality.

F. Effect of focal loss

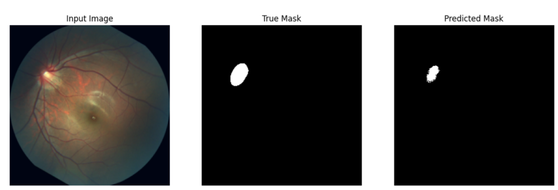

Fig. 13. Effect of focal loss

The Dice Coefficient for the model's performance utilizing the Focal Loss method and the SGD optimizer was 0.9221. By concentrating on difficult, wrongly classified instances, Focal Loss addresses class inequality in particular. With complex segmentation applications like medical imaging, where it is essential to precisely identify examples that are difficult to classify, this property makes it more valuable. The table shows that while Focal Loss modifies the cross-entropy loss by favoring difficult cases over correctly classified ones, improving learning in critical areas, traditional cross-entropy and a Combined cross-entropy with dice score approach perform marginally better than Focal Loss in terms of dice score, landing at 0.9425 and 0.9426, respectively. Although Focal Loss had a lower Dice score, its capacity to manage datasets with notable imbalances and intricate segmentation needs highlights the significance of customizing loss functions for unique project issues.

V. CONCLUSION

High Dice coefficients were obtained when a UNet convolutional neural network was used to segment the optic disc and cup in retinal pictures, which is essential for diagnosing glaucoma. When training the optic disc and cup together, the accuracy was higher, with a Dice score of 0.9425 as opposed to 0.9286 when training them separately. SGD optimizer fared better than other evaluated optimizers, with the greatest Dice coefficient, indicating that it can improve automated glaucoma detection.

REFERENCES

[1] J. B. Jonas, S. Amini, J. M. Goldstein, et al., ”Glaucoma,” The Lancet, vol. 390, no. 10108, pp. 2183–2193, Nov. 2017. [Online]. Available: https://doi.org/10.1016/S0140-6736(17)31469-1
[2] O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany, 2015, pp. 234-241.
[3] T. M. Lehmann, C. Gonner, and K. Spitzer, ”Survey: Interpolation methods in medical image processing,” IEEE Transactions on Medical Imaging, vol. 18, no. 11, pp. 1049-1075, Nov. 1999.
[4] O. Ronneberger, P. Fischer, and T. Brox, ”U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Munich, Germany, 2015, pp. 234-241.
[5] F. Milletari, N. Navab, and S.-A. Ahmadi, ”V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, 2016, pp. 565-571.
[6] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, ”Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images,” in Advances in Neural Information Processing Systems, vol. 25, Lake Tahoe, NV, 2012, pp. 2843-2851.
[7] G. Litjens, T. Kooi, B. E. Bejnordi, et al., ”A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60-88, Dec. 2017.
[8] A. Esteva, B. Kuprel, R. A. Novoa, et al., ”Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115-118, Feb. 2017.