Glaucoma Diagnosis Using Hybrid Neural Encoder Decoder Based Unet Hybrid Inception

Akanksha Bali

Vibhakar Mansotra
ORIGINAL STUDY

Glaucoma Diagnosis Using Hybrid Unet Inception

Akanksha Bali*, Vibhakar Mansotra

Department of Computer Science and IT, University of Jammu, Jammu, Jammu and Kashmir, India

Abstract

The anatomy of the retinal fundus picture, which includes the optic disc, affects how glaucoma characteristics are extracted. One of the main causes of blindness in the globe is glaucoma. Correct glaucoma screening in its early phases is difficult since the condition only becomes apparent when its symptoms are severe. If glaucoma is not treated, the optic nerve head will be harmed, resulting in vision loss. Nevertheless, there are currently not enough eye experts accessible, and eye screening is personal, time-consuming, and labor-intensive. Therefore, this is a real issue that can be addressed by automatically diagnosing glaucoma using deep learning algorithms. Routine glaucoma screening is thus recommended and required. This necessitates the use of an automated segmentation system that can precisely identify the areas with margins and help ophthalmologists monitor and diagnose the severity of glaucoma early on. A hybrid density-ED-UHI encoder-decoder-based U-Net hybrid inception model with 15-fold cross-validation is presented in the recommended research for the diagnosis of glaucoma. The proposed model was trained and validated using the REFUGE (Retinal Fundus Glaucoma Challenge) dataset. For the segmentation of glaucoma, the model achieves 99.9% accuracy, 97.2% area under the curve, 99.9% sensitivity, and 93.7% specificity. This remarkable outcome implies that fundus pictures’ blood vessel segmentation may be used as a substitute for automatically detecting glaucoma.

Keywords: Glaucoma, InceptionV3, REFUGE, ResNet18, ResNet34, Segmentation, UNet

1. Introduction

The optic nerve is harmed by a series of eye conditions known as glaucoma, which results in progressive and permanent vision loss. A common risk factor is increased intraocular pressure, yet it can also happen when eye pressure is normal. As symptoms frequently appear later than expected, early identification and treatment are essential to preventing vision loss or blindness. One symptom of glaucoma is an advancement in intraocular pressure within the eyeball and another symptom is poor vision that occurs from damage to the optic nerve driven by optic neuropathy. A case of glaucoma may harm the visual nerve. When it reaches a high degree, it becomes clear. When a patient's eyesight is reduced by 70%, glaucoma can be identified. As a result, it is critical to get glaucoma checked in the eyes on a frequent basis. The three types of glaucoma are open-angle or primary, common pressure, and angle closure. Primary glaucoma is the most prevalent type of glaucoma. Over time, the eye's seepage system becomes insufficient for this kind of glaucoma. This decrease in the seepage structure causes the eye's weight to steadily rise. Angle-closure glaucoma is a less prevalent kind of the condition. This glaucoma is blocking the back of the eye. Even when the eye weight is within normal limits, normal-strain glaucoma damages the optic nerve (Bali and Mansotra, 2023a).

To diagnose glaucoma, ophthalmologists can measure parameters such as corneal thickness or intraocular pressure and conduct visual field tests. Furthermore, color retinography, a commonly utilized kind of retinal imaging, allows direct observation of crucial illness symptoms such as abnormalities in the optical disc, peripapillary atrophy, or retinal nerve fiber layer anomalies. The deformation of the optical disc is the symptom of glaucoma that has been studied the most in terms of early identification. Medical image processing-based glaucoma diagnostic techniques are currently

Received 27 December 2023; revised 3 April 2024; accepted 5 April 2024.
Available online 30 May 2024

* Corresponding author at: Department of Computer Science and IT, University of Jammu, Jammu, 180016, Jammu and Kashmir, India. E-mail address: akankshabali5@gmail.com (A. Bali).

https://doi.org/10.58491/2735-4202.3214
2735-4202/0 2024 Faculty of Engineering, Mansoura University. This is an open access article under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).
becoming more popular than more conventional research. In these cases, it is important to take note of a number of features of the ocular retinal structure, such as the peripapillary atrophy, cup, optic nerve head, and the surface of the retinal nerve fibers. An image of a glaucoma-affected eye and a healthy eye are shown in Fig. 1 (Bali and Mansotra, 2021a; Bali and Mansotra, 2022).

The five primary components of the automatic detection technique are segmentation, classification, feature extraction, feature selection, and picture preprocessing. An important step in picture improvement is preprocessing, which involves eliminating any extraneous portions of the retina to improve accuracy. Numerous filters, such as mean, median, gaussian, and others, can be employed for preprocessing. Several morphological procedures are then utilized for the extraction and selection of features, after which several algorithms are used for the segmentation and classification (Bali and Mansotra, 2021b).

According to current methodology, many approaches are used to identify glaucoma depending on several modalities of retinal imaging. The deep learning (DL) technology will be used in the suggested way to diagnose glaucoma. The following defines the paper’s primary contribution:

(a) This study segments retinal lesions semantically using a modified U-Net model.
(b) The encoder used in the model is a deep network encoder called ‘InceptionV3.’ The model employs a unique up sampling method that resizes the convolution nearest neighbor by using pixelwise periodic shuffling convolution.
(c) In comparison with traditional approaches, our method produces retinal images devoid of checkerboard artifacts and accelerates network convergence. Comparing the proposed model to other research, cutting-edge results were obtained for glaucoma identification.
(d) Using a hybrid U-Net architecture with an inception design and several kernel extractions, the work addresses deep encoding.
(e) The study evaluates the glaucoma model’s performance for a thorough cross-platform analysis.

This is how the rest of the paper is organized. In Section II, relevant literature on automated glaucoma segmentation is reviewed. The approach that is suggested is discussed in Section III. Section IV discusses the work’s findings and comments. Finally, conclusions are given in Section V.

2. Related work

Numerous investigations have been carried out to use DL to automatically detect glaucoma. For the optic cup (OC) segmentation, a U-Net architecture based on the DL algorithm and a persistent transfer learning model were employed, according to Sudhan et al. (2022). With DenseNet-201 deep convolutional neural network (CNN), the features for glaucoma prediction are extracted. The generated model achieves accuracy in training and testing of 96.90 and 98.82%, respectively. The ORIGA dataset was used in this study for evaluation. Veena et al. (2022) introduced the segmentation method for glaucoma detection. The authors used a DL design with enhanced two CNN models for OC and optic disc (OD) independently to provide a reliable outcome using the DRISHTI-GS database. In this case, the accuracy of the current CNN model’s OD and OC segmentation was 98% and 97%, respectively. There are several uses for this approach in medical picture segmentation. Huazhu et al. (Fu et al., 2018) suggested a novel automated glaucoma screening network called the Discaware Ensemble Network (DENet) and developed a DL method to quickly diagnose glaucoma from the fundus picture and to obtain additional information from the image. This study’s primary flaw is that it underperformed when compared with the dataset SINDI. Gheisari et al. (2021) created an MLP (Multilayer Perceptron)–based ANN with GLCM (Gray Level Co-occurrence Matrix) feature extraction for glaucoma detection. It achieves 93.4% accuracy in effectively classifying the fundus images.

Many segmentation systems, such as the well-known U-Net and SegNet, are based on encoder-decoder architectures. Sevastopolsky et al. (Sevastopolsky, 2017) used vertical CDR as metrics used for OC and OD glaucoma and suggested a change to U-Net to improve CDR results, and resulted in improvement in the segmentation performance even with minimal data as a cutting-edge framework. Shankaranarayana et al. (2017) created adversarial learning-based network to identify glaucoma by using fully convolutional networks (FCNs), and resulted in segmentation performance enhancement.

Fig. 1. Healthy eye and glaucoma eye.
for the joint OD and OC segmentation. The encoder-decoder structure based on patch-based adversarial learning suggested by authors to address the domain-shifting problem. Its performance in OC segmentation was lacking (Shankaranarayana et al., 2017). Very few frameworks use non-medical images to train the network prior to official training. Cropping, contrast adaptation, and histogram equalization are just a few of the intensive preprocessing steps needed for this kind of automation in glaucoma screening (Orlando et al., 2017).

Deep learning-based methods typically use the same steps to identify glaucoma. Usually, segmentation is carried out first, and then clinical analysis is done to check for glaucoma. By employing various streams, deep learning-based algorithms have the ability to extract more pertinent information from the image for the purpose of directly detecting glaucoma. Rather than using widely used datasets like the Drishti and RIM-ONE databases, this methodology was thoroughly tested using the SCES and the SINDI datasets. By employing encoder-decoder designs to obtain more pertinent information, it requires less training time. The use of prior weight initialization allows for fine-tuning (Fu et al., 2018b). Fu et al. (2018a) modified U-Net to jointly segment the OD and OC, leading to the development of a well-known M-Net model. It is possible to feed multi-level input, which leads to improved prediction for layer scaling. With this encoder-decoder design, multi-level input may be fed, and a better prediction can be obtained. It encountered a decline in performance when segmenting low-contrast and blurry images. However, domain-shifting issues arise when the same trained network is tested using distinct datasets. The patch-based adversarial learning technique for domain modification was used to tackle domain-shifting issue. Results were achieved in this work by testing the network on other datasets after it had been trained on one. Additionally, the same dataset was used for testing and training another network. The results that are shown were also obtained by preprocessing. Nonetheless, Wang et al. (2019) have observed a void in the enhancement of OC’s segmentation ability. An additional approach is put forth that combines two distinct networks in an encoder-decoder fashion. The model’s encoder is the well-known U-net, while the model’s decoder is the residual network-34 (ResNet-34) architecture. This hybrid model produces better outcomes with much less training time. It’s an easy, two-step process. The head of the optic nerve is found first. A weighted loss is used for segmentation in the second stage. Preprocessing is applied after initializing the weight (Almubarak et al., 2020).

To lessen the workload for eye specialists, the authors of the research, Sreng et al. (2020), suggested an automated two-fold glaucoma diagnostic technique. In the first step, the DL model that was implemented, DeepLabv3+, identified in order to extract the OD from the full image. To differentiate between healthy individuals and those with glaucoma in the fragmented OD area, three deep CNN techniques have been employed in the second phase. Lastly, the authors finally concluded that a mix of DeepLabv3 and MobileNet models is the best option for OD segmentation after evaluating their algorithms on five public datasets totaling 2787 images. The accuracy of the REFUGE dataset was 95.59% and 95.75% on the ensemble (P1) and ensemble (P2), respectively, had been achieved. The REFUGE dataset yielded an AUC of 95.10% and 94.32% on the ensemble (P1) and ensemble (P2), respectively. Bian et al. (2020) presented a model for using CDR to segment the OD and OC. Here, the OD and OC segmentation results are 93% and 88%, respectively using the REFUGE fundus database. However, the primary drawback is that inaccurate information from the first step of OD segmentation might also damage the second stage.

Segmenting the OC accurately is quite difficult, and proper OD segmentation is also required. To increase computational efficiency, decrease network costs, and enhance the performance of OD and OC pixel-wise segmentation by limiting the loss of spatial information, Haider et al. (2022) proposed the separable linked segmentation network (SLS-Net) and the separable linked segmentation residual network (SLSR-Net), each of which had a separable convolutional link. The results of OD segmentation using the SLS-Net approach are as follows: accuracy of 0.998, dice coefficient of 0.962, sensitivity of 0.986, specificity of 0.999, and OC segmentation yielding dice coefficient of 0.880, sensitivity of 0.826, and accuracy of 0.99. In OC segmentation, accuracy of 0.99, dice coefficient of 0.884, sensitivity of 0.827, and specificity of 0.999 have been attained; in OD
segmentation, the SLSR-NET approach has yielded an accuracy of 0.998, dice coefficient of 0.953, sensitivity of 0.930, and specificity of 0.999.

Using both pixel-level and image-level labels during network training, Hervella et al. (2022) suggested a novel multi-task technique for the simultaneous segmentation of the OD and OC and the classification of glaucoma that increases the overall performance. The DRISHTI-GS and REFUGE public datasets were validated. The Refuge dataset yielded a classification AUROC of 97.60, and the segmentation dice values for cups and discs were 88.25, and 95.85 respectively. To lower the total processing cost of the system, Tabassum et al. (2020) presented a Cup Disc Encoder Decoder Network for the mix segmentation of the OD and OC. Additionally, they got rid of the postprocessing and pretreatment steps. Segmenting OD and OC involved modeling a semantic pixel-wise labeling challenge. Results of the model's training indicate that the proposed CDED-Net system yields state-of-the-art OC and OD segmentation results. An OD IOU of 0.8837 and an OC IOU of 0.8111 and 0.985 were obtained on the REFUGE dataset.

Septiarini et al. (2023) suggested a CNN with a single-shot multi-box detector (MobileNetV2) technique that automatically partitions the OD area in fundus pictures using both public (REFUGE) and private datasets containing retinal fundus images. After that, several preprocessing actions were implemented and a U-Net model with 128 \times 128 input data was used to segment the OD. The intersection over union (IOU) of 97.12, the dice score of 98.38, and the FI-score of 98.54 for the REFUGE dataset have all been attained. To preserve important textures and make it easier to recognize the regions of interest, Wang et al. (2021) provided a unique sub-network for a DL network that automatically identified the OD regions using conventional U-Net architecture to increase the consistency and accuracy of OD area segmentation on fundus pictures with color. Three publicly accessible datasets named as MESSIDOR, ORIGA, and REFUGE were used for training the developed network, average values of 93.42 for the Dice similarity coefficient (DSC), 87.96 for the intersection over union (IOU), and 93.7 for the Matthew’s correlation coefficient (MCC) were obtained. For multiclass multilabel classification on the ODIR-19 dataset, Bali et al. (Bali and Mansotra, 2021c) proposed a one versus rest classifier and achieved 91.34% validation accuracy on glaucoma diagnosis. Bali et al. (Bali and Mansotra, 2023b) proposed DDFex-hybrid approach of BeeHive model, CGAN, and PSO to categorize eye diseases and obtained 98.79% accuracy, 95.99% sensitivity, 99.79% specificity, and 97.16% accuracy, 96.81% F1 score in the RFMiD and ODIR dataset, respectively.

3. Methodology

The following Fig. 1 shown below is the suggested procedure.

3.1. Dataset used

REFUGE is the dataset that is used for experimentation in this work collected from IEEE data port. This collection contains 1200 images, 1080 of which are of healthy people and 120 of which are of glaucoma patients (Fu et al., 2019). The glaucoma dataset is a collection of clinical data and anonymized eye measurements from glaucoma patients and healthy people. The findings of visual field tests, measures of the optic nerve, and demographic information are frequently included, assisting in the creation of management and diagnostic plans.

3.2. Implementation details

To train models, a GPU with 8 GB of RAM is utilized. Being among the first works to try picture segmentation using U-Net, the research of (Ronneberger et al., 2015) both made this attempt. The topic of these publications is how to segment retinal lesions in glaucoma pictures using the U-Net design. Since then, the U-Net design has been extensively utilized in numerous medical picture segmentation applications, including the detection of glaucoma. Originally developed for generic biomedical image segmentation tasks, the U-Net architecture is currently widely utilized in numerous different medical imaging domains, such as Glaucoma.

The workflow steps make up the procedure depicted in Fig. 2. The Glaucoma dataset was gathered via IEEE Data port. The training, testing, and validation sets of the dataset are split up in the following ratios: 70:20:10. Following data preprocessing, all of the photos have been scaled to 128*128. After that, data normalization was carried out to increase model accuracy by assigning each variable an equal weight, preventing any one variable from influencing model performance in a particular way just because it contains huge numbers. Since our algorithm can only process numeric data, we have performed a single hot encoding after normalization to transform the categorical data into numerical data. This allows us to apply our approach without encountering any difficulties.
Since the region of interest differs depending on the application, no one algorithm can be used to segment all the photos. Therefore, I used a variety of hybrid segmentation techniques, including U-Net res18, U-Net res34, U-Net inception, and so forth. A new design called inception net is a mix of two cutting-edge DL algorithm architectures: U-Net architecture and inception architecture. The inception layers within the U-Net take the place of the default convolution layers. The processing of biomedical pictures, which have evolved from traditional CNN, is the purpose of U-Net. CNN focuses on image classification tasks where image is the input, and a label is the result. However, in biological circumstances, we need to identify the area of abnormality in addition to determining whether a disease exists; hence, U-Net is used to classify each pixel in the image.

Selecting the appropriate kernel size for the convolution process becomes difficult because of the significant variation in the information's position. As a result, the inception module uses a variety of filter sizes, with smaller filters covering little information and larger filters covering more information. Following that, the model is assessed and verified using the validation dataset. For example, if the training accuracy is 10% higher than the validation, overfitting is present; if it is 10% lower than the validation accuracy, underfitting is present. Hyperparameter tuning, in which different parameters are automatically altered, has been done in both cases. When there is no longer any overfitting or underfitting, the iterative mode is continued, the model is saved, and 10-fold cross-validation is conducted on an 80:20 ratio.

Cross-validation is a DL technique that evaluates the model's output on data that is not seen. The testing set is one of the folds or subsets of the data, and the training set is the remaining. The last stage is averaging the results from every testing phase to generate a more trustworthy assessment of the model's output. Cross-validation's primary goal is to avoid overfitting, which happens when a model performs poorly on unknown data despite having been trained excessively well on training data. Subsequently, other performance metrics were computed, including the IOU score, dice score, sensitivity, recall, and f1 score.

4. Results and discussion
The technique obtains relatively minimal loss rates over any given set of data. The algorithm's average train loss of 0.115092 indicates that it successfully minimizes its prediction errors throughout training. The test and validation losses of 0.193428 and 0.208596, respectively, as shown in Table 1 and Fig. 3, seem to suggest that the algorithm performs a reasonable job of minimizing the disparities in the expected and actual ophthalmic divisions, even though they are slightly bigger. Fig. 4 shows that the accuracy scores for the train and test sets are both rather high at 0.999567 and 0.998914, respectively. This indicates that the system
achieves a high degree of overall precision in classifying pixels as either ophthalmology or non-glaucoma regions. The model's ability to accurately classify pixels is further evidenced by the binary accuracy scores, which have test and train values of 0.998914 and 0.999567, respectively.

However, a noticeable difference between the test and training sets of data is evident in the AUC scores. The train AUC of 0.988646 indicates that the system performs well in distinguishing between places with ophthalmology and those without glaucoma in the initial training set. However, the

Table 1. Data split analysis glaucoma.

<table>
<thead>
<tr>
<th>Split</th>
<th>Loss</th>
<th>Accuracy</th>
<th>Binary accuracy</th>
<th>Auc</th>
<th>Specificity</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.115092</td>
<td>0.999567</td>
<td>0.999567</td>
<td>0.988646</td>
<td>0.999928</td>
<td>0.957831</td>
</tr>
<tr>
<td>Test</td>
<td>0.193428</td>
<td>0.998914</td>
<td>0.998914</td>
<td>0.935631</td>
<td>0.999698</td>
<td>0.966844</td>
</tr>
<tr>
<td>Valid</td>
<td>0.208596</td>
<td>0.998545</td>
<td>0.998545</td>
<td>0.922519</td>
<td>0.999702</td>
<td>0.9755411</td>
</tr>
</tbody>
</table>

(a) Loss Training Validation Plot

(b) Accuracy Training validation Plot

Fig. 3. Glaucoma training analysis.

DENSE-ED-UHI: ENCODER DECODER BASED UNET HYBRID INCEPTION COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
<th>Valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>accuracy</td>
<td>0.999567</td>
<td>0.998914</td>
<td>0.998545</td>
</tr>
<tr>
<td>binary accuracy</td>
<td>0.999567</td>
<td>0.998914</td>
<td>0.998545</td>
</tr>
<tr>
<td>auc</td>
<td>0.988646</td>
<td>0.935631</td>
<td>0.922519</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.999928</td>
<td>0.999698</td>
<td>0.999702</td>
</tr>
</tbody>
</table>

Fig. 4. UHI metric analysis glaucoma.
analysis's AUC of 0.935631 indicates that the potency of biases in unobserved data is lower. This suggests that the framework might have some issues applying its findings to people with glaucoma who are unknown and new.

In general, the method exhibits exceptional accuracy and dichotomous correctness in classifying glaucoma regions. However, the discrepancy in the AUC ratings raises doubts regarding the model's capacity to successfully apply to new glaucoma cases. Further research and evaluation are necessary to determine the concept's practicality and robustness.

The framework achieves a median loss of 0.139275, indicating that the algorithm successfully reduces its predicting errors on a normal basis. As can be seen from the lowest and maximum loss estimates of 0.149321 and 0.128920, respectively, there is not much fluctuation in loss rates. This demonstrates that the simulation consistently does an excellent job of minimizing the discrepancies between the expected and actual glaucoma classifications.

Reliability ratings demonstrate the model's ability to correctly classify photons as belonging to the ophthalmology or non-glaucoma sectors. The model's exceptionally high average precision and minimum accuracy ratings (0.999345) demonstrate its ability to classify data successfully. In Figs. 5 and 6, the prediction is shown for inception mask and inception overlap and compared with Resnet.

These indicators have narrow ranges, and their small standard deviations suggest that the algorithm's classification skill is stable. The AUC values aid the model's ability to distinguish between regions with and without conjunctivitis, as shown in Table 2. The statistical model does a decent job at distinguishing between these sites, with a mean AUC of 0.972007. The algorithm's 0.968623 25th percentile score and 0.974657 75th percentile number demonstrate a consistently high level of biased capacity across various datasets.

![Fig. 5. Mask prediction and BBox using Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception (Proposed Model).](image-url)
The sensitivity and specificity metrics, in turn, assess how well the model performs in accurately recognizing both positive and negative categories. The average specificity of 0.999849 shows that the technique is successful in identifying non-glaucoma areas, while the average sensitivity of 0.9372308 shows that it can properly identify glaucoma locations.

The statistical analysis highlights the Glaucoma Localization model's effective operation when utilizing the Dense-ED-UHI: Encoder Decoder-based U-Net Hybrid Inception structure, as shown in Fig. 7. The model achieves low loss values and good scores for specificity, AUC, dichotomous reliability, and sensitivity. These findings suggest that the model does a good job of differentiating between glaucomatous and non-glaucomatous areas. To assess the durability and generalization of the framework in practical situations, more testing on a wider range of datasets will be

<table>
<thead>
<tr>
<th>Table 2. Data statistical analysis glaucoma.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>25%</td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>75%</td>
</tr>
<tr>
<td>Max</td>
</tr>
</tbody>
</table>

Fig. 6. Comparison with other state of the art backbone (Resent50).
necessary. It is critical to remember that those conclusions depend on the particular data set under consideration.

4.1. Conclusion

The novel Glaucoma classification methods conceptually segment using Inception and apply a revolutionary up-sampling methodology that uses pixel-wise periodic shuffling convolution, together with a tailored UHI or U-Net model. With a median accuracy of 99.9%, the mathematical models are incredibly accurate and absolutely right when it comes to glaucoma. The models’ 97.2% AUC values show how well they can distinguish between favorable and unfavorable cases. The 93.7% sensitivity values show how well the models are able to detect true positives. The models’ high specificity values of 0.999 show that they accurately categorize unfavorable instances. In summary, segmenting retinal sickness can be greatly aided by the application of DL models, particularly the UHI system, and the possible integration of the categorization network for broad category categorization. These simulators have proven to be incredibly robust, dependable, and capable of automating and
assisting in the timely diagnosis and identification of retinal issues. Better healthcare could result from this discipline’s research and development.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors contribution

AB and VM created the idea for the study. AB led the writing of this paper. AB analyzed and interpreted the data. AB contributed to the implementation of the paper as well as to the identification of theoretical problems. AB contributed to the writing of the paper as well as participated in revising this manuscript. VM encouraged AB to investigate and supervise the findings of this work. All authors contributed significantly to the writing of the paper, discussed the results, reviewed the final draft, and approved the final version of this manuscript.

Conflicts of interest

There are no conflicts of interest.

References