

# Enhancing Choroidal Nevi Segmentation in Fundus Images Using YOLO

Mehregan Biglarbeiki University of Calgary Calgary, Canada mehregan.biglarbeiki@ucalgary.ca	Roberto Souza University of Calgary Calgary, Canada roberto.souza2@ucalgary.ca	Emad Mohammed Wilfrid Laurier University Waterloo, Canada emohammed@wlu.ca	Ezekiel Weis University of Alberta Edmonton, Canada ezeziel@ualberta.ca
Carol L. Shields Wills Eye Hospital Philadelphia, USA carolshields@gmail.com	Sandor R. Ferenczy Wills Eye Hospital Philadelphia, USA sandor@shields.md	Behrouz Far University of Calgary Calgary, Canada far@ucalgary.ca	Trafford Crump University of Calgary Calgary, Canada tcrump@ucalgary.ca

**Abstract**—Choroidal nevus often appears as a darkly pigmented, benign ocular lesion, which may progress to malignant forms, such as choroidal melanoma. Prompt and precise diagnosis of choroidal melanoma cannot be overstated as untreated cases can lead to vision loss and even life-threatening metastasis, underscoring the importance of regular screening of the eye. However, this procedure is performed manually, which can be time-consuming and prone to human errors. Recent advancements in deep learning show potential for detecting eye diseases, including choroidal nevi. However, these models require extensive labelled data, which can be difficult to acquire due to the associated labelling costs. In this paper, we utilize two approaches to tackle these challenges. Firstly, we leverage a pre-trained YOLOv8 segmentation model and train it on both patches and full-size high-resolution colour fundus images. This strategy effectively expands the dataset size and allows the model to focus on the fine details of lesions within individual patches while understanding the general shape of the lesions by analyzing the entire image. Secondly, we use data augmentation to further expand the dataset size and tackle the class imbalance problem. Additionally, through the utilization of post-processing techniques, we enhance the predicted masks by addressing any potential flaws. This approach resulted in a 0.833 Dice Coefficient Score and a 0.714 Intersection Over Union (IOU) in our initial dataset and a Dice score of 0.764 and IOU of 0.618 on our second test set collected from a different site. In both datasets, this approach surpassed the models trained exclusively on full-size images.

**Index Terms**—Choroidal Nevus, Deep Learning, YOLOv8, Segmentation

## I. INTRODUCTION

Choroidal nevi are benign melanocytic lesions located in the back of the eye that can transform into uveal melanoma, a potentially fatal type of eye cancer [1]. Therefore, patients with choroidal nevi require regular screening. However, with a prevalence range from 5% to 25% in the adult population, there are significantly more patients with choroidal nevi than the capacity of eye care providers to monitor [1].

Eye care providers often use high-resolution fundus images to identify and monitor choroidal nevi [2]. While this manual process is effective, it is labour-intensive and time-consuming [1]. In recent years, deep learning techniques have significantly advanced medical image analysis, specifically in the field of

segmentation. Therefore, applying these techniques to fundus images could provide a solution for automating the screening and monitoring of choroidal nevi.

Prior research has showcased the efficacy of deep learning in detecting retinal diseases from fundus images [3]. However, due to the limited availability of labelled data, there are only a few studies on the automatic detection of choroidal nevus in fundus images [4], [5]. This scarcity poses a challenge for deep learning models, which rely on substantial data for training to achieve optimal performance [6].

Researchers have explored various approaches, including transfer learning and data augmentation to tackle this challenge by enabling models to work effectively with limited labelled data [7]. Transfer learning leverages features learned from one task to enhance generalization in another. By initially training a model on a large dataset and then fine-tuning it on a different but related problem, data requirements can be significantly reduced. This approach also requires fewer computational resources, as training from scratch is unnecessary [8]. However, while the COCO dataset [9] is widely used for pre-training in segmentation tasks, its domain gap with medical datasets limits the performance of models pre-trained on COCO in medical applications. Data augmentation techniques increase the dataset size and allow the model to generalize better to unseen data [10]. However, applying colour and geometric transformations requires careful consideration, as they may not always preserve data labels [10]. Therefore, to further overcome the constraints of training a deep learning model on limited data, a proposed approach is to train the model on data patches to increase the dataset size [11]. Additionally, dividing images into smaller patches enables the model to focus on desired areas with enhanced detail [12].

This study investigates training a YOLOv8 [13] model on both patches and full-size colour fundus images to segment choroidal nevi. Additionally, it assesses the impact of different augmentation techniques, including geometric and colour transformations and random cropping, particularly on minority class (i.e., lesion pixels), while maintaining data labels, to ad-

dress class imbalances. Furthermore, morphological operations are explored to refine the predicted segmentation masks. The approach is also tested on a second dataset from a different site to determine its generalizability over different datasets. These findings could overcome the current limitations of deep learning in automating fundus image analysis for choroidal nevi segmentation and facilitate eye screening protocols.

The rest of the article is structured as follows: Section II provides a review of related literature. Section III outlines the dataset, the model architecture, and the preprocessing and post-processing techniques used. Section IV showcases the results obtained from our approach. Section V presents the key findings and discusses challenges, potential solutions, and future directions. The paper concludes in Section VI.

## II. LITERATURE REVIEW

In recent years, significant progress has been made in utilizing deep learning models to segment and detect ocular diseases from fundus images. However, to our knowledge, no prior work has focused explicitly on segmenting choroidal nevi lesions from fundus images. Although some efforts have been made to classify these lesions in fundus images as identified in [3] and [14], the field remains relatively unexplored due to the limited availability of public data, presenting challenges for deep learning models in achieving reliable results. In the following section, we delve into previous works utilizing deep learning models for classifying, detecting, or segmenting medical images, particularly in diagnosing ocular diseases.

In [12], the authors utilized a binary classification model on a small private dataset of choroidal nevi and healthy cases to study the impact of training the model with patched data. They achieved an accuracy of 92.61%, which showed a 7.4% improvement compared to the model trained solely on full-size images. In [5], a dataset with 798 ultra-widefield retinal images of three classes (control, choroidal nevi, and uveal melanoma) was used. They employed a pre-trained DenseNet121 model to prevent overfitting with limited data and tested the effect of colour fusion techniques. The intermediate fusion method achieved an accuracy of 89.72%, with the model trained only on the red channel achieving an accuracy of 83.79%. Lam et al. [15] compared the performance of five state-of-the-art models for binary classification of normal and abnormal retinal images using a subset of 243 images of the Kaggle Diabetic Retinopathy dataset. To expand the dataset, they employed two techniques: sliding window extraction, where each window served as input, and manual patching of images with lesions. Among these models, Inception-v3 performed the best, achieving an accuracy of 98% on the Ophtha dataset.

Moving forward, Wang et al. [16] proposed a segmentation model using Nest-U-Net and a specialized patch extraction strategy to help the model focus on vessel details. They tested this approach on publicly available datasets DRIVE and STARE, achieving an accuracy of 95.12% on the DRIVE dataset. Skouta et al. [17] introduced a modified Convolutional Neural Network (CNN) U-Net for segmenting retinal hemorrhages in fundus images from two public datasets. They

initially extracted the green channel from RGB images and applied various geometric transformations due to limited data, achieving a Dice coefficient score of 0.865. Singh et al. [18] utilized five deep learning-based networks to segment the retinal blood vessel tree by randomly patching the data obtained from the STARE dataset. The LadderNet model yielded the best result, achieving an accuracy of 97.1%.

The Literature Review section in this paper offered a thorough overview of prior research in classifying and segmenting medical images, focusing on deep learning applications in ocular diseases. Through this review, some gaps have become evident: notably the absence of prior research dedicated to segmenting choroidal nevi lesions in fundus images. Moreover, most prior studies exclusively tested their methods on datasets used for initial model training, without assessing generalization to other datasets. This lack of evaluation leaves it uncertain whether their approach can be generalized effectively. Furthermore, previous studies have primarily focused on detecting retinal vessels, retinal hemorrhage, and diabetic retinopathy, which are characterized by colours resembling red tones. Using the green channel of RGB images for segmenting these regions has proven effective, as it enhances these red regions. However, this approach is not applicable to our context, as our lesions are distinguished by dark-pigmented gray/green tones. These identified gaps highlight the unique challenges in choroidal nevi lesion segmentation, emphasizing the need for approaches specifically adapted to these lesion characteristics.

## III. MATERIALS AND METHODS

In this study, we conducted an initial experiment to compare the performance of a YOLOv8 segmentation model with a U-Net [19] model, both trained on image patches. Subsequently, we conducted two additional experiments to explore the effects of training the YOLOv8 model on both patches and full-size images. Finally, we analyzed the effects of applying our proposed post-processing techniques on the predicted masks. The complete methodology, including preprocessing and post-processing techniques, will be detailed further in this section.

### A. Dataset

In this study, we utilized a dataset of 574 3916×3918 pixels colour fundus images captured by a Zeiss Clarus 500 camera, designed for ultra-widefield fundus imaging. The dataset was obtained from the Alberta Ocular Brachytherapy Program, a referral centre for uveal melanoma in Edmonton, Canada. An ocular oncologist labelled 237 images as "Lesion", accurately tracing the lesions within them, with the remaining images marked as "Control." This dataset served as the foundation for training, validation, and testing our model. Furthermore, we utilized 124 fundus images with varying dimensions, ranging from 3000×3000 to 4000×4000 pixels, from the Wills Oncology dataset, collected from Wills Eye Hospital in Philadelphia, USA. Among these images, 53 were annotated as "Lesion", while the remaining images were categorized as "Control." These images were captured using the Zeiss Clarus

camera with different settings than the previous dataset and were exclusively used to validate our model's performance.

### B. YOLOv8

A YOLOv8n-seg model, with approximately 3.2M parameters, underwent initial pre-training on the COCO dataset before fine-tuning for the specific task of choroidal nevi segmentation. This model which is an extension of the YOLOv8 object detection model [13], is designed to segment objects in images by predicting bounding boxes and associated class probabilities using a single CNN [20]. The YOLOv8 model has three main components: backbone, neck, and head. The model uses CSP-Darknet53 as the backbone and it consists of a 3x3 convolution, and a C2f module designed with skip connections for improved performance. The backbone also includes a spatial pyramid pooling fusing (SPPF) module [13]. The neck module contains the path aggregation network (PANet) and feature pyramid network (FPN), enabling the detection of objects of various sizes within an image. The head of YOLOv8 consists of multiple convolutional layers followed by a series of fully connected layers [21]. The YOLOv8 segmentation model uses a weighted summation of five different losses to leverage the strengths of each loss type to improve overall model robustness and segmentation performance.

$$\text{Total\_Loss} = w_1 \cdot \text{Box\_loss} + w_2 \cdot \text{DF\_loss} + w_3 \cdot \text{CIoU\_loss} + w_4 \cdot \text{Seg\_loss} + w_5 \cdot \text{BCE\_loss} \quad (1)$$

Equation 1 [22], [13], comprises several components: the Box\_loss, which evaluates the deviation of predicted bounding box coordinates and dimensions from ground truth; the Distribution Focal (DF) loss addresses class imbalance; the Complete Intersection over Union (CIoU) loss enhances bounding box accuracy; the Seg\_loss quantifies the dissimilarity between predicted segmentation masks and their corresponding ground truth masks; and the Binary Cross-Entropy (BCE) accounts for classification loss [22].

For all experiments, the model underwent training for 60 epochs with a batch size of 16 for YOLOv8 and 4 for U-Net, with early stopping implemented to prevent overfitting. The initial learning rate was set to 0.001, and we utilized AdamW as the optimizer. Additionally, we monitored the loss curves throughout training to avoid overfitting.

### C. Preprocessing

In all experiments, the Alberta Ocular Brachytherapy dataset was split into three subsets: 50% for training (274 images), 20% for validation (120 images), and 30% for testing (180 images). The data splitting was stratified to maintain a balanced representation of healthy and lesion cases in each subset.

1) *Experiment 1*: In the initial experiment, both U-Net and YOLOv8 models were used for training. Initially, all images within the three subsets and their corresponding masks were resized from their original dimensions to 3000×3000 pixels for YOLOv8 training, and to 512×512 pixels for U-Net training. These dimensions were selected for the U-Net model due to hardware constraints, as it is more memory-intensive compared to the YOLOv8 model. Subsequently,

each image in the training and validation subsets, along with their respective binary masks, was uniformly divided into a grid of 4 patches. For YOLOv8 training, each patch was sized at 1500×1500 pixels, while for U-Net training, each patch was sized at 256×256 pixels. For YOLOv8, various patch sizes were tested: 300×300 pixels, 600×600 pixels, 1000×1000 pixels, 1500×1500 pixels, and 1900×1900 pixels. While smaller patch sizes resulted in more samples, the model struggled with correctly segmenting the patches that consisted entirely of lesions without any background. Larger patch sizes (1000, 1500, and 1900 pixels) solved this issue. Therefore, we selected a patch size of 1500×1500 pixels, making a trade-off between predicted mask quality, training speed, and memory efficiency, with a stronger consideration for mask quality. This yielded 1096 training images and 480 validation images. For testing, a sliding window technique divided the images into 9 patches with a 50% overlap, resulting in 1620 images.

Since initial labels were assigned to the full-size images, a two-step procedure was used to reassign labels to patches and correct their masks accordingly. First, we implemented a categorization approach based on lesion size. Four different lesion size groups were defined: Large Lesions (>10% of the image size), Medium Lesions (between 1% and 10% of the image size), Small Lesions (between 0.3% and 1% of the image size), and Minute Lesions (<0.3%) Subsequently, "Lesion" labelled patches were identified by comparing the ratio of white pixels, which represent lesions in the binary masks, in each patch to those in its corresponding full-size mask. If this ratio fell below an experimentally determined threshold, the masks were replaced with empty masks, indicating the patch as a "Control" patch. While the patching process expanded the dataset and improved lesion detail recognition, it introduced a mild class imbalance issue, resulting in fewer "Lesion" patches than "Control" patches. Different data augmentations addressed the class imbalance in the underrepresented "Lesion" class and enhanced the model's capability to segment a broader range of lesion scenarios. These augmentations involved randomly cropping full-size images initially labelled as "Lesion" twice, ensuring that a part of the lesion would be included in each cropped patch. Additionally, colour and geometric transformations such as 90-degree rotations, flips (both horizontal and vertical), and adjustments in brightness and hue were applied to all patches identified as "Lesion." Furthermore, image translation, saturation adjustment, and mosaic augmentation (up to the final ten epochs) were randomly applied during training. Following all the preprocessing steps, a training set of 3117 patches and a validation set of 1448 patches was created.

2) *Experiment 2*: This experiment was exclusively performed using the YOLOv8 model. All images and their corresponding masks within the three subsets were resized from their original dimensions to 1500×1500 pixels for fine-tuning. The model also underwent the same data augmentations that were applied during the training of the model in the previous experiment.

3) *Experiment 3*: In the third experiment, a YOLOv8 model was trained using both patches and full-size images. This in-



volved combining the patches processed in the first experiment with the full-size images from the second experiment. The same augmentations employed during training the model in the first and second experiments were applied.

#### D. Post-processing

The first post-processing step applied to the predicted masks, was specific to models in experiments 1 and 3, which were trained on patches. After testing the models on patches generated from the full-size images with a 50% overlap, we applied an automatic method to identify each set of 9 patches that belong to each full-size image using their file names. Following this, we reassembled the patches using the OR operation on overlapped areas to form a full-size mask. Using the OR operation for patch reassembly decreased false negatives, as it accounted for the scenario where if a lesion is missed in one patch, it might be segmented in the neighboring patch that shares an overlap.

In Experiment 3, the model was trained on both patches and full-size images. To evaluate its performance, we initially tested the model using image patches. The reassembly process to create a full-size mask from these predicted patches followed the same method as in Experiment 1. Additionally, we tested the model on full-size images, resulting in another set of full-size masks. Each image then had two masks: one from the reassembled predicted patches and another from testing on the full-size image. These were then combined to form a single predicted mask for each image. Upon examination, we observed that if a patch within the reassembled mask is mistakenly predicted as empty, combining the reassembled mask with its corresponding full-size mask using the Average operation results in a smaller predicted mask compared to the ground truth. To address this, we developed a method to detect the presence of horizontal or vertical edges that indicate areas where a patch was incorrectly predicted as empty within the reassembled masks. We applied the Canny edge detection algorithm and HoughLine Transform to detect the boundaries of the mask and identify any vertical or horizontal edges within these boundaries. If such edges were detected, we utilized the OR operation to combine the masks; otherwise, we employed the Average Operation to obtain the best version.

The second post-processing step, common to all experiments, aimed to refine the predicted mask and improve the overall quality of the segmentation results. Morphological operations, including Connected Component Analysis (CCA), closing, and opening operations, were employed for this purpose. In the first stage, we applied CCA to designate the first pixel as the initial component and then searched for neighboring pixels to include them in this component. A new component was established if no further connected pixels were found but unassigned pixels remained. We chose to only retain the largest component, which likely represented our region of interest [23]. Moving to the second stage, we used closing, consisting of a dilation followed by an erosion, effectively filling small gaps in the segmented parts while preserving their shape and size [24]. Finally, in the third stage, we applied

the opening operation, where the mask underwent erosion followed by dilation. This helped in smoothing the contours and removing noise or thin protrusions, as well as eliminating sharp peaks or capes present in the segmented masks [25].

#### E. Experimental Setup

This study utilized the Dice coefficient and IOU, also known as the Jaccard index, which are widely used metrics in the context of segmentation. These metrics were used to evaluate the performance of segmentation models by measuring the overlap between the predicted mask and the ground truth. Higher Dice and IOU scores indicate a closer alignment between the predicted mask and the ground truth mask [26].

### IV. RESULTS

In this section, we initially provide a comparative analysis of the results achieved by both YOLOv8 and U-Net models, each trained on patches of data. Following this, we present the performance results of the YOLOv8 model across two additional experiments, where it was tested on both the Alberta Ocular Brachytherapy and the Wills Oncology datasets. Additionally, we investigate the impact of various post-processing techniques on these results.

#### A. Experiment 1 - YOLOv8 vs U-Net

Both YOLOv8 and U-Net models were evaluated on patches generated using a sliding window technique with a 50% overlap on the test set images. The results in Table I show that the YOLOv8 model achieved a Dice score of 0.735, outperforming the U-Net model with a Dice score of 0.652 on the Alberta Ocular Brachytherapy dataset. Consequently, we proceeded with the YOLOv8 model for the remaining experiments due to its superior performance compared to the U-Net model. Additionally, the YOLOv8 model was also tested on the Wills Oncology dataset. In this experiment, the model primarily encountered challenges in segmenting large lesions. Figure 1 illustrates the predicted patches and the subsequent process of reassembling and refining the full-size mask for choroidal nevi segmentation.

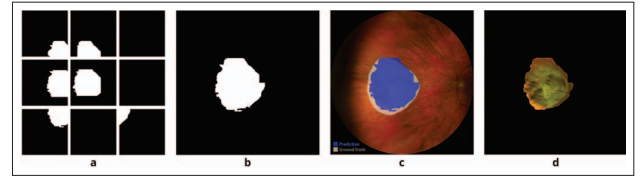


Fig. 1. (a) Predicted patches with 50% overlap in original order and position. (b) Patches reassembled into the original mask using the OR operation. (c) Comparison of the full-size mask (blue) with the ground truth mask (white). (d) Extracted lesion after mask refinement through morphological operations.

#### B. Experiment 2

The YOLOv8 model was trained using full-size images and subsequently tested on full-size images from both datasets. The results, including Dice and IOU scores, are summarized in Table I. The model in this experiment faced challenges in accurately segmenting smaller lesions and those with brighter edges. Figure 2 displays the predicted mask

alongside the ground truth mask, with the segmented lesion post-morphological operations.

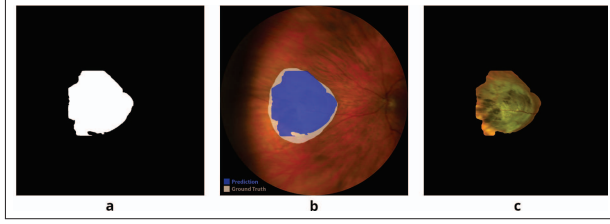


Fig. 2. (a) Predicted mask. (b) Comparison of the predicted mask (blue) with the ground truth mask (white). (c) Extracted lesion after mask refinement through morphological operations.

### C. Experiment 3

In response to the challenges posed by previous experiments, the YOLOv8 model in this experiment was trained and evaluated on full-size images and patches. Table I presents the Dice and IOU scores for this model evaluated on both datasets. The final mask was generated using the Average operation to combine the reassembled mask created from testing the model on patches, and the full-size mask from testing the model on full-size images. Figure 3 illustrates the process of reassembling patches and combining the reassembled mask with the full-size mask using either OR or Average operations, resulting in the segmented lesion post-morphological operations. Furthermore, table II demonstrates the impact of the horizontal and vertical edge detection technique on selecting between the OR or Average operation for combining masks, resulting in higher Dice and IOU scores. Additionally, while morphological operations slightly increased the Dice and IOU scores, they also visually enhanced the predicted masks.

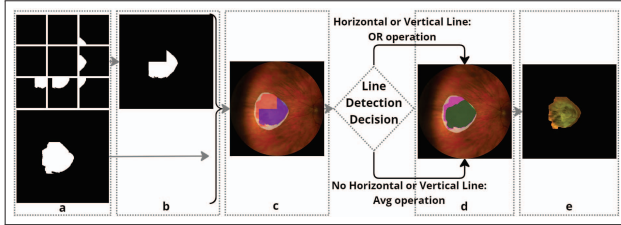


Fig. 3. (a) Top: Predicted patches with 50% overlap from testing on patches. Bottom: Predicted full-size mask from testing on full-size image. (b) Patches reassembled into a full-size mask using the OR operation. (c) Predicted reassembled mask (blue), full-size mask (red), and ground truth mask (white). (d) Average of reassembled and full-size masks (green), OR (purple). (e) Extracted lesion using OR mask post-morphological operations.

## V. DISCUSSION

Early detection of choroidal nevi is crucial, giving the potential to transform into malignant melanoma. However, differentiating between choroidal nevi and melanoma, especially smaller choroidal melanomas, poses a challenge due to overlapping features such as size, colour, and location [27].

This highlights the importance of precise and regular eye screening. Automated choroidal nevi segmentation using deep learning models offers an efficient alternative to manual review

TABLE I  
PERFORMANCE COMPARISON OF MODELS ACROSS THREE EXPERIMENTS.  
\*THE U-NET MODEL WAS NOT EVALUATED ON THE WILLS ONCOLOGY DATASET. THE BEST METRICS ARE SHOWN IN BOLD.

Test Datasets	Metrics	#1 U-Net	#1 YOLO	#2	#3
Alberta Ocular	Dice	0.652	0.735	0.772	<b>0.814</b>
	IOU	0.484	0.581	0.629	<b>0.686</b>
Wills Oncology	Dice	*	0.696	0.727	<b>0.732</b>
	IOU	*	0.534	0.571	<b>0.577</b>

TABLE II  
IMPACT OF DIFFERENT POST-PROCESSING ON DICE AND IOU SCORES IN EXPERIMENT 3 (PATCH AND FULL-SIZE IMAGE COMBINATION).

Test Datasets	Metrics	OR	Avg	Avg or OR	Morph ops
Alberta Ocular	Dice	0.801	0.814	<b>0.827</b>	<b>0.833</b>
	IOU	0.668	0.686	<b>0.705</b>	<b>0.714</b>
Wills Oncology	Dice	0.746	0.732	<b>0.755</b>	<b>0.764</b>
	IOU	0.595	0.577	<b>0.606</b>	<b>0.618</b>

by facilitating lesion detection, localization, and the creation of masks outlining lesions' boundaries.

Initially, we utilized a U-Net model, a commonly employed baseline for medical image segmentation, for comparison with the YOLOv8 segmentation model. However, based on the results presented in Table I, we transitioned to a pre-trained YOLOv8 model due to its superior segmentation performance, faster training speed, and memory efficiency. Additionally, training the model with larger patch sizes enhanced segmentation quality, further supporting our choice of YOLOv8 due to its compatibility with larger input sizes. According to Table I, the model trained on both full-size images and patches outperformed those trained only on full-size images or patches. This strategy increased the number of training samples and improved the model's understanding of overall lesion shapes from full-size images while enabling a finer focus on lesion details from patches, resulting in more accurate boundary segmentation.

Furthermore, transfer learning prevented overfitting, while data augmentations enhanced the model's generalizability to unseen data by expanding the dataset and addressing class imbalances. The sliding window method with a 50% overlap for patching testing images in Experiments 1 and 3 ensured smooth mask reassembling, eliminating visible gaps and capturing missed lesions in overlapping areas.

Experiment 3 utilized line detection techniques to identify vertical and horizontal edges in the reassembled mask. Selecting between the OR or Average operation for combining the reassembled and full-size masks yielded a mask closely resembling the ground truth. Finally, applying morphological operations improved the visual quality of the masks and improved the Dice and IOU scores, indicating a closer match between the predicted and ground truth masks. Despite overall performance improvements, the model struggles with accurately segmenting low-contrast lesions and those with a surrounding yellow halo. False positives occur in two scenarios: drusen (small yellow dots of lipids, proteins, and cellular debris) appearing without lesions, often overlaying them and causing confusion, and shadow-like abnormalities around the optic disc mistakenly identified as choroidal nevi

lesions during segmentation.

All proposed approaches enhanced choroidal nevi segmentation in both the Alberta Ocular Brachytherapy and Wills Oncology datasets. Despite a slight performance decrease on the Wills Oncology dataset, likely due to dissimilar lesions such as low-pigmented bright ones not present in our initial training set, these approaches still resulted in fewer false positives and false negatives.

Future work could involve testing the proposed approach on diverse ocular disease datasets to assess its generalizability. Another direction could be using datasets like Kaggle EyePacs, which include 30,000 images of diabetic retinopathy [28], for pre-training before fine-tuning on our dataset to potentially improve model performance. Ongoing research could also evaluate the effectiveness of the recently developed YOLOv9 model using this approach and compare results with YOLOv8.

## VI. CONCLUSION

This study centered on automating choroidal nevi segmentation from fundus images, a relatively underexplored area compared to other ocular diseases. Our study demonstrated that training a YOLOv8 model on both patches and full-size images, along with data augmentation and our proposed post-processing techniques, outperformed alternative methods when dealing with a limited dataset. Specifically, we achieved a Dice score of 0.833 and an IOU of 0.714 on the Alberta Ocular Brachytherapy dataset and a Dice score of 0.764 and an IOU of 0.618 on the Wills Oncology dataset.

## ACKNOWLEDGMENT

This research was funded by a grant from the New Frontiers Research Fund - Explorations grant (NFRFE/00227-2021).

## REFERENCES

- [1] C. L. Shields, S. E. Lally, L. A. Dalvin, M. S. Sagoo, M. Pellegrini, S. Kaliki, A. K. Gündüz, M. Furuta, P. Mruthyunjaya, A. T. Fung, J. S. Duker, S. M. Selig, A. Yaghy, S. R. Ferenczy, M. B. Eydelman, and M. S. Blumenkranz, "White Paper on Ophthalmic Imaging for Choroidal Nevus Identification and Transformation into Melanoma," *Translational Vision Science & Technology*, vol. 10, pp. 24–24, 02 2021.
- [2] M. Solnik, N. Paduszynska, A. M. Czarnecka, K. J. Synoradzki, Y. A. Yousef, T. Choragiewicz, R. Rejdak, M. D. Toro, S. Zweifel, K. Dyndor, et al., "Imaging of uveal melanoma—current standard and methods in development," *Cancers*, vol. 14, no. 13, p. 3147, 2022.
- [3] S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, and V. Lakshminarayanan, "Ophthalmic diagnosis using deep learning with fundus images—a critical review," *Artificial intelligence in medicine*, vol. 102, p. 101758, 2020.
- [4] A. K. Dadzie, S. P. Iddir, M. Abtahi, B. Ebrahimi, D. Le, T. Son, M. J. Heiferman, and X. Yao, "Deep learning for automated diagnosis of uveal melanoma," in *Ophthalmic Technologies XXXIV* (D. X. Hammer, K. Bizheva, and G. Schuele, eds.), vol. 12824, p. 128240B, International Society for Optics and Photonics, SPIE, 2024.
- [5] X. Yao, A. Dadzie, S. Iddir, M. Abtahi, B. Ebrahimi, D. Le, S. Ganesh, T. Son, and M. Heiferman, "Color fusion effect on deep learning classification of uveal melanoma," 09 2023.
- [6] M. M. Taye, "Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions," *Computers*, vol. 12, no. 5, 2023.
- [7] L. Alzubaidi, J. Bai, A. Al-Sabaawi, J. Santamaría, A. S. Albahri, B. S. N. Al-dabbagh, M. A. Fadhel, M. Manoufali, J. Zhang, A. H. Al-Timemy, et al., "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *Journal of Big Data*, vol. 10, no. 1, p. 46, 2023.
- [8] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. a review," *Applied Sciences*, vol. 10, no. 6, p. 2021, 2020.
- [9] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2015.
- [10] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [11] K. Upadhyay, M. Agrawal, and P. Vashist, "Characteristic patch-based deep and handcrafted feature learning for red lesion segmentation in fundus images," *Biomedical Signal Processing and Control*, vol. 79, p. 104123, 2023.
- [12] M. Biglarbeiki, E. Mohammed, R. Souza, B. Far, E. Weis, and T. Crump, "Choroidal nevi classification in fundus images using a patch-based deep learning approach," in *2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology*, pp. 177–178, 2023.
- [13] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," Jan. 2023.
- [14] S. Narayanan, "Deep learning of fundus images and optical coherence tomography images for ocular disease detection—a review," *Multimedia Tools and Applications*, pp. 1–45, 2024.
- [15] C. Lam, C. Yu, L. Huang, and D. Rubin, "Retinal Lesion Detection With Deep Learning Using Image Patches," *Investigative Ophthalmology & Visual Science*, vol. 59, pp. 590–596, 01 2018.
- [16] C. Wang, Z. Zhao, and Y. Yi, "Fine retinal vessel segmentation by combining nest u-net and patch-learning," *Soft Computing*, vol. 25, pp. 1–14, 04 2021.
- [17] A. Skouta, A. Elmoufidi, S. jai andalousi, and O. Ouchetto, "Hemorrhage semantic segmentation in fundus images for the diagnosis of diabetic retinopathy by using a convolutional neural network," *Journal of Big Data*, vol. 9, 06 2022.
- [18] L. K. Singh, M. Khanna, S. Thawkar, and R. Singh, "Deep-learning based system for effective and automatic blood vessel segmentation from retinal fundus images," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 6005–6049, 2024.
- [19] 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: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.
- [20] A. Dumitriu, F. Tatui, F. Miron, R. T. Ionescu, and R. Timofte, "Rip current segmentation: A novel benchmark and yolov8 baseline results," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1261–1271, 2023.
- [21] X. Yue, K. Qi, X. Na, Y. Zhang, Y. Liu, and C. Liu, "Improved yolov8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage," *Agriculture*, vol. 13, no. 8, 2023.
- [22] Z. Ameli, S. J. Nesheli, and E. N. Landis, "Deep learning-based steel bridge corrosion segmentation and condition rating using mask rcnn and yolov8," *Infrastructures*, vol. 9, no. 1, 2024.
- [23] B. Preim and C. Botha, "Chapter 4 - image analysis for medical visualization," in *Visual Computing for Medicine (Second Edition)* (B. Preim and C. Botha, eds.), pp. 111–175, Boston: Morgan Kaufmann, second edition ed., 2014.
- [24] G. Wang, Y. Wang, H. Li, X. Chen, H. Lu, Y. Ma, C. Peng, Y. Wang, and L. Tang, "Morphological background detection and illumination normalization of text image with poor lighting," *PLoS One*, vol. 9, no. 11, p. e110991, 2014.
- [25] K. A. M. Said, A. B. Jambek, and N. Sulaiman, "A study of image processing using morphological opening and closing processes," *International Journal of Control Theory and Applications*, vol. 9, no. 31, pp. 15–21, 2016.
- [26] A. Reinke, M. Eisenmann, M. D. Tizabi, C. Sudre, T. Radsch, M. Antonelli, T. Arbel, S. Bakas, M. J. Cardoso, V. Cheplygina, K. Farahani, B. Glocker, D. Heckmann-Nötzel, F. Isensee, P. Jannin, C. Kahn, J. Kleesiek, T. Kurc, M. Kozubek, and A. Noyan, "Common limitations of image processing metrics: A picture story," 04 2021.
- [27] C. L. Shields, M. Furuta, E. L. Berman, J. D. Zahler, D. M. Hoberman, D. H. Dinh, A. Mashayekhi, and J. A. Shields, "Choroidal Nevus Transformation Into Melanoma: Analysis of 2514 Consecutive Cases," *Archives of Ophthalmology*, vol. 127, pp. 981–987, 08 2009.
- [28] E. Dugas, J. Jared, and W. Cukierski, "Diabetic retinopathy detection," 2015.