The Optic Disc Detection and Segmentation in Retinal Fundus Images Utilizing You Only Look Once (YOLO) Method

Zahraa Jabbar Hussein[1,2] and Enas Hamood Al-Saadi[3]

[1]*Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq*
[2]*Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq*
[3]*Department of Computer, College of Education for Pure Science, University of Babylon, Babylon, Iraq*

Abstract: Automated analysis of retinal images is a crucial diagnostic method for the early detection of disorders affecting the eyes, such as glaucoma, diabetic macula edema (DME), and retinopathy brought on by diabetes. The current study introduces a reliable technique for segmenting and detecting optic discs using a deep learning-based technique. This is comparable to the initial stage of creating a diagnostic system supported by a computer for diabetic macula edema in retinal images. The suggested approach is predicated on the recommended approach using the YOLO (You Only Look Once) algorithm for detection objects and segmentation for bounding boxes that belong to the same category, comparing the Intersection over Union (IOU) values of each bounding box with those of the others. If IOU values are higher than thresholds, they will consider them the same targets and maintain the boundary boxes with the highest reliability. Three retinal image databases that are accessible to the public are used to quantitatively assess the technique: Messidor-1, Messidor-2, and the IDRID Database. The technique yields a success rate of 99.5% for optic disc identification, a precision of 99.9%, and a recall of 100% for Messidor-1, and testing for all Messidor-2 and IDRID databases accepts accuracy of 99.1% and 98.7% respectively. When it comes to the identification and border extraction of the optic disc, this special technique has demonstrated a notable improvement over the previous approach.

Keywords: Fundus images, You only look once, Diabetic, Optic discs, Diabetic macular edema

1. INTRODUCTION

The center of the retina has a circular optical nerve area known as the optic disc (OD). Sharp focus and homogeneous lighting are necessary for retinal images. The macula appears dark, and the vessels are deep red in optimal illumination. In addition, hard exudate (HE) and OD are also vivid yellows. Diseases alter retinal images, making it more challenging to identify structures. Figure 1 explains normal eyes and the main components [1]. Additionally, Figure 2 displays images of the retinal fundus that are normal and pathological [2]. The optic disc may appear differently, and abnormal areas may be visible in the retinal image due to the presence of illnesses. Recent advancements in digital image processing and gathering have allowed for the use of medical images to detect and treat a variety of eye conditions.

Papers New technical developments are developed every day. The techniques employed in the medical sector are part of this progression. Data on OD diameter may be applied to the research of glaucoma and diabetic retinopathy.

Additionally, because some illnesses, such as diabetic macular edema, rely on hard exudates for disease diagnosis and have the same coloristic characteristics, they require detection and segmentation using OD. Therefore, a regular eye exam is required to identify these problems. Automated equipment can perform this kind of inspection [3].

This paper aims to build an effective model for detecting and segmenting the optic disc, which causes problems and errors in the process of diagnosing eye diseases for several reasons (size, color, texture, etc.). For example, one of the eye diseases is hard exudate (HE), which is similar to OD in brightness and yellow color. Therefore, the diseases change the images of the retina, making it more difficult to identify the disease. The contribution lies in presenting this effective model, which consists of several stages of segmentation with high accuracy for all types of images in various datasets. This proposed model is considered efficient in dealing with all eye diseases that require optic disc removal.

The phrase "optical disk detection" describes a procedure in retinography where the middle of the OD is located and is identified using computer technology.

The technique of delineating and subsequently removing the OD from the images is known as segmentation of the OD. High-intermediate gray-scale pixels can be used to place the OD since it is the portion of the retina that is most lit [4]. This method is not the most effective, despite the fact that numerous high-intensity areas, such as exudates, might be seen in the retinal images and cause the OD to appear in the wrong place.

Both domestic and international scientists have presented a number of disc localizations and segmentations procedures based on the unique characteristics of the fundus OD.
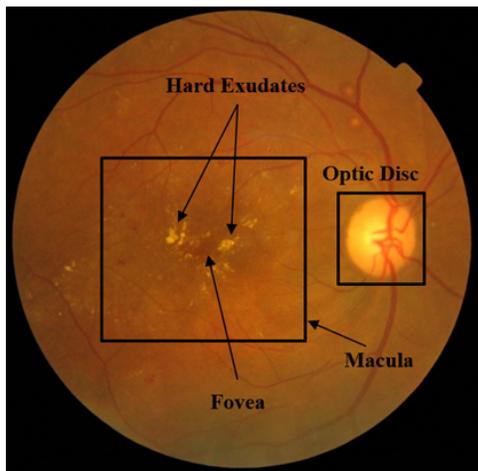

Figure 1. An image of the retina and its essential parts

in Figure 2 the first row (a) has two images from the Messidor-1 dataset: the fundus retinal image in normal does not have any hard exudats. In the second row (b), there is the hard exudate, which appears spread out in yellow because the patient has diabetic macula Edema or diabetic retina.
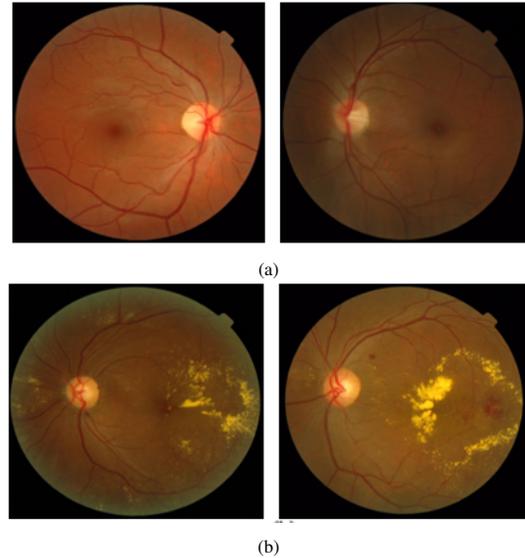

Figure 2. (a) first row normal retinal fundus images and (b) second row abnormal retinal fundus images

While early research concentrated on determining the precise OD location, more recent work has tried to predict the general shape. The following describes many methods for identifying and dividing the OD [5]. In retinal imaging, OD is frequently identified using geometric data gathered from the vasculature [6]. Hoover and Goldbaum suggested a method for detecting OD that heavily depended on the joining of numerous vascular tree branches [7]. Using this approach, it was discovered that the OD served as the system's nerve center. They are able to locate the intersection techniques for locating and segmenting the OD. Welfer et al. use this knowledge to help identify the OD because all retinal arteries have the OD as their origin [8]. Welfer et al. developed this same paradigm by employing multiscale morphological techniques. Even if it is challenging to locate the main blood vessels, where the OD is located may be identified by examining the OD, the geometric connection, and the major blood vessels [9]. The retinal vasculature's direction and brightness are utilized to estimate the OD. They recommend a two-step process [10]. Projecting the image's characteristics onto the x-axis yields the x-location of the OD first; the y-location of the OD is then gained by comparing the x-location to the original image. Image intensity and template matching have been used in several studies to pinpoint the OD.

Youssif et al. offer a method for normalizing contrast after first adjusting the initial green channel's strength to normal [11]. A diagram illustrating a retinal vascular tree's course is used to find the most likely location of vessels in an image using a template match. Muangnak and others. Utilize a hybrid optic-disc detection methodology. In this approach, the vessel transforms and vessel-vector-based phases are combined. In the analysis of portraits, a map

showing the path of a retinal vascular tree is used to find the most likely location of vessels in an image using a template match. Muangnak and others. Utilize a hybrid method for optic-disc detection. In this approach, the vessel transforms and vessel-vector-based phases are combined in the analysis of portraits [12]. Meng et al. employ convolutional neural networks (CNNs) to detect the OD [13]. Reza created a method that does not rely on blood vessels or a backdrop mask to identify OD automatically in retinal fundus photos.

In the current work, the author exploited an OD and a circle operator [14]. The authors of the study, Lu and Lim, identified the presence of OD by contrasting the OD's intensity with that of surrounding vessels. To overcome the drawbacks of the technique described, the authors included a line operator [15]. To identify OD and the corresponding sources, Roy Chowdhury et al. assessed the difference in intensity between the surrounding blood vessels and the OD [16]. By analyzing the vessel data and the retinal image's intensity, Zou et al. utilized the model for verification to calculate the OD [17]. Ramavath and Babu used the feature match approach to detect if an inquiry image is comparable to the original image used for an electoral OD. Here, background noise was removed from the image using the median filter, and an algorithm based on the firefly method was used to find the OD [18]. Abdullah et al. entered the special qualities as their input [19], and released a conventional image processing method that uses the Hough transform to identify spherical objects and their radii, perimeters, and outlines. Given that dark blood veins arise from the light zone known as the OD, Bharkad developed an equilateral filter to lessen this fluctuation and aid in future segmentation applications [20]. Walter et al. suggested an operator for shadow correction for the purpose of minimizing the effect of minute changes in the background; this operator is applied to the green channel [21]. In particular, the macula and the OD were distinguished and separated by the Canny edge filter and Otsu threshold used by Seo et al. [22].

In transformation and geodetic reconstruction, in particular, Stapor et al. introduced partitioning the region of OD and excavation using an approach based on mathematical morphology [23]. Using the biggest local variance, Kande et al.'s research identified the location of the OD center [24]. Due to its distinctive shape, Lupascu et al. identified the OD thanks to its round form and very elevated luminosity in contrast to the remainder of the image's structure [25].

This research attempts to build a deep learning technique called Your-Only-Lock-Once Networks (YOLOX) for autonomously recognizing and segmenting optical discs (ODx). Finding the precise position of one of the key goals was the optic disc, since it serves as the starting point for many studies whose main objective is to find and identify common retinal problems. These disorders' auto-diagnosis mainly relies on OD recognition. Samawi H. et al. recommended morphological methods and intensity thresholding for segmenting the optic disc [5]. The development

of an automated, reliable system that can be used with images that include pathologies is another goal of this effort. Two publicly accessible datasets of retinal images were utilized to evaluate the suggested techniques' robustness. The organization of this article is as follows: The databases exploited in the current study are listed in Section 2. Is recent work in Section 3. The recommended methods and materials for these measurements are also characterized in Section 4, and Section 5 Measures of the Efficiency of Identification and Division OD displays the outcomes of several strategies' experiments. The findings are described in Section 6.

This paper is considered a prelude to another paper in the works to discover eye diseases. In order to obtain high accuracy in classifying eye diseases (diseases that are similar in color to OD), the OD must be cut off. Therefore, one of the motivations for this model is to obtain an image free of various problems, which helps in the process of building an effective and accurate model that helps doctors diagnose eye diseases.

2. RELATED WORK

In the current section, we first clarify a recent method used in segmentation optic discs, then provide a thorough explanation of the architecture of the YOLOv5 model.

A. Recent Works

Shaikha et al. in 2019 [26] This paper discusses the technique for identification and division of optic discs in retinal fundus photos, concentrating on its significance in automatic diabetic retinopathy (DR) detection using the three datasets, DIARETDB0, DIARETDB1, and DRIVE. The authors utilize a histogram template to find the optic disc center using a matching algorithm accurately. They emphasize the importance of preprocessing steps, such as using binary masks to remove noisy regions and backgrounds from the images. Additionally, the authors highlight the varying weights assigned to different color planes (red, green and blue) in the similarity score calculation for optic disc segmentation. Template matching and histogram analysis are key techniques employed in this study. In the result faced by the misdetection of OD, hence the effect of the segmentation wrong part of OD Reagion. Interestingly, the result has an accuracy of 96%, 98%, and 98%, respectively.

Jiang et al. in 2020 [27] proposed segmentation of the optic disc aimed at improving glaucoma detection. Assuming elliptical shapes for the disc regions, we propose an end-to-end joint (RCNN) region-based convolutional neural network. Atrous convolution is incorporated to enhance feature extraction. Joint RCNN includes a disc proposal network (DPN) for generating bounding box proposals. This module aids in selecting suitable bounding boxes for both optic disc detection and comprehensive experiments demonstrate that our joint RCNN model accuracy was 0.85% and 0.901, respectively.

Maiti et al. in 2022 [28] proposed a modified convo-

lution network. This is a deep learning algorithm aimed at automatically identifying and dividing the objective of the optic disc in fundus photos to assist in the detection of diabetic retinopathy and streamlining the screening process for glaucoma. The optic disc's segmentation is pivotal for assessing retinal pathologies accurately. The study explores different encoder architectures like ResNet34, DenseNet121, VGG11, VGG13, VGG16, VGG19, and InceptionV3 to identify the most efficient one. Ultimately, the VGG16 architecture is selected as the encoder, complemented by a symmetric decoder structure to enhance object segmentation. The proposed approach's effectiveness is evaluated across multiple fundus image datasets, including DRIVE, STARE, CHASE-DB1, DIARETDB0, DIARETDB1, MESSIDOR, and IDRiD. Evaluation metrics such as intersection-over-union, dice coefficient, accuracy, and sensitivity are employed to gauge the model's performance. Boasting an overall accuracy of 99.44% and significant potential for early diagnosis and intervention in diabetic retinopathy and glaucoma.

Nazir et al. in 2021 [29] propose an automated framework for detecting glaucoma using deep learning based on Mask-RCNN on blurred retinal images. Glaucoma detection is challenging due to subtle symptoms, hence the need for automated detection. The framework includes data augmentation, blur introduction and ground-truth annotation generation. DenseNet-77 is integrated into Mask-RCNN for feature extraction, facilitating OD localization and segmentation. The evaluation on the one dataset is ORIGA, with a total of 30. The results show recall, F-measure, and IOU 0.965, 0.963, and 0.97, which protected the patient from reaching the stage of blindness.

Wisaeng In 2023 [30], this study, proposes an approach combining Fuzzy K-Means Thresholding (FKMT) and Morphological Operations with Pixel Density Features (MOPDF) to address these challenges. Initially, the input retinal images are threshold and fuzzy K-Means clustering is used to coarsely segment them, distinguishing the OD from neighbouring organs with similar intensities. The segmentation is then fine-tuned by morphological processes that use pixel density feature computations, minimizing false detections in tiny OD pixels. Lastly, the Sobel edge detection algorithm accurately identifies the OD region. Following optimization, our method achieves high sensitivity, specificity, and accuracy scores: DiaretDB0 reported 96.74%, 96.78%, and 96.92%; DiaretDB1 reported 97.12%, 97.10%, and 97.75%; and the (STARE) dataset reported 97.19%, 97.47%, and 97.43%.

Septiarini et al. (2023) [31] suggest a deep learning based automated optic disc region segmentation technique based on a convolutional neural network (CNN). We exploited both a public dataset and a private dataset that included retinal fundus images; the Retinal Fundus Glaucoma Challenge (REFUGE) collection included 350 images. Using the original images downsized to 640 × 640

input data, the technique creates a region-of-interest (ROI) images using a CNN with a single-shot multi-box detector (MobileNetV2). Augmentation, scaling, and normalization were among the pre-processing techniques applied. Subsequently, a U-Net model with 128 × 128 input data was employed for optic disc segmentation. The outcome was for the private dataset, the F1-score, dice score, and intersection over union are, respectively, 0.9880, 0.9852, and 0.9763; for the REFUGE dataset, they are, 0.9854, 0.9838, and 0.9712.

Bazi et al. (2024) [32] present two key contributions. Firstly, it introduces an apparatus with segmentation heads based on transformers to recognize discs in images of the retinal fundus. Following training on both the original and cropped images via the optical disc, the model presents two different topologies reliant on the multiscale Swin transformer and the monolithic vanilla vision transformer. The test image is initially cropped using the first head of inference, and furthermore, refined using the second head to produce a segmented map. With references to state-of-the-art techniques, the architecture exhibits promising results and is fully trainable, removing the requirement for pre- or post-processing chores. Second, a leave-one-out assessment technique is used to test the models' generalization ability on a brand-new benchmark issue that includes eight retinal datasets. Average dice scores for the disk are 85.31% for the hierarchical Swin-based model, and 85.33% for the monolithic ViT-based model.

Islam et al. in 2023 [33] We suggest a deep learning-driven method to automatically pinpoint the optical disc area in human retinal images, utilizing diverse publicly accessible datasets. This technique frames the task as an image segmentation challenge and incorporates an attention-based residual U-Net architecture. Our experimentation demonstrates that our method achieves pixel-level accuracy of over 99% in detecting the optical disc region, along with approximately 95% in Matthew's correlation coefficient. Comparison with alternative U-Net variations employing different encoder CNN architectures confirms the superior performance of our proposed approach across multiple evaluation metrics. This underscores the efficacy of our method in precisely identifying the optical disc in human retinal images, offering significant insights into an individual's health status.

Veena, H.N. et al. in 2022 [34] This study proposes an efficient framework for glaucoma diagnosis, focusing on segmenting the optic disc to calculate the Disc Ratio (CDR). Leveraging deep learning techniques, the proposed system employs novel CNN architectures tailored specifically for this task. It utilizes two distinct CNN models to segment the Optic Disc (OD), aiming to achieve improved accuracy. The model is trained and evaluated on the publicly available DRISHTI-GS database. Impressively, the proposed system achieves a segmentation accuracy of 98% for the optic disc. This underscores the effectiveness of the deep learning-based approach in accurately identifying and delineating

crucial structures relevant to glaucoma diagnosis.

B. YOLOv5 Deep Network Model

YOLO is a computer vision object detection algorithm. For object detection, a neural network is utilized in a single pass, directly predicting bounding boxes and class probabilities. YOLO's holistic approach considers the entire image, enabling it to capture context and make accurate predictions [35], [36].

Generally, the architecture of YOLO versions has improved significantly since the initial version was released. The progress was cumulative, and the new versions with improved speed and accuracy were published due to the many copies presented, each of which used the prior versions.

YOLO version one (YOLOv1) [37] was presented by Joseph Redmon et al. in 2016 as a real-time object detection method. Its architecture was a custom network, including two fully linked layers and 24 convolutional layers based on the Google-Net framework. The Yolov1 algorithm splits the source picture into a grid of S×S. Every grid cell in the image predicts several bounding boxes, objectness (confidence that the box contains an object), and probabilities of classes for those boxes.

The objectivity multiplied by the probability of the class generates the confidence score for the class-specific (indicates how well the class object fits within the expected box) for each class object and, consequently, for each bounding box. For simplicity, each cell predicts several bounding boxes. Each bounding box has coordinates, a class, and a confidence score.

The box's width, height, and center (x, y) are its coordinates. The low-confidence detections are filtered out using a threshold to eliminate the number of bounding boxes. Lastly, non-suppression is employed as a post-processing approach to keeping only the most precise detection of each object and eliminate overlapping boxes [39].

YOLOv5 instead, a series of improvements and extensions made to the previous versions of the YOLO architecture released in January 2022 by Ultralytics. Most changes in YOLOv5 related to model scaling and architecture were found in the code and the documentation of Ultralytics YOLOv5. Three factors led us to select YOLOv5 as our first learner.

YOLOv5 first combined Darknet with the cross-stage partial network (CSP-Net) to form CSPDark-Net, which serves as the network's central hub. With large-scale backbones, CSP-Net resolves the issue of recurring gradient information by ensuring accuracy and speed of inference. While also minimizing model size, modifying the gradient of the feature map, reducing FLOPS (floating-point operations per second), and model parameters. Rapidity and precision are essential for sperm cell detection, and the size of the model affects how well it infers information

on edge devices with constrained resources. Second, the YOLOv5 employed a route aggregation network (PA-Net) as its neck in order to enhance information flow. With an enhanced bottom-up strategy, PA-Net leverages a unique feature pyramid network (FPN) design to improve the transmission of low-level features. In addition, adaptive feature pooling guarantees that relevant data from every feature level reaches the subsequent subnetwork by linking the feature grid for each level of feature. Furthermore, by enhancing accurate localization signals in the lower layers, PA-Net significantly improves the object's positional precision. Lastly, the YOLO layer, the brains of YOLOv5, enabling the model to handle five distinct sizes by generating three alternative feature map sizes to enable multi-scale prediction: small (YOLOv5s), medium (YOLOv5m), large (YOLOv5l), extra-large (YOLOv5x), and nano (YOLOv5n), Figure 3. illustrates the YOLOv5 architecture.

3. CHALLENGING AND PROBLEMS FACED WITH OPTIC DISC DETECTION (ODD)

A. Variable Appearance

As demonstrated in Figure 4, the appearance of the optic disc can fluctuate significantly across people and in various photographs, making it challenging to develop a single method that is efficient in all circumstances.
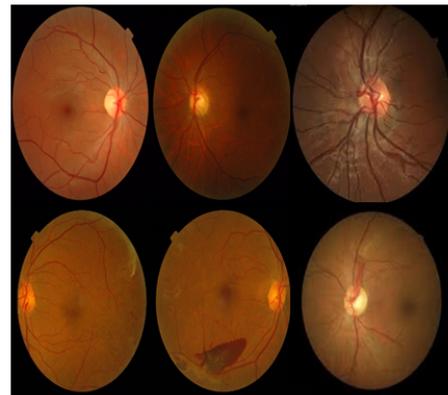

Figure 4. The different locations and size of the OD

B. Presence of Blood Vessels

Blood veins that surround the optic disc may resemble the disc in appearance, confusing detectors and producing false-positive results.

C. Low Contrast

In some images, the contrast between the optic disc and the backdrop may be poor, making it difficult for algorithms to distinguish the disc properly.

D. Noise

Eye movements, blinking, and other imaging conditions might create noise and aberrations in fundus images that compromise the precision of ODD algorithms.

E. *Class Imbalance and Computational Complexity*

Algorithm performance may be affected by datasets with an uneven distribution of positive and negative instances. Real-time detection with large datasets can also be computationally challenging.

4. DATA SETS

The effectiveness and adaptability of the suggested models to various data sets were evaluated using three publicly available data sets.

A. *Messidor-1*

This dataset consists of 1,200 fundus images from three ophthalmologic departments in France that have been
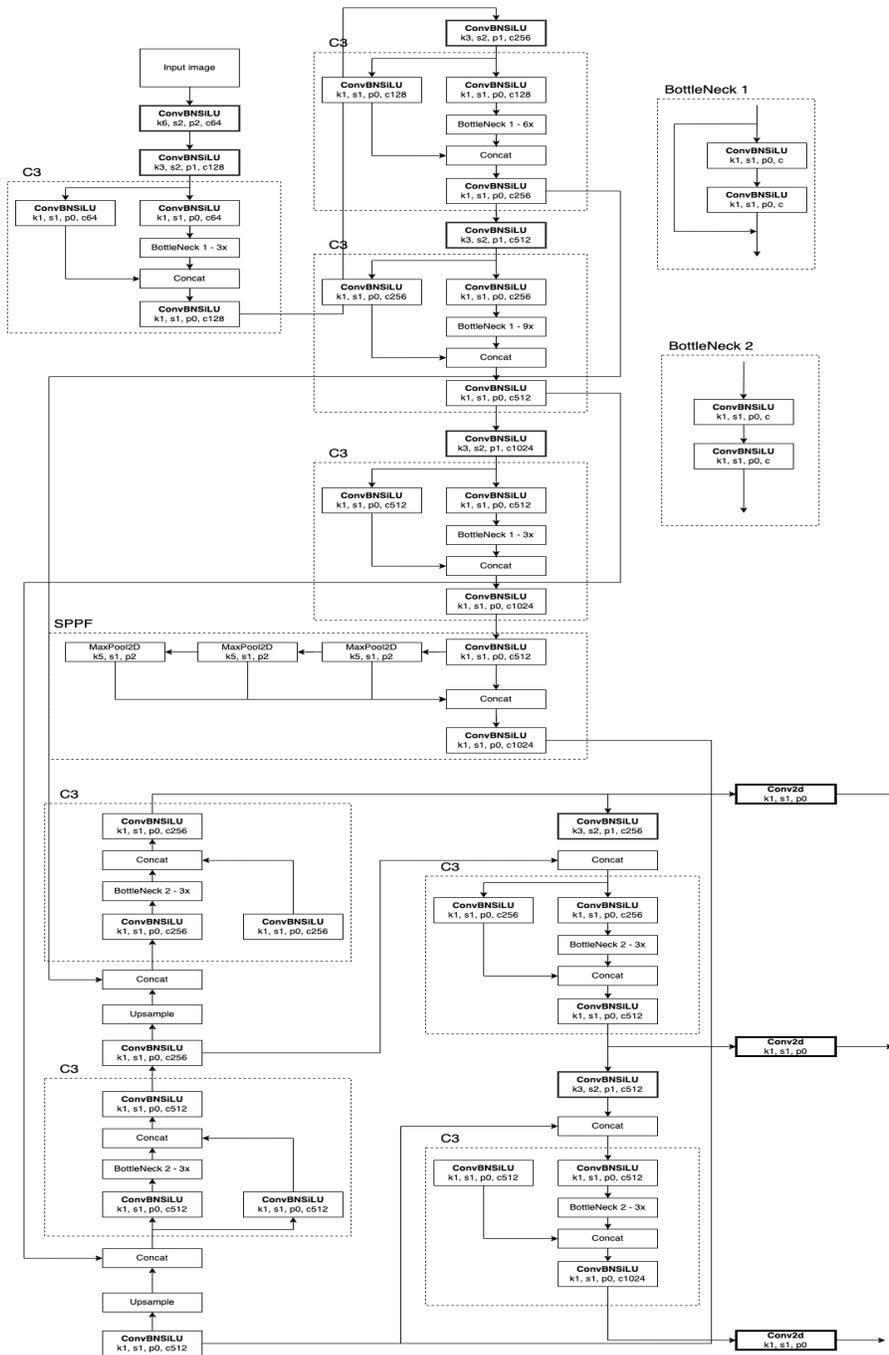

Figure 3. The network architecture of YOLOv5 [38]

labeled with two eye diseases: potential causes of diabetic retinopathy as well as diabetic macular edema. A 45-degree perspective nonmydriatic retinal camera, the Topcon TRC NW6, was used to take the images. We employed a four-point scale (0 for no HE, 1 for HE more than 1 DD from the macula, and 2 for HE inside 1 DD of the macula), as opposed to Messidor-1's use of a three-point scale (0 for no HE, 1 for HE greater than 1-disc diameter (DD) from the macula, and 2 for HE within 1 DD of the macula). The dimensions of each image are 2240 by 1488, as previously stated [40].

B. Messidor-2

The Messidor-2 collection contained 1,748 fundus images, five images were excluded. As seen in Figure 5 there are several issues with images, including those lacking the macula, the entire optic disc, or just part of it. We trained 1,743 graded fundus images for our investigation. Furthermore, the dimensions of every image are 2240 by 1488 [40].

Figure 5. Exclusion criteria of images

C. DRID (Indian Diabetic Retinopathy Image Dataset)

Having an average population of Indians makes it the first database. Annotation is done at the pixel level and encompasses both healthy retinal structures and typical lesions associated with diabetic retinopathy. Information about the amounts of diabetic macular edema and diabetic retinopathy in each image is included in the data collection. It is, thus, an ideal environment for the purpose of creating and evaluating image processing algorithms to detect diabetic retinopathy early on. In total, 455 images are included, with each image having a dimension of 4288 by 2848 pixels [41].

5. METHODOLOGY AND MATERIALS

In this section's proposed methodology, the prepressing step plays a significant role in getting more accurate results in the detection of the optic disc, which also depends on the proper detection and treatment of OD. Figure 6 explains the complete system flow. Here, the fundus image with a size of 2240 × 1488 pixels is given as an input. Then the median filter was applied in the pre-processing step to improve the quality, noise, and contrast of an image while preserving important image features such as edges, corners, and texture details. then upload this dataset to Roboflow for labeling OD, then split this data to train 60%, valid 20%, test 20%, download code format YOLOv5 PyTorch, and

train this dataset by using the YOLOv5 algorithm to locate and detect the region of the optic disk. After detecting the required region, then the region will be cropped, as shown in Figure 7. The detailed explanation of each section is explained below.
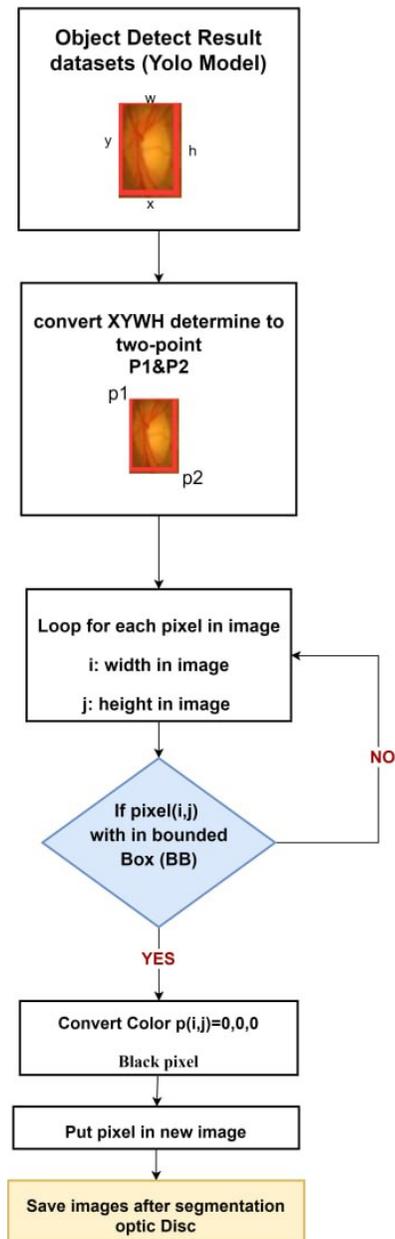

Figure 7. Design of the suggested segmentation division scheme

A. Preprocessing

In the two-step preprocessing, as:

- The images were recalled, and then the image file

formats were converted from GIF to JPG, then used.

- Median filter: images that show the fundus of the eye and are used in ophthalmology can also suffer from various types of noise that can reduce image clarity, making it challenging to analyze the retinal structures. For noise reduction, the median filter technique can enhance fundus images by reducing noise and smoothing the image while preserving the edges and key features and improving the visibility of important features and other details.

B. Workflow Design Sign Detection YOLO

Finding a collection of accurately labeled signs is necessary to prepare a YOLO data set. YOLOv5 is an object identification model with a feature extractor based on convolutional layers. The feature extractor part of the method uses the images from the dataset as input to extract patterns, giving a thorough understanding of the classes in the other dataset. To complete this study, the data set must be transformed to the YOLOv5 format.

The annotation is in YOLO format, which requires the presence of a TXT file with the same image name that contains the kind of class and four points that indicate the
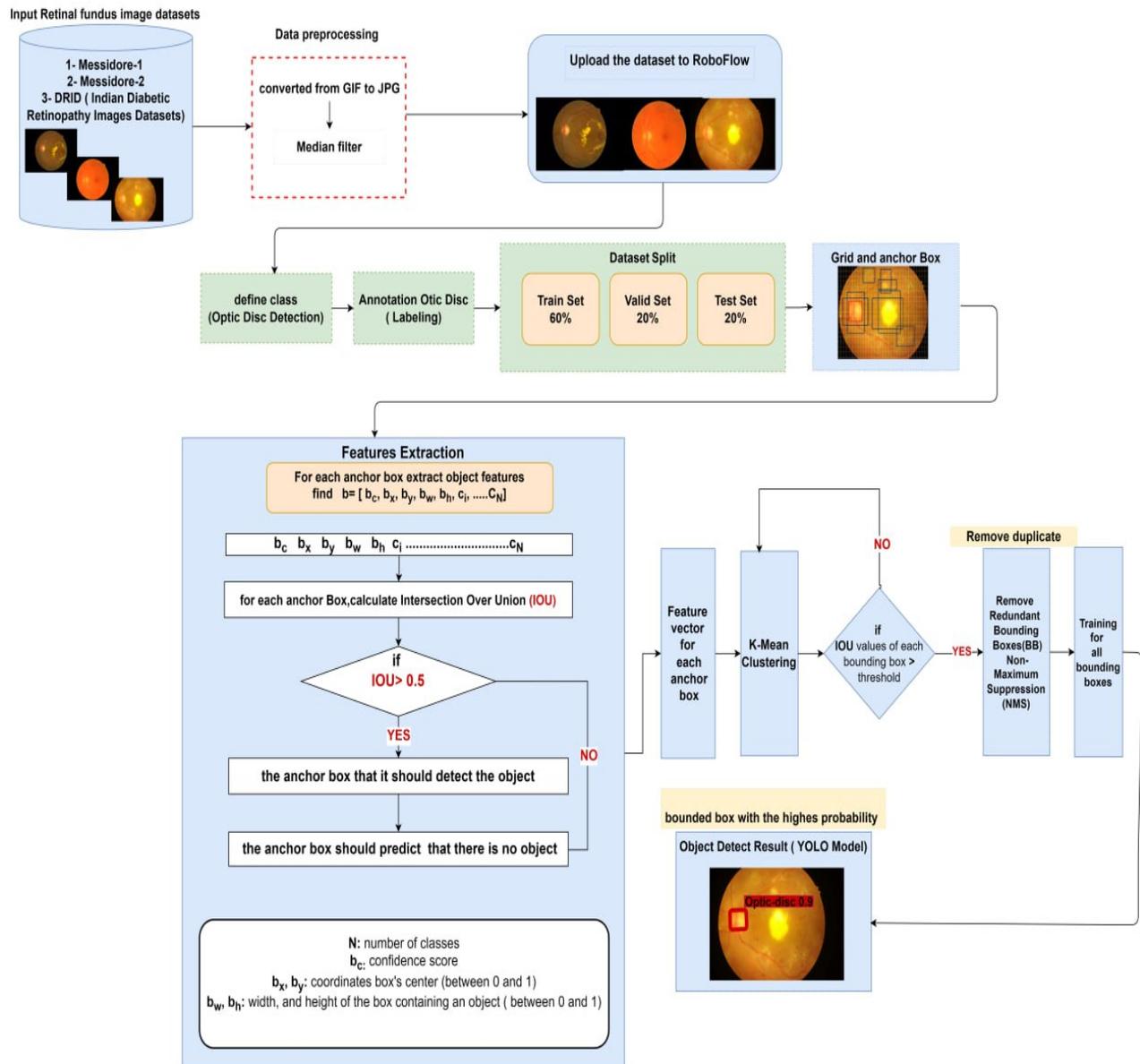

Figure 6. Design and Training Phases of the Proposed System to detection OD

bounding box. Algorithm 1. Discuss the steps of converting the selected data set into the YOLOv5 format.

Algorithm 1 Convert dataset to YOLOv5 format.

1) Step 1: Input Dataset
2) Step 2: Upload the dataset to RoboFlow
3) Step 3: define class_list
4) Step 4: annotation Optic Disc(labeling)
   ((Draw a bound box on the sign as accurately as possible. Then, choose a class from class_list))
5) Step 5: Choose the required formats (byTourch YOLOv5)
6) Step 6: export dataset to computer ((Download zip file (dataset).
7) Step 7: Output Dataset in YOLOv5 format (images, annotation files .txt, image information file)

C. Data Annotation and Labeling

The main component of labeling is drawing enclosing boxes around every item in the photo to show its precise location. The bounding-box annotations manage the object detector's YOLOv5 learning throughout the training phase. The detector drew a box around each thing it hoped to find, labeling each box with the kind of object it hoped to foresee within. A text file with one row for each sign seen in the image will be required for every image. Each row's box is a representation of the bounding box for an item whose class, x and y coordinates, width, and height were:

Class: displays the object type that was provided in the bordered box.

X and Y coordinates: reflect the bounded box's center point's X and Y coordinates, normalized between 0 and 1 (also known as the object's origin).

Width and height: indicating the box's dimensions, which are its height and breadth. (normalized to a range between 0 and 1)

D. Splitting the Data Set

There were 1200 images in Messidor-1 following the labeling process. There were 1744 images in Messidor-2 and 455 images in DRID that have been previously described. Train, valid, and test images (20 percent validation set, 20 percent testing set, and 60 percent training set) were separated into three groups for each data set. The weight value and biases of the model were established using the training set. The validation set, on the other hand, kept track of performance and noticed when the model started to overfit.

E. Grids and Anchor Boxes

Because they have lower pixel allocations and the same contrast as hard exudates (HE), it is challenging to detect optic discs in the real world. You-Only-Lock-Once (YOLO), a deep learning model, was employed to recognize objects

in order to resolve this issue. 2020 presentation of (open-source research towards potential vision AI techniques). The YOLOv5 has YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, which are the four variations. YOLOv5s has the smallest network structure, the least complexity, and the quickest detection speed among the four YOLOv5 versions. The images are initially split up into N grids, each having a size of S x S (19×19). The detection and localization of items contained in each N grid are their responsibility. The K-binding box and its confidence score are predicted in each grid cell. Remember that each grid can only recognize one object at a time. However, if multiple objects are in a single network, the concept of an anchor box must be used. Anchor boxes are preconfigured boxes with predetermined widths and heights of various shapes that assist in producing many outputs in a single grid as opposed to a single output and are chosen based on the size of the item in our training database. In contrast to other techniques, the YOLO algorithm splits the image into many cells according to the number of items covered. the process of making the anchor box and grid.

F. Extracting Features

The center of the bounding box is (x, y), the box's dimensions, and the confidence score to show how accurately the box believes the object can be predicted. These five parameters are: X is width, Y is height, and the confidence score indicates the accuracy of the confidence score to show how the box contains objects in the model. Show Figure 6 steps in detail.

G. K-mean

An anchor box is utilized to identify numerous targets inside a single grid cell. To enable greater IOU values, the K-means software does latitude clustering in an effort to resemble the anchor boxes and nearby ground truth as much as feasible. The ground and the bounding box of truth have the following relationship:

$$d(\text{box\_centroid}) = 1 - IOU\ (\text{box, centroid}) \qquad (1)$$

IoU (box, centroid) is the process of determining the intersection over union (*IoU*), a statistic used to gauge how accurately matching items are detected from the centroid to the box. As the value of the IOU rises, the distance falls. The stages of K-means clustering are explained in Algorithm 2.

Algorithm 2 The K-means clustering algorithm

---

1) **Input:** Points of data D K is the cluster number.
   First: Place k randomly chosen cluster centers at $(w_i, h_i), \ i \in 1, ..., \ k\}$.
   where: wi, hi reflect each anchor box's height and breadth.
2) **Second:** calculate the separation between each cluster centroid and each ground truth where:

$$d(box\_centroid) = 1 - IoU \ (box, \ centroid)$$

   Because the anchor box's location is movable, the clustering center and each ground truth's center point coincide.
3) **Third:**Recalculate the centroid for every cluster.
4) **Forth:** Repeat steps 2-3 until the clusters converge.
5) **Output:** Data Points with cluster memberships

H. Remove redundant boundary boxes (non-maximum suppression (NMS))

The YOLO approach uses non-maximum suppression to eliminate bounding boxes with low probability values. It is required to first go through the probability score assigned to each selection before selecting the one with the highest likelihood. The next stage is to avoid non-maximum suppression (NMS) by ignoring bounding boxes that have a high intersection with a bounding box that is likely to contain an item, as illustrated in the following three steps:

Algorithm 3 Steps non-maximum suppression (NMS)

---

1) **First:** for bounding boxes that belong to the same category, compare the *IoU* values of each bounding box with those of the others.
2) **Second:** If *IoU* values are higher than thresholds, they will consider them the same targets and maintain the boundary boxes with the highest reliability.
   If *IoU* > 0.5, tell the anchor box where the object should be detected.
   Else
   If *IoU* < 0.5, then the anchor box ought to indicate if an object is absent.
3) **Third:** Continue with steps 1 and 2 until all boxes are preserved.

I. Training YOLOV5 to Detect Optic Disc

The Messidor-1 data sets were employed throughout the training phase. Location and class predictions are made using YOLOv5. The actual anchor boxes and class labels are contrasted with the predicted boundary boxes and labels. The difference between the output of the model and the observed forecast is then determined using the function of loss. The loss function known as the Generalized Intersection over Union (*IoU*) is one of the YOLOv5 loss functions in which the You Only Live Once (YOLO) algorithm can back-propagate.

J. Detect Optic Disk by YOLOv5

These characteristics may be used in order to remove the symbol from the supplied photos, retrieve the interesting region, and eliminate the backgrounds of the images once the OD is identified using the YOLOv5 object detector and the bounding box features are obtained.

1) Segment the bounding box(es)

Every bounding box that YOLO finds has a vector (x, y, w, h, class) (center point coordinates (x, y), width, and height) in it, as was previously said. In Figure 6 the center point, width, and height are divided into two points that indicate the box's top left and bottom right points, respectively.

The OD was removed from the original images and turned into a black box after the coordinates box. Indicate the box's bottom right corner. The OD was excised from the original images and transformed into the black box seen in Figure 8 once the coordinate box had been produced.
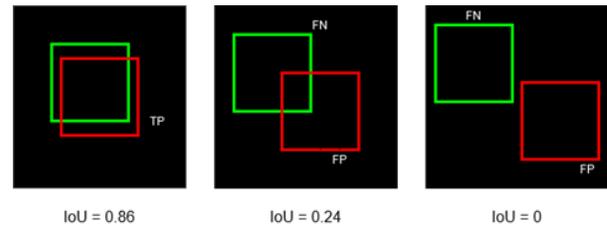

Figure 8. IOU performance [29]

6. RESULTS AND DISCUSSION

The Python3 programming language and the Colab Pro Google environment (GPU), GPU RAM 51GB, and Disk 166.8 GB are used to effectively assess the effectiveness of an object detector.

The basic stage in image processing systems (classification, object detection, segmentation, etc.) is pre-processing. It has a very effective role in increasing the accuracy of the system. The proposed system starts with image preprocessing through the use of a median filter. This filter works to remove noise, improve image contrast, and reduce blood vessels by taking the intermediate value and expanding the filter size, as it combines the background with the object.

It is essential to consider and evaluate relevant metrics. The task is a formidable barrier, as it necessitates delineating that every recognized item in the picture has bounding boxes all around it. The evaluation of detection performance is conducted using widely recognized measures. The accurate identification of an object presented in an image is addressed as a true positive (TP). The occurrence of an incorrect identification of an item is referred to as a false positive (FP). A false negative (FN) refers to an item in the image that goes undetected by the network. The precision measure is used for evaluating the number of

true positive forecasts based on predicted positives, where predicted positive equals (TP) and (FP). The division of all samples predicts true positive labels on all samples. In object identification, boundary box predictions are used to gauge decision performance rather than class predictions, which is where accuracy and recall come into play. All samples have true positives and false positives. It shows that the percentage of increasing behavior is accurately forecasted depending on FP. This criterion is measured according to Equation 2 [42].

$$Precision = \frac{TP}{(TP + FP)} \qquad (2)$$

The Recall (True Positive Rate) measure is used for evaluating the number of true positive forecasts based on actual positives, where actual positives equal true positive and false negatives. The division of all samples predicts true positive labels on all samples with true positive and false negative; this criterion is measured according to Equation 3 [43].

$$Recall = \frac{TP}{(TP + FN)} \qquad (3)$$

F1-score measure: the precision can be increased by increasing the recall parameter, or vice versa. In measuring the method's performance, precision and recall are very important. It is a combination of precision and recall. It determines to what extent and with what, precision the algorithm has been successful in predicting the increasing behavior. Importance The F1-score in accuracy for each class by calculating the resolution for each class is calculated separately; it can be calculated using Equation 4 after obtaining precision and recall.

$$F1_{\_score} = 2 * \frac{Precision * Recall}{Precision + Recall} \qquad (4)$$

The correctness of a predicted bounding box is ascertained using the intersection over the Union IoU relative to actual reality. A numeric value between 0 and 1 indicates the extent to which the predicted and bounding boxes for ground truth overlap [44]. If there is no overlap between any boxes, the IoU value is 0. A value of 1 for IoU means that the union of the two boxes is the same as their overlap, which means they fully overlap.

Calculated using Equation 5 [45]. The (IoU) according to the predicate class in training YOlOv5 was excellent; the more precise the bounding box, the higher the IoU value. Relying on the 0.5 criteria to evaluate the accuracy of a projected bounding box's IoU, the result for training YOLOv5.

$$IoU = \frac{TP}{TP + FP + FN} \qquad (5)$$

A common metric for assessing the precision of object detectors and classifiers, like YOLO, is Mean Average Precision (mAP). The region beneath the recall vs. precision curve that is provided to us is the typical accuracy for each class for every model. The average of this number

across all classes is known as mAP. Depending on the current detection barriers, the average point score (mAP) finds its value by averaging all classes and/or the total IoU thresholds. This criterion is measured according to Equation 6 [46]. Using diagnostic training data from the datasets (MESSIDOR1), the accuracy of the demonstration of the suggested technique was evaluated at 99.5

$$mAP = \frac{1}{|classes|} \sum_{c \in classes} \frac{TPc}{|FPc| + |TPc|} \qquad (6)$$

There were 1135 labels in the training dataset that we produced. Figure 9 displays the graph of width and height. We will increase the size of both datasets in relation to the present size of marked images. Better accuracy, less overfitting, and improved performance are the outcomes of this strategy.

Figure 10 demonstrates the coordinates of the central point distribution range (x, y) and the height and breadth of each genuine bounding box for each pest picture in the training dataset we acquired through the use of the K-means clustering approach. Furthermore, we deduced that throughout the training dataset, the majority of the pest targets are situated at the heart of the original images based on color depth, which shows that darker areas correspond to larger concentrations of pests. Moreover, the scatter plot, which depicts the correlation between the width and height as the x and y axes, shows that the majority of every bounding box's height and breadth are higher than the diagonal line, indicating a proportionate connection.

There are three initial sets of values for the bounding box that the model will select that, according to the range of sizes for these bounding boxes, are most similar to the size values in the dataset. These three sets of numbers line up with the three YOLOv5 scale detecting heads. Only modest adjustments to the size of the bounding box are required when utilizing the three sets of values for the anchor box to identify pests of varying sizes, which significantly lowers bounding box loss and boosts the model's detection effectiveness. Figure 11 displays a sample of the detection findings. Figure 12 shows accurate detection of training YOLOv5. Figure 13 shows measures of the Messidor-1 dataset. Figure 14 displays the result of segment OD.
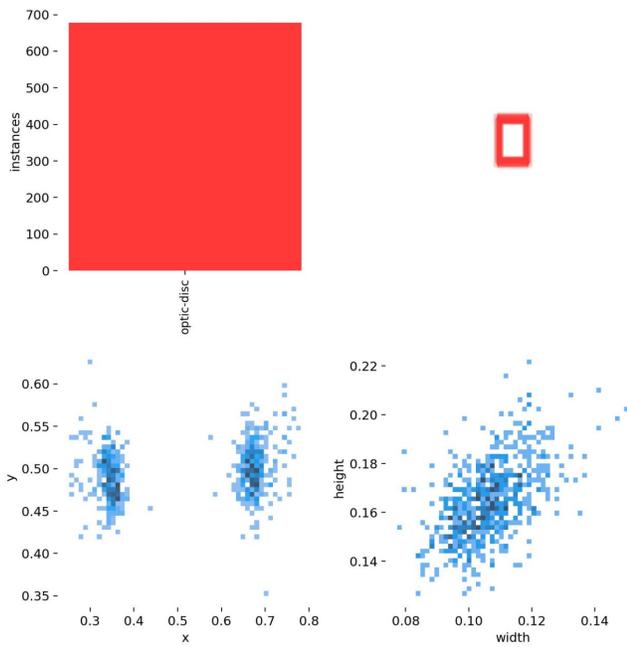
Figure 9. Described the width and height of the boxes as the labels' dimensions. designed using the Seaborn package
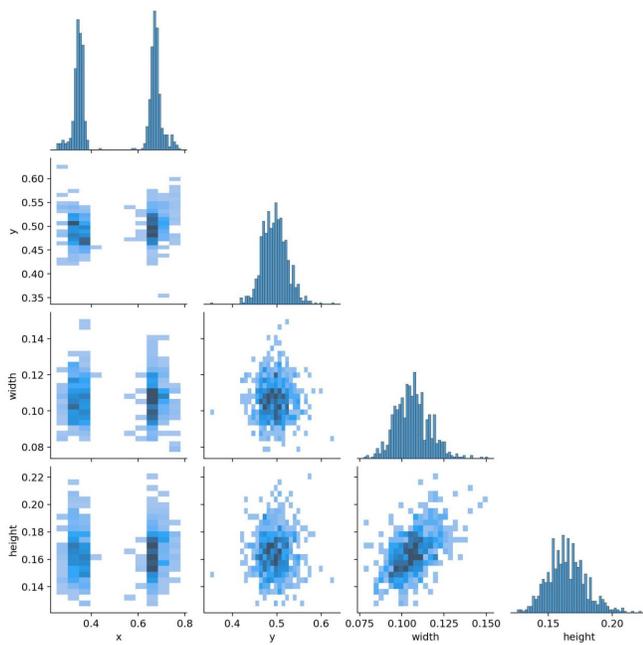

Figure 10. For each target pest in the dataset, determine its center point coordinates, breadth, height, and correlations using the K-means clustering approach
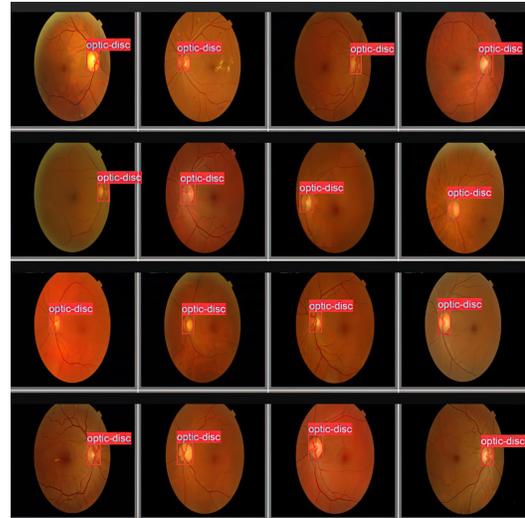

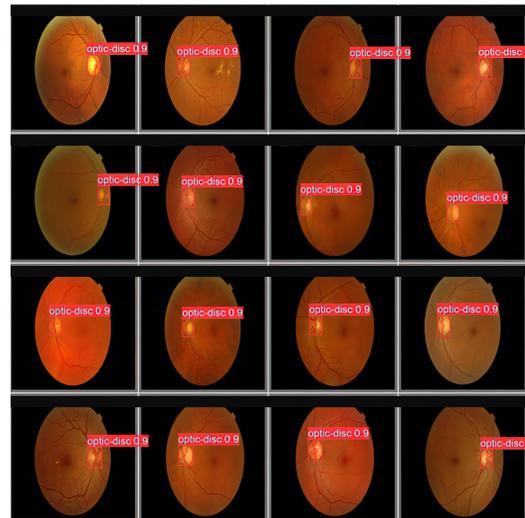Figure 11. The result for detection location OD

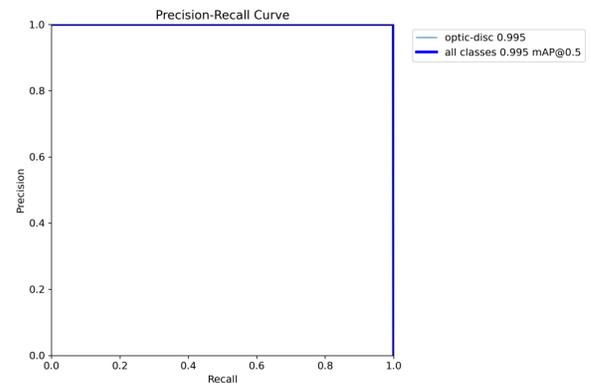Figure 12. The accuracy (IOU) detection in training YOLOv5-200

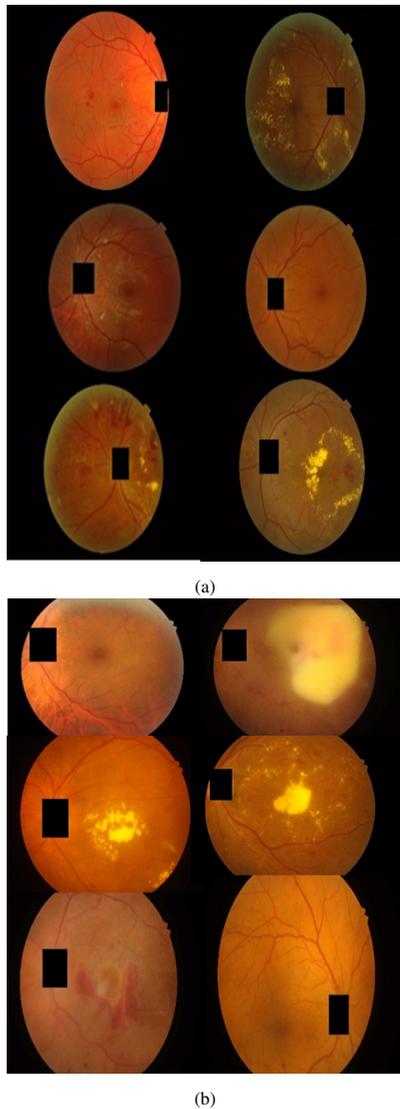Figure 13. YOLOv5 precision, recall, and mAP 0.5 in (Messidor-1) datasets

(a)

(b)

Figure 14. The test segment of optic disc in different locations Images Datasets: (a) Messidor-1 and Messidor-2, (b) DRiD (Indian Diabetic Retinopathy)

Table I gives details on the precise experimental settings that were utilized to train the novel network model that this study proposes. The training set up was 200 epochs with a batch size of 16. We changed the Messidor-1 dataset's input image size to 2240 × 1488. We started the learning rate at 0.01 and used SGD as the optimizer. Set at 0.99 and 0.0005, respectively, were the momentum and weight loss. By speeding up the model's training phase, this optimizer was able to attain the best detection performance. Training using the YOLOv5-pytorch dataset yielded pretrained weights.

The Messidore dataset's blue line shows the change in each pest's detection accuracy; generally, the F1 of each pest is directly correlated with the confidence. The suggested

model achieves an accuracy of up to 1.00. Furthermore, the training progress, accuracy, recall, mAP 0.5, and mAP 0.5 0.95 are shown in Figure 15.
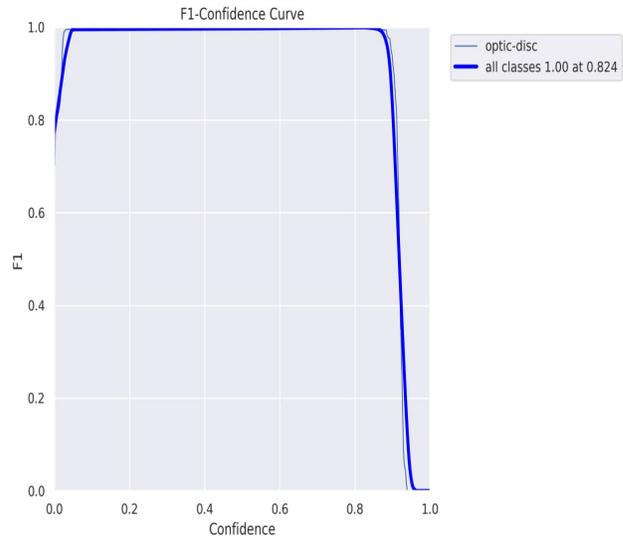

Figure 15. F1_curve in the training of Messidor-1 datasets

The optimal value (approaching one) is reached by mAP 0.5, recall, and precision.

TABLE II. Control parameter experiments

Models YOLOv5X	Size images	Testing accuracy (mAP 0.5)
Messidor-1	2240×1488	99.7
Messidor-2	2240×1488	99.1
DRID	4288×2848	98.7

Figure 16 demonstrates how training loss and detection accuracy are evolving over time in YOLOv5X for every epoch on the datasets used for training and validation. According to the figures, as the number of training epochs increases, there is a consistent decline in the boundary loss (box_loss), classification loss (cls_loss), and confidence score loss (obj_loss). Additionally, the corresponding values for, in the first 100 epochs, there is a discernible improvement in detection precision, recall, and mAP, then slowly increase from 50 to 150 epochs before finally reaching convergence.

Table III: We compare our proposed approach with many existing optic disc identification techniques in comparative experiments with other models. In addition, consider the strengths and weaknesses of each model.

After comparing the original model's mAP0.5 (IOU), mAP0.5-0.95, and F1-score with those of other modules in the YOLO version, Table IV shows how the detection metrics have evolved.

TABLE IV. Comparative experiments with other version of YOLO deep learing

Version YOLO	mAP0.5(IOU)	mAP0.5–0.95	F1- score
YOLOv3	90%	70.8%	90.1%
YOLOv4 pytorch	93%	74%	93.3%
YOLOv7 pytorch	92%	72.9%	92.2%
yolov8	98%	73.2%	92.8%
YOLOv9	98%	73.3%	92.4%
Proposed method	99%	80.2%	99.5%

The diversity and availability of study data play a major role in the success of any model. One of the biggest limitations that researchers face, especially in this topic, is the availability of a sufficient database for training when using deep learning algorithms. The accuracy of deep learning in producing a model capable of distinguishing or detecting an object is related to the amount of data available and the number of learning cycles. The greater these two characteristics, the more optimal and highly accurate the model is. The data collected is limited, so an augmentation was made to expand the amount of data. This method provides a variety of data in terms of the variety of parameters used, such as rotation, scaling, displacement, etc.

We tried to obtain data from a government and private hospital, but there was a privacy issue here, as it belonged to patients, and we were not allowed to give it. This is to demonstrate the applicability of the suggested model to private data outside the scope of the study. However, referring to Table III, it becomes clear how general the model is in its application to global data that was used by a group of researchers and showed effective results compared to those researches.

Also, the difference in the injuries of patients or the images taken of the eye is sometimes from the right side and sometimes from the left side, and this requires precision in treatment.

7. Conclusions

The eye is considered the most important part of the human body and may be exposed to external diseases such as allergies, causing apparent redness, or internal diseases such as cataracts or glaucoma, which cause blindness in its advanced stages. To identify these diseases with high accuracy, we have presented a model that is very effective in determining OD, which usually causes an error in the detection of the disease because it has the same color and texture as the disease. The proposed model begins with the optimization process, which contributes to removing noise

TABLE I. Control parameter experiments

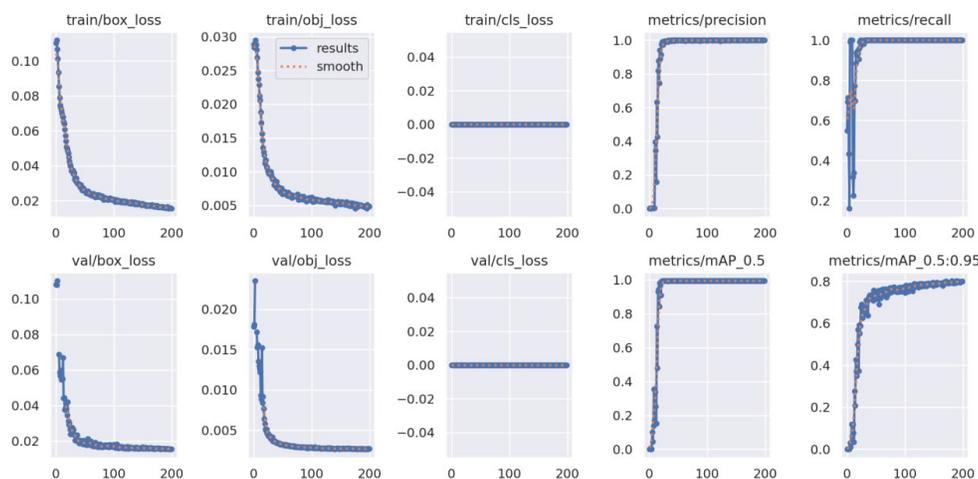| Models YOLOv5 Size	No. of layers	No. of parameters	Time of training in hours	Precision %	Recall %	mAP 0.5	mAP 0.5-0.95	F1- score
YOLOv5s	157	1760518	0.396	98.1	100	98.5	78.3	99.04
YOLOv5m	212	20852934	1.123	99.1	100	99.0	79.1	99.54
YOLOv5l	267	46108278	1.627	99.5	100	99.1	79.9	99.74
YOLOv5x	322	86173414	2.966	99.9	100	99.5	80.2	99.94

Figure 16. All Training results for YOLOv5x Performance Matrixes on Mesidor datasets for 200 epochs.

TABLE III. Comparative experiments.

Model	Year	Dataset	No. of image	Strengths	weaknesses	Resolution of image	Resize image	Training, Valid, Test	Accuracy
histogram template matching algorithm and OD size [26]	2019	DIARTDB0 DIARTDB1 DRIVE	130 89 40	Use binary masks for preprocessing.	The study does not discuss potential limitations or challenges encountered during the preprocessing techniques, and it does not compare the performance of their modified models with other state-of-the-art approaches.	1500×1152 1500×1152 768×584	190×190 190×190 80×80	-	0.96% 0.98% 0.98%
Joint-RCNN [27]	2020	ORIGA SCES	650 1676	Use deep convolutional layers for feature extraction from images.	Not explaining what techniques are used in preprocessing and how many databases there are after the augmentation	3072×2048 pixels	800×800	80%,10%,10%	0.85% 0.90%
modified convolution network [28]	2022	MESSIDOR DRIVE IDRiD	1200 40 455	use the deep algorithm for detection.	In this study there is a lack of extensive analysis or discussion regarding the modified CNN model.	2240×1488 768×584 4288×2848	-	80%,10%,10%	0.94%
Mask-RCNN [29]	2021	ORIGA	650	Localization and segmentation of OD using Custom Mask-RCNN	Due to the fixed size (28 × 28) of the segmentation masks, the method may struggle to generalize well to unseen or complex datasets. Different datasets may have varying levels of complexity, variability in image quality, and diverse anatomical structures, making it challenging for the model to adapt effectively across different datasets.	3072×2048	-	70%, -,30%	0.96%
histogram template matching algorithm and OD size [26]	2019	DIARTDB0 DIARTDB1 DRIVE	130 89 40	Use binary masks for preprocessing.	The study does not discuss potential limitations or challenges encountered during the preprocessing techniques, and it does not compare the performance of their modified models with other state-of-the-art approaches.	1500×1152 1500×1152 768×584	190×190 190×190 80×80	-	0.96% 0.98% 0.98%

TABLE III. Comparative experiments (*Continued*).

Model	Year	Dataset	No. of image	Strengths	weaknesses	Resolution of image	Resize image	Training, Valid, Test	Accuracy
Joint-RCNN [27]	2020	ORIGA SCES	650 1676	Use deep convolutional layers for feature extraction from images.	Not explaining what techniques are used in preprocessing and how many databases there are after the augmentation	3072×2048 pixels	800× 800	80%,10%,10%	0.85% 0.90%
modified convolution network [28]	2022	MESSIDOR DRIVE IDRiD	1200 40 455	use the deep algorithm for detection.	In this study there is a lack of extensive analysis or discussion regarding the modified CNN model.	2240×1488 768×584 4288×2848	-	80%,10%,10%	0.94%
Mask-RCNN [29]	2021	ORIGA	650	Localization and segmentation of OD using Custom Mask-RCNN	Due to the fixed size (28 × 28) of the segmentation masks, the method may struggle to generalize well to unseen or complex datasets. Different datasets may have varying levels of complexity, variability in image quality, and diverse anatomical structures, making it challenging for the model to adapt effectively across different datasets.	3072×2048		70%, -,30%	0.96%
Fuzzy K–Means Threshold (FKMT) and Morphological Operation with Pixel Feature Density (MOPDF) [30]	2023	DiaretDB0 DiaretDB1	616 images were randomly selected	The proposed method's strength lies in its comprehensive evaluation using both pixel-based and image-based criteria, which provides a holistic assessment of its performance. Addition used the Sobel edge detector.	The algorithm lacks any preprocessing, which affects the accuracy and generalizability of a new database.	1500×1152	500× 650	-	96.74%
CNN with a single-shot multibox detector (MobileNetV2)[31]	2023	private dataset (REFUGE) public	350 200	Resized and converted toa binary image using a preprocessor; furthermore, a data augmentation technique	Use private datasets and not generalizability on public data.	640×640 4288×2848	128× 128 640×640	70%,20%,10%	0.98%

TABLE III. Comparative experiments (*Continued*).

Model	Year	Dataset	No. of image	Strengths	weaknesses	Resolution of image	Resize image	Training, Valid, Test	Accuracy
Vision transformers: 1.anilla vision transformer (ViT) [32] 2.multiscale Swin transformer [32]	2024	eight retinal datasets (public& private)	The study explores two different transformer architectures, namely the vanilla Vision Transformer (ViT) and the Swin Transformer, as backbones for the segmentation model.	While the paper mentions the use of three retinal image databases, it lacks details on the characteristics of these datasets, such as the number of images, diversity, and potential biases.	-	-	-	-	93.55% 94.45%
U-Net [33]	2023	IDRiD Drishti-GS1 RIM-ONE	-	Focus on deep learning models.	The study does not discuss potential limitations or challenges encountered during the preprocessing techniques, and it does not compare the performance of their modified models with other state-of-the-art approaches.	-	-	-	0.95%
Convolutional Neural Networks (CNN's) [34]	2022	DRISHTI-GS	101	use deep learning CNN algorithms	Take 50 retinal fundus images from dataset. It is relatively small for training deep learning models, especially when aiming for high accuracy	2896×1944	128×128	50%,-,51%	0.98%
Proposed method	-	Messidor-1 Messidor-2 DRID	1200 1743 455	preprocessing steps Deep Learning Approach Wide Applicability: The technique is designed for the early detection of various eye disorders, including glaucoma, diabetic macula edema (DME), and retinopathy caused by diabetes. The study employs three publicly accessible retinal image databases with high resolution for quantitative assessment and testing High Success Rate and Precision	1. The model does not test on the private dataset. 2. Some diseases, such as optic nerve tumors, eye infections, and retinal atrophy, can affect the damage to the optic disc. For this reason, cannot be detected by the optic disc.	2240×1488 2240×1488 4288×2848	640×640 640 × 640 640×640	60%,20%,20%	0.99%

and enhancing the appearance of the photo so that it is treatable and reduces the error rate in detection. Then followed the stage of using the YOLO technique with segmentation of the OD area. When these two steps were combined, effective segmentation appeared, even for difficult images. The YOLO model was used for several versions, starting with the small (s) with an accuracy ranging from 98 to 99 and ending with the (x) version with an accuracy ranging from 99 to 99.9, demonstrating proficiency in detection and segmenting OD and on three global rules (Messidor-1, Messidor-2, DRID). The best accuracy was obtained for Messidor-1, which was 99.7 compared to Messidor-2, 99.1, and DRID, 98.7, which is considered lower because of the images' large size, which was reduced and some information was lost. Despite the challenges that emerged from the lack of data, the amount of noise, and the similarity of the disease to the disc, the system showed effectiveness in identifying and detecting OD. This model proposed in this paper is considered the first stage of disease detection. We have conducted another study under construction to identify three types of eye diseases. It is difficult to diagnose. Accuracy reached 99.7% with this model, and without it, it ranged between 70 and 80.

REFERENCES

[1] J. M. Molina-Casado, E. J. Carmona, and J. García-Feijoó, "Fast detection of the main anatomical structures in digital retinal images based on intra-and inter-structure relational knowledge," *Computer methods and programs in biomedicine*, vol. 149, pp. 55–68, 2017.

[2] X. Zhang, A. M. Tohari, F. Marcheggiani, X. Zhou, J. Reilly, L. Tiano, and X. Shu, "Therapeutic potential of co-enzyme q10 in retinal diseases," *Current medicinal chemistry*, vol. 24, no. 39, pp. 4329–4339, 2017.

[3] C. C. Kwan and A. A. Fawzi, "Imaging and biomarkers in diabetic macular edema and diabetic retinopathy," *Current diabetes reports*, vol. 19, no. 10, p. 95, 2019.

[4] H. J. Samawi, A. Y. Al-Sultan, and E. H. Al-Saadi, "Optic disc segmentation in retinal fundus images using morphological techniques and intensity thresholding," in *2020 International Conference on Computer Science and Software Engineering (CSASE)*. Duhok, Iraq: IEEE, 2020, pp. 302–307.

[5] L. Tan and J. Jiang, "Image processing basics," in *Digital Signal Processing*, 3rd ed., L. Tan and J. Jiang, Eds. Cambridge, MA, USA: Academic Press, 2019, ch. 3, p. 649–726.

[6] Z. J. Hussein and E. H. Al Saadi, "Recent developments used for early detection of diabetic macular edema by retinal images: A survey," in *2022 Iraqi International Conference on Communication and Information Technologies (IICCIT)*. Basrah, Iraq: IEEE, 2022, pp. 335–341.

[7] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 951–958, 2003.

[8] D. Welfer, J. Scharcanski, C. M. Kitamura, M. M. Dal Pizzol, L. W. Ludwig, and D. R. Marinho, "Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach," *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 124–137, 2010.

[9] D. Welfer, J. Scharcanski, and D. R. Marinho, "A morphologic two-stage approach for automated optic disk detection in color eye fundus images," *Pattern Recognition Letters*, vol. 34, no. 5, pp. 476–485, 2013.

[10] A. E. Mahfouz and A. S. Fahmy, "Fast localization of the optic disc using projection of image features," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3285–3289, 2010.

[11] A. A.-H. A.-R. Youssif, A. Z. Ghalwash, and A. A. S. A.-R. Ghoneim, "Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter," *IEEE transactions on medical imaging*, vol. 27, no. 1, pp. 11–18, 2007.

[12] N. Muangnak, P. Aimmanee, and S. Makhanov, "Automatic optic disk detection in retinal images using hybrid vessel phase portrait analysis," *Medical & biological engineering & computing*, vol. 56, pp. 583–598, 2018.

[13] X. Meng, X. Xi, L. Yang, G. Zhang, Y. Yin, and X. Chen, "Fast and effective optic disk localization based on convolutional neural network," *Neurocomputing*, vol. 312, pp. 285–295, 2018.

[14] M. N. Reza, "Automatic detection of optic disc in color fundus retinal images using circle operator," *Biomedical Signal Processing and Control*, vol. 45, pp. 274–283, 2018.

[15] S. Lu and J. H. Lim, "Automatic optic disc detection from retinal images by a line operator," *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 88–94, 2010.

[16] S. Roychowdhury, D. D. Koozekanani, S. N. Kuchinka, and K. K. Parhi, "Optic disc boundary and vessel origin segmentation of fundus images," *IEEE journal of biomedical and health informatics*, vol. 20, no. 6, pp. 1562–1574, 2015.

[17] B. Zou, C. Chen, C. Zhu, X. Duan, and Z. Chen, "Classified optic disc localization algorithm based on verification model," *Computers & Graphics*, vol. 70, pp. 281–287, 2018.

[18] S. A. Ramakanth and R. V. Babu, "Approximate nearest neighbour field based optic disk detection," *Computerized medical imaging and graphics*, vol. 38, no. 1, pp. 49–56, 2014.

[19] M. Abdullah, M. M. Fraz, and S. A. Barman, "Localization and segmentation of optic disc in retinal images using circular hough transform and grow-cut algorithm," *PeerJ*, vol. 4, p. e2003, 2016.

[20] S. Bharkad, "Automatic segmentation of optic disk in retinal images," *Biomedical Signal Processing and Control*, vol. 31, pp. 483–498, 2017.

[21] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, "A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina," *IEEE transactions on medical imaging*, vol. 21, no. 10, pp. 1236–1243, 2002.

[22] J. Seo, K. Kim, J. Kim, K. Park, and H. Chung, "Measurement of ocular torsion using digital fundus image," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1. San Francisco, CA, USA: IEEE, 2004, pp. 1711–1713.

[23] K. Stapor, A. Świtonski, R. Chrástek, and G. Michelson, "Segmentation of fundus eye images using methods of mathematical morphology for glaucoma diagnosis," in *Computational Science-*

ICCS 2004: 4th International Conference. Kraków, Poland: Springer, 2004, pp. 41–48.

[24] G. B. Kande, P. V. Subbaiah, and T. S. Savithri, "Segmentation of exudates and optic disk in retinal images," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing.* Bhubaneswar, India: IEEE, 2008, pp. 535–542.

[25] C. A. Lupascu, D. Tegolo, and L. Di Rosa, "Automated detection of optic disc location in retinal images," in *2008 21st IEEE International Symposium on Computer-Based Medical Systems.* Jyvaskyla, Finland: IEEE, 2008, pp. 17–22.

[26] H. K. Shaikha and A. B. Sallow, "Optic disc detection and segmentation in retinal fundus image," in *2019 International Conference on Advanced Science and Engineering (ICOASE).* Zakho - Duhok, Iraq: IEEE, 2019, pp. 23–28.

[27] Y. Jiang, L. Duan, J. Cheng, Z. Gu, H. Xia, H. Fu, C. Li, and J. Liu, "Jointrcnn: a region-based convolutional neural network for optic disc and cup segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 335–343, 2019.

[28] S. Maiti, D. Maji, A. K. Dhara, and G. Sarkar, "Automatic detection and segmentation of optic disc using a modified convolution network," *Biomedical Signal Processing and Control*, vol. 76, p. 103633, 2022.

[29] T. Nazir, A. Irtaza, and V. Starovoitov, "Optic disc and optic cup segmentation for glaucoma detection from blur retinal images using improved mask-rcnn," *International Journal of Optics*, vol. 2021, pp. 1–12, 2021.

[30] K. Wisaeng, "Automatic optic disc detection in retinal images using fkmt-mopdf." *Computer Systems Science & Engineering*, vol. 46, no. 1, p. 2569–2586, 2023.

[31] A. Septiarini, H. Hamdani, E. Setyaningsih, E. Junirianto, and F. Utaminingrum, "Automatic method for optic disc segmentation using deep learning on retinal fundus images," *Healthcare Informatics Research*, vol. 29, no. 2, p. 145, 2023.

[32] Y. Bazi, M. M. Al Rahhal, H. Elgibreen, and M. Zuair, "Vision transformers for segmentation of disc and cup in retinal fundus images," *Biomedical Signal Processing and Control*, vol. 91, p. 105915, 2024.

[33] M. T. Islam, F. Ahmed, M. Househ, and T. Alam, "Optical disc segmentation from retinal fundus images using deep learning," in *Healthcare Transformation with Informatics and Artificial Intelligence.* IOS Press, 2023, pp. 628–631.

[34] H. Veena, A. Muruganandham, and T. S. Kumaran, "A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 6187–6198, 2022.

[35] A. S. Mahdi and S. A. Mahmood, "Wildfire detection system using yolov5 deep learning model," *International Journal of Computing and Digital Systems*, vol. 14, no. 1, pp. 10 149–10 158, 2023.

[36] Y. A. Al Khafaji and N. K. El Abbadi, "Traffic signs detection and recognition using a combination of yolo and cnn," in *2022 Iraqi International Conference on Communication and Information Technologies (IICCIT).* Basrah, Iraq: IEEE, 2022, pp. 328–334.

[37] S. Hossain, H. Anzum, and S. Akhter, "Comparison of yolo (v3, v5) and mobilenet-ssd (v1, v2) for person identification using ear-biometrics," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1259–1271, 2024.

[38] M. Dobrovolny, J. Benes, J. Langer, O. Krejcar, and A. Selamat, "Study on sperm-cell detection using yolov5 architecture with labaled dataset," *Genes*, vol. 14, no. 2, p. 451, 2023.

[39] M. E. Khebbache, S. Bitam, and A. Mellouk, "A cost-effective deep active learning for object detection in automated driving systems," *International Journal of Computing and Digital Systems*, vol. 14, no. 1, p. 839–854, 2023.

[40] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay *et al.*, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis and Stereology*, vol. 33, no. 3, pp. 231–234, 2014.

[41] P. Prasanna, P. Samiksha, K. Ravi, K. Manesh, D. Girish, S. Vivek, and M. Fabrice, "Indian diabetic retinopathy image dataset (idrid)," *IEEE Dataport*, vol. 2, 2018.

[42] A. Tripathi, M. K. Gupta, C. Srivastava, P. Dixit, and S. K. Pandey, "Object detection using yolo: A survey," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I).* Uttar Pradesh, India: IEEE, 2022, pp. 747–752.

[43] L. Wang, L. Feng, J. Wang, J. Li, H. Li, F. Zeng, L. Sun *et al.*, "A variable-clustering-based feature selection to improve positive and negative discrimination of p53 protein in colorectal cancer patients," *Computational and Mathematical Methods in Medicine*, vol. 2022, p. 9261713, 2022.

[44] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, pp. 1–13, 2020.

[45] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," in *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision.* Taipei, Taiwan: Springer, 2017, pp. 198–213.

[46] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *International symposium on visual computing.* Las Vegas, NV, USA: Springer, 2016, pp. 234–244.

Zahraa Jabbar Hussein is a doctoral student under the supervision of Dr. Enas Hamood, obtained a Bachelor's Degree in Computer Science from the University of Babylon, College of Science for Women, Computer Department in 2007, and then got a Master's Degree from University Kebangsaan Malaysia (UKM) Faculty of Information Technology, Computer Since Department in 2013. She is currently working towards her Ph.D. in Computer Science Department, Faculty of Computer Science and Mathematics, University of Kufa, Iraq. Do as lecturer at the University of Babylon/College of Science for Women/Computer Department Since 2008 until now. Her research interests include Artificial Intelligence, Networking and Data Mining.

Enas Hamood Al-Saadi obtained a computer science bachelor's degree, a master's degree, and a doctorate from Babylon University. She teaches as an associate professor at Babylon University's College of Education. Her areas of interest in study are biomedical, computational intelligence, voice processing, image processing, and steganography. She has conducted numerous studies in the field of computer science.