Pertanika J. Sci. & Technol. 34 (2): 889 - 911 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

OD-SAM: Automated Zero-shot Segment Anything Model for Optic Disc Segmentation

Bhuvaneswari S* and Subashini P

Department of Computer Science, Centre for Machine Learning and Intelligence, Avinashilingam Institute for Home Science and Higher Education for Women, 641043 Coimbatore, Tamil Nadu, India

ABSTRACT

The severe stage of diabetic retinopathy (DR) is often identified by the presence of hard exudates on ophthalmologic examination. For precisely detecting exudates, Optic Disc (OD) segmentation is significant due to the high similarity between exudates and the OD. The primary objective of this OD-SAM study is to examine the feasibility of a zero-shot framework for OD segmentation, intending to improve subsequent hard exudate detection in retinopathy screening. The proposed work integrates automatic OD localisation using the peak-end thresholding approach, prompt-based optic disc segmentation using the segment anything model (SAM), a multi-criteria decision-making (MCDM) approach for optimal OD mask selection, and ellipse fitting to smooth the disc boundary. Subsequently, the OD is removed to evaluate its impact on YOLO-based hard exudate detection. The efficiency of this proposed framework is tested on the IDRiD dataset. It achieved an 86.2% overlap and a 90.7% dice coefficient. The results show the effectiveness of the proposed approach for OD segmentation and highlight improvements in exudate detection after OD removal. This study can support lesion analysis and grading in an automated diabetic retinopathy screening and severity assessment.

Keywords: Deep learning, diabetic retinopathy, hard exudate, localisation, optic disc, segment anything model, segmentation

ARTICLE INFO

Article history:
Received: 21 July 2025
Accepted: 17 March 2026
Published: 17 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.13

E-mail addresses:
bhuvaneswari_cs@avinuty.ac.in (Bhuvaneswari S)
subashini_cs@avinuty.ac.in (Subashini P)
*Corresponding author

INTRODUCTION

Diabetes can affect various parts of the body and cause serious complications, including diabetic retinopathy (Halder et al., 2020), which affects the retina of the eye. This condition weakens the light-sensing cells and breaks the connections between retinal cells, damaging blood vessels and

e-ISSN: 2231-8526

breaking down the protective layer. This serious condition leads to vision impairment (Antonetti et al., 2021). However, with timely diagnosis and treatment, the chance of developing vision loss can be reduced (Valverde et al., 2016). Retinopathy screening using a retinal fundus image is a time-saving and cost-effective approach (Teoh et al., 2023). The severity of retinopathy is classified as non-proliferative (NPDR) or proliferative diabetic retinopathy (PDR) (Usman Akram et al., 2014), depending on the extent of retinal lesions, microaneurysms, haemorrhages, hard and soft exudates, and cotton wool spots. Exudates and cotton-wool spots are identified as bright lesions, whereas haemorrhages and microaneurysms are identified as dark lesions (Kumar et al., 2020). Therefore, the type and location of the lesions play an important role in grading.

Hard exudates are considered an important indicator among the factors used to assess disease severity (Liu et al., 2025). The International Clinical Diabetic Retinopathy (ICDR) scale also includes a factor of hard exudates to indicate the severity of the disease (Cleland, 2023). An automated screening system should be developed that can effectively identify hard exudates to help diagnose the disease and initiate treatment at the right time (Joshi & Karule, 2018). Since the brightness of OD is almost identical to that of hard exudate, it is very important to remove OD before detecting exudate (Theera-Umpon et al., 2019). Figure 1 shows that OD can be misidentified due to its similar appearance while identifying exudates. Therefore, removing the OD is very helpful in increasing the accuracy of hard exudate identification (Fu et al., 2023).

There are numerous techniques for segmenting OD from fundus images. However, OD is not easily identifiable due to its low contrast relative to the retinal region, its variability in appearance, and its blood vessels, making it extremely difficult to automatically determine the disc's boundaries at the pixel level (Motevali et al., 2021). Image segmentation is crucial for identifying regions and object edges within an image. However, effective segmentation requires suitable image preprocessing, which simplifies the data, removes unwanted data,

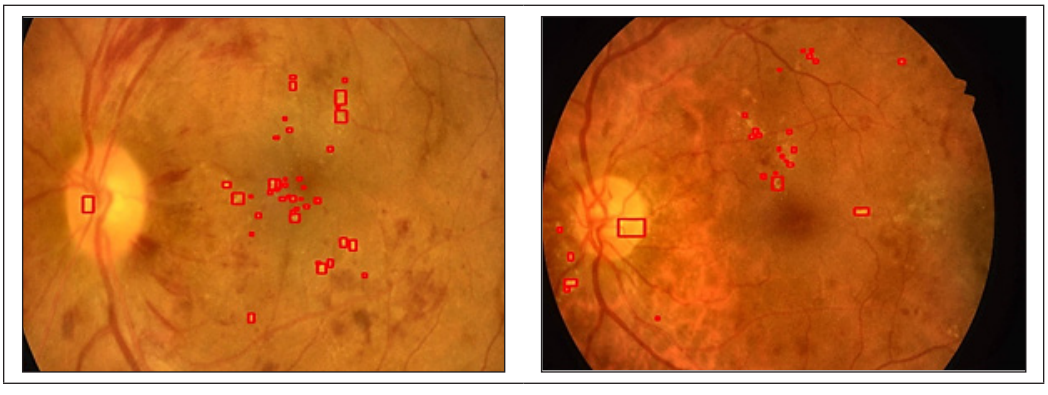

Figure 1. False positive prediction of hard exudate

retrieves important information, and creates an annotated dataset, which is crucial for deep learning-based image segmentation (Chen et al., 2024). Nevertheless, the effectiveness of supervised learning models is currently constrained by the limited availability of manually labelled image data (Hasan et al., 2021).

Inspired by the zero-shot segmentation method, in the present work, the SAM (Kirillov et al., 2023) is proposed to segment the OD from fundus images to address the lack of an annotated dataset and improve OD region segmentation. The procedure for the work carried out is given in Figure 2.

The following are the study's primary contributions:

- The segment anything model is proposed in this study to overcome the limitations of sparsely represented datasets and improve OD segmentation. It performs based on the vision transformer for segmenting the regions.
- SAM was originally developed for problems in real-world image segmentation. The multi-criteria decision-making (MCDM) technique has been used to apply this SAM approach to medical image segmentation. The goal of this adaptation is to address the challenges of both over- and under-segmentation.
- The proposed work integrates the shape-based ellipse fitting (Stojmenovic & Nayak, 2007) approach to enhance the OD boundary. Ellipse fitting helps smooth jagged boundary edges caused by imperfections in the segmentation process.

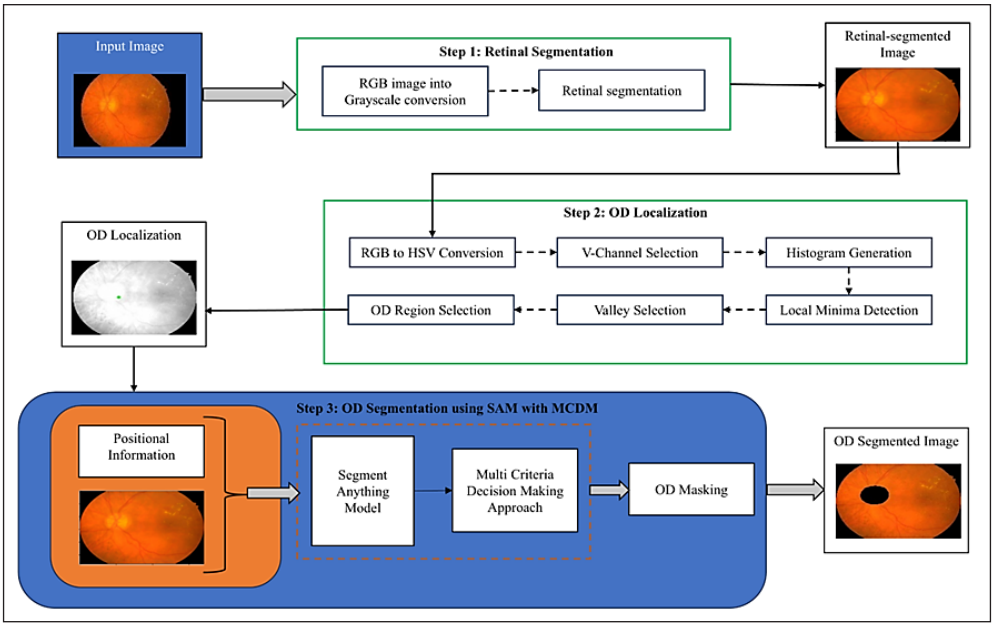

Figure 2. The general pipeline of the work

This paper is structured as follows. Section 2 contains the literature based on OD segmentation. The materials and methods used for this study are given in Section 3. Section 4 presents the experimental results obtained from this study and the discussions. Section 5 contains the conclusion.

RELATED WORK

This section describes works that support OD segmentation using traditional image processing approaches and deep learning techniques.

Traditional Image Processing Approaches for OD Segmentation

In their study, Dharmawan et al. (2020) proposed a new optic disc localiser method. The template-based, map-based vessel density and maximum entropy-based methods have been used to find the centre of the optic disc. Additionally, the modified Dolph-Chebyshev type I function is designed to detect boundary conditions. Finally, circular Hough transform, polar transform, and B-spline approximation are used to obtain the final OD boundary. Testing the proposed approach on the MESSIDOR and DRIVE datasets yielded overlap accuracies of 89.5% and 87.3%, respectively. The proposed method is robust to variations in image quality. But the fixed template match may not adapt well to significant variance in OD appearance. The paper (Ramani & Shanthamalar, 2020) presents an improved image-processing algorithm for automated optic disc segmentation to detect glaucoma early. For this purpose, an improved region-based pixel density calculation method for localisation, and circular Hough transform with Hough peak and red channel superpixel segmentation methods have been proposed to segment the optic disc. The model achieved segmentation accuracies of 99.2% to 99.73% across datasets. The proposed method is tested across multiple datasets and is computationally fast. Even though it is less flexible compared to DL methods. A study by Dietter et al. (2019) proposed a new paradigm to detect and segment the optic disc, as it is a key indicator in retinal image analysis and retinal disease detection. To deal with the strong technical artefacts during OD detection, a two-stage methodology has been proposed. First, the vessel's orientation and brightness are determined to locate the vessel. The modified score function was then used to detect the best OD border. Experimental results show that the best segmentation accuracy is across different datasets. The proposed method is robust against artefacts, but the segmentation performs poorly on distorted vessel structures. Abdullah et al. (2020) presented a new approach for optical disc localisation and segmentation. In their approach, the median filter and Otsu have been used to accurately determine the OD location. The red channel has then been used to remove the veins, and morphological operations have been applied to reduce noise. From the resulting optic disc region, fuzzy clustering, the mean-shift algorithm, the active contour model, and ellipse fitting were used to segment the disc.

The methods on the DRIVE, DIRATEDB1, DRIONS-DB, and STARE datasets achieved segmentation accuracies of 84.3%, 84.56%, 81.01%, and 78.35%, respectively. The proposed method helps handle noise and structural irregularities. Due to severe pathologies and poor image contrast, segmentation accuracy was reduced. Since OD is an indicator for diagnosing the macula and is very important in many analyses of retinal disease, Xue et al. (2022) have proposed a new hybrid level set model to segment the optic disc. For this purpose, the area-based term has been used to detect the average pixel value difference between the inside and outside of a contour. The shape-based term has also been used to identify the prior shape model and the contour. Experimental results showed that the proposed method achieved an overlap accuracy of 92.75% and 81.79% on the DRISHTI-GS and TMUEH datasets, respectively. This approach is robust to handle parapapillary atrophy (PPA) and bright artefacts. The proposed method heavily depends on the preprocessing quality and non-deep learning methods.

Deep Learning Approaches for OD Segmentation

Semi-supervised and transfer learning are proposed by Bengani et al. (2020) for automated OD segmentation in glaucoma detection. The unlabelled retinal images were used for convolutional autoencoder training, and the trained network was converted to a segmentation network. This network was then trained on retinal images using the OD ground-truth image, leveraging transfer learning. The segmentation study on the DRISHTI GS1 and RIM-ONE datasets reported a Jaccard score of 93.14% and 88.24%, respectively. The proposed method is effective with limited labelled data. But it requires a high convolutional autoencoder (CAE) training. Septiarini et al. (2023) have proposed to automatically segment the optic disc needed to diagnose glaucoma. To do this, they used two convolutional neural network architectures: the single-shot multibox detector and MobileNetV2 to segment the OD area. Before this, preprocessing such as augmentation, normalisation, and resizing was applied. Experimental results indicated that the proposed method is robust in handling diverse image quality and pathology. It achieved an overlap accuracy of 97.63% on the private dataset and 97.12% on the REFUGE dataset, but it suffers from over-segmentation with PPA and requires high computational time for large datasets. Sevastopolsky (2017) presented the modified U-Net convolutional model for automated optic disc and cup segmentation for optic nerve head examination. This modified model has fewer filters, allowing it to be lightweight. The approach achieved an overlap accuracy of 89% and 89% on datasets such as DRIONS-DB and RIM-ONE v.3. It was tested across multiple public datasets. However, the optic cup is highly sensitive, posing a challenge and limitation for this study. In a study by Zhang et al. (2021), a transferable attention U-Net model was proposed for optic disc and cup segmentation in cup-to-disc ratio-based glaucoma detection. Two discriminators and attention modules have been

applied to locate and extract invariant features from the datasets. The method used an attention mechanism and an adversarial training method to address the domain adaptation problem. Experimental results reported a higher dice score across datasets. The proposed approach is robust in handling different camera resolutions, contrasts, and field-of-view, even though it requires a labelled source domain and more complex adversarial training.

Yin et al. (2021) have proposed a level-set-based deep learning model to perform OD and OC segmentation with weak boundaries. This combination uses a level set loss function to achieve smooth, segmented boundaries and consistent regions. To reduce ambiguity, joint region-based CNNs, which assume the disc and cup are approximately elliptical, and M-Net knowledge, which shows that the optic cup is always inside the OD, are used to improve segmentation. It reports superior performance consistently. The usage of two separate computationally intensive networks is a limitation of the study.

From the literature study, it has been observed that the majority of optic disc segmentation studies have largely focused on glaucoma detection because the diagnosis of glaucoma is particularly dependent on the unique characteristics of the OD region. OD segmentation is also vital for diabetic retinopathy tasks to prevent false positives when recognising exudates. Thus, the current study seeks to bridge this gap in hard exudate detection in diabetic retinopathy, particularly in the context of OD localisation and segmentation. Furthermore, this work presents an effective solution to eliminate reliance on templates and reduce the need for large, annotated datasets, which are commonly encountered in existing image processing studies and in most deep learning techniques for OD segmentation.

MATERIALS AND PROPOSED METHODOLOGY

This section gives a detailed description of the dataset and the proposed method. SAM is a prompt-driven model. Therefore, it requires prompts to segment objects (Mazurowski et al., 2023). Thus, the study primarily focuses on identifying the optic disc to obtain the location data using traditional image processing techniques. Later, the SAM model uses the obtained location data and the input image to separate the OD from the retinal images.

Dataset Collection

The study obtained 81 colour fundus photographs from the publicly available IDRiD benchmark dataset (Porwal et al., 2018). For the segmentation study, the dataset gives 54 images for training purposes and 27 images for testing purposes with their ground-truth images. Most other popular datasets for diagnosing diabetic retinopathy usually give each image a severity grade or image-level annotation but lack precise annotation information (Chincholi & Koestler, 2023). However, the IDRiD database provides both ground-truth images and grading information, all of which were assessed by medical professionals.

Retinal Segmentation

Before OD localisation, the retinal region is segmented from the fundus images because images from the IDRiD database contain unwanted background information. Thus, the bright retinal region is segmented from the background to assist the model in concentrating on the important information in the foreground. Then, the segmented retinal input image (I_{BGR}) is converted into the HSV colour space (I_{HSV}) to isolate the Value (V) channel image for further analysis, as brightness information is more viable in the V channel (Nagpal et al., 2024). The splitting can be expressed as shown in Equation 1: -

$$(I_H,\ I_S,\ I_V) = \text{split}(I_{HSV}) \quad [1]$$

Optic Disc Localisation using Peak-end Thresholding Approach

Following retinal segmentation, the peak-end thresholding method is used on v-channel images for locating the optic disc. The goal of the OD localisation is to provide a reference point for the OD, which is then used to subdivide the OD area with the SAM. Several studies investigate the use of retinal vessels for optic disc localisation. However, vessel segmentation-based methods can lead to increased computational complexity in the OD localisation process (Kamble et al., 2017). For OD localisation, this work used a peak-end thresholding technique that combined connected component analysis with an inverted histogram. Figure 3 shows the OD localisation approach.

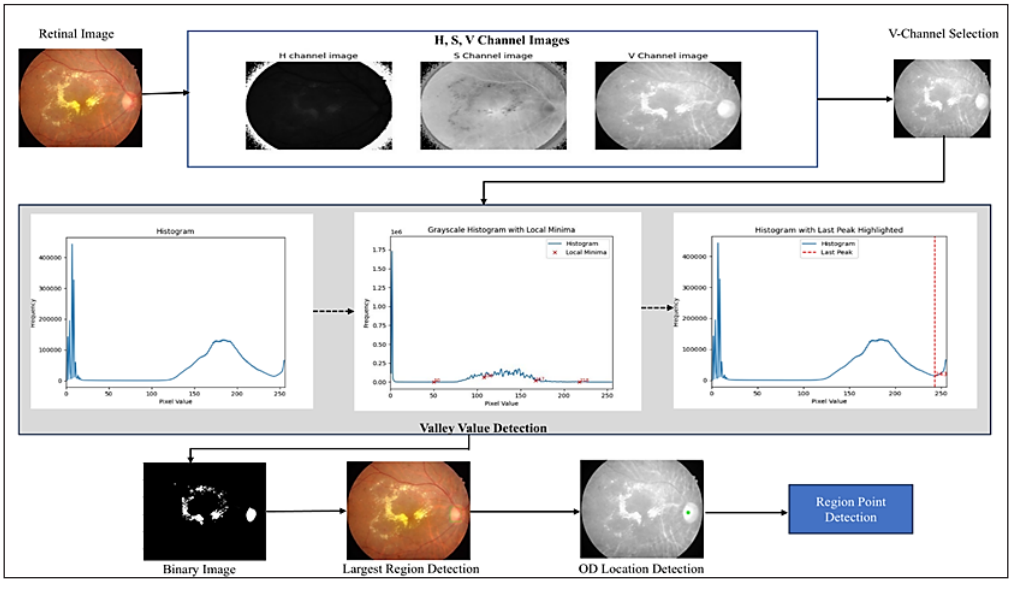

Figure 3. OD localisation

The OD is generally regarded as a bright area within the image. Histogram inversion and local minima detection techniques help identify the bright area associated with the optic disc in retinal images. To achieve this, a technique that involves detecting local minima in the histogram of the image is employed. The histogram inversion transforms the bright region of the OD into a local minimum, allowing the positions in the inverted histogram to be located. The detected minima are separated by a minimum distance to avoid false detections of closely spaced features. Once the local minima are identified, the value corresponding to the most significant minimum, typically the last detected peak, is used for further processing. Subsequently, this threshold value is employed in binary segmentation to isolate the bright OD region.

This approach, however, can sometimes yield false positives, as it may also capture high-intensity exudates. To accurately locate the OD region and distinguish it from other bright features, the largest connected component analysis is used. Because typically OD exhibits a larger area compared to other features. This approach effectively identifies the OD by assessing the size and shape of the connected components. Then, the positional information of an OD is obtained from the largest component region. In some situations, misvocalisation is occasionally possible due to large exudate clusters.

Proposed Optic Disc Segmentation using SAM

This work utilised a zero-shot learning SAM method for OD segmentation. Typically, the SAM prompt encoder, which provides positional encodings, requires real-time interactions. However, in this study, the process is automated through the prior OD localisation approach to eliminate the need for manual intervention (Guan et al., 2025). The optic disc usually appears bright, but this is not constant throughout. Due to factors such as differences in illumination, blood vessel darkness, or the fundamental anatomic structure of the optic disc, certain regions may look brighter than others. Effectively recognising and segmenting the optic disc might be difficult due to the uneven brightness in image processing. Therefore, a zero-shot learning SAM is proposed to segment OD from retinal images, ensuring effective region detection and segmentation, even when some regions are less bright than others.

Network Architecture

This study used the promptable (Misera et al., 2026) segmentation network (SAM), built on transformer-based vision models. The vision transformer serves as a backbone network in the SAM to extract features from the image. The model generates feature maps at multiple scales using a Feature Pyramid Network (FPN) to identify objects at different levels and their boundaries. The transformer-based design uses contextual information from the input image to improve the segmentation result. To do this, the SAM model is integrated with a prompt encoder that effectively extracts and combines the input image's spatial and

semantic elements to process positional encoding along with free-form text and encoding specific elements, a Masked Auto-encoder (MAE), and an image encoder that transforms the input image into image embeddings.

Enhanced transformer decoder blocks help to focus on significant areas within the inputs via self-attention and cross-attention mechanisms. To enhance segmentation precision, the model combines information from both the prompts and the image by utilising the enhanced transformer decoder blocks. The self-attention enables a focus on the most important image parts, and cross-attention helps to establish an interface between input embeddings and prompt embeddings. After the self-attention and cross-attention processes have been carried out, the generated embeddings are processed by the mask decoder. The image embeddings are then up-sampled by the model to provide a final mask with the same resolution as the original image. This up-sampled embedding and output token are used by the dynamic mask prediction head to predict the mask's probability. On the other hand, a single prompt presents many masks, such as the object's part, a smaller sub-part, and the entire object. For that, the model evaluates the generated masks according to a confidence score using overlap measures. Utilising the most appropriate and precise mask, it reduces the chance of inaccurate segmentation. The overview of the segmentation process using the SAM model is illustrated in Figure 4a.

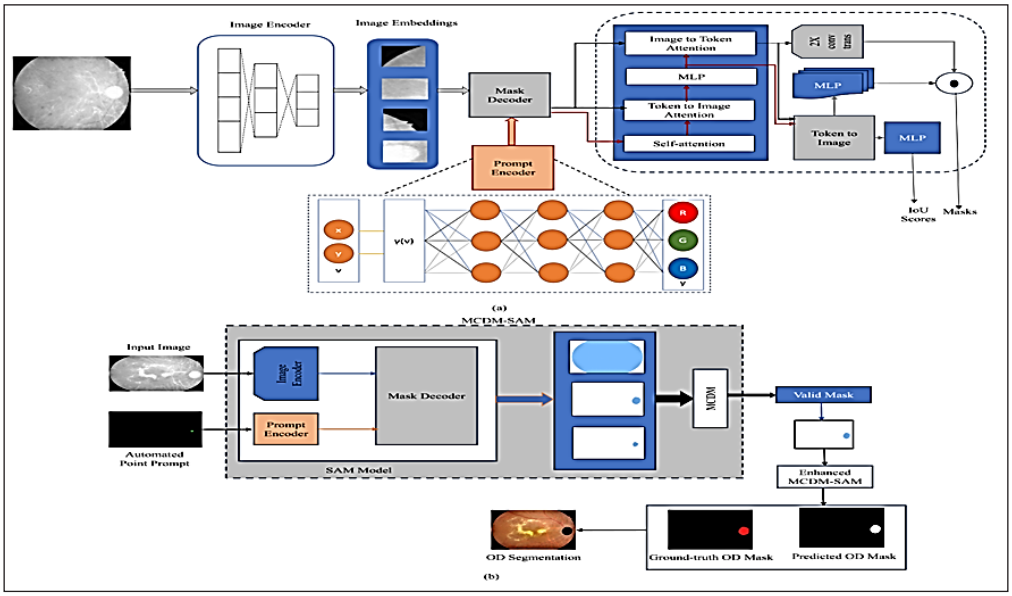

Figure 4. (a) Illustration of the OD segmentation process in the SAM architecture; (b) OD segmentation in the MCDM-SAM

Optic Disc Segmentation

After the above-mentioned retinal segmentation and OD localisation process, the image is converted into the desired colour channel, and then the converted image is fed into the proposed SAM model with the positional information. Positional information obtained from the OD localisation is assigned to the SAM model as a point prompt. Although the models used in this work are well-established, the current study focuses on their innovative usage in retinal segmentation, particularly in medical image segmentation. Figure 4b provides an overview of the suggested OD segmentation method. This proposed approach ensures precise and efficient segmentation, streamlining the process without requiring real-time human input.

The SAM model, which is initialised with the ViT-B, is given the input image. The image encoder of the SAM is specially designed with the Vision Transformer (ViT) to handle more complex and volumetric data, and it is scaled from ViT-B (Base) to ViT-H (Huge). In comparison with ViT-B, ViT-H can capture more significant and detailed features. Since it modestly improves accuracy and substantially raises processing requirements (Kirillov et al., 2023). Therefore, for computational efficiency, the ViT-B model was used as the image encoder in this study, although ViT-L and ViT-H also exhibit considerable performance. After processing the given V-channel image, the pre-trained Vision Transformer with the MAE approach He et al. (2022) creates a 64x64 embedding that preserves rich, downscaled information from the high-resolution input. SAM needs positional information in addition to the image to generate a mask on it. To analyse the positional data, SAM provides a prompt encoder.

The prompt encoder supports real-time user interaction. Instead of real-time interactions, the points that are identified automatically during the OD localisation process are transferred to the prompt encoder in the present study. These predetermined points, encoded by their positions and associated foreground labels, will serve as prompts to guide the segmentation process. The prompt encoder will translate these fixed coordinates into embeddings that precisely match the image's features accordingly. The mask decoder will predict the final segmentation mask using a transformer-based architecture that integrates the prompt token and image embeddings. The model generates multiple segmentation masks to capture different interpretations and address an issue in a single prompt. During interpretation, the model usually produces three optimal masks. The model ranks predicted masks by their Intersection over Union (IoU) with the object and evaluates the loss for each to obtain the most accurate prediction.

However, as shown in Figure 5, the successful SAM segmentation is not always granted to the maximum accuracy mask. Because the original SAM structure was designed for the segmentation of 2-D natural images. As a result, when evaluated with medical images, the model exhibits performance degradation (Gong et al., 2023). Therefore, to improve

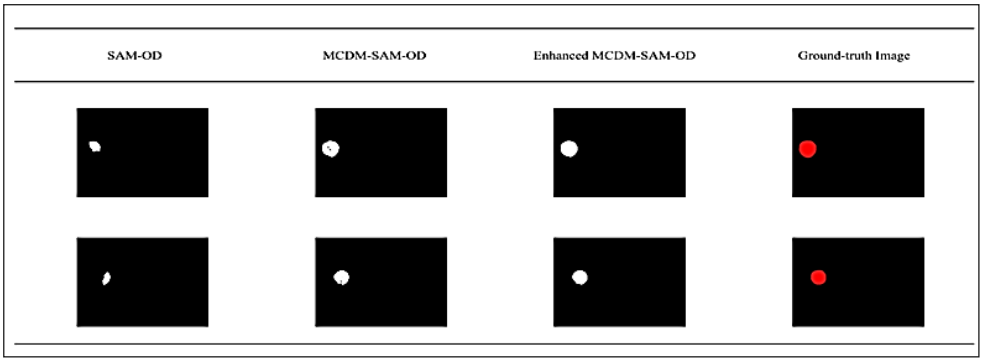

Figure 5. Results of OD segmentation: columns (1) using SAM; (2) SAM with MCDM; (3) enhanced MCDM-SAM; and (4) ground-truth image

optic disc segmentation, this study suggests using the Multiple Criteria Decision Making (MCDM) technique with the SAM (MCDM-SAM). MCDM aids in choosing the optimal solution among several choices for making informed decisions when multiple criteria are involved (Ali & Broumi, 2024; Barretta et al., 2023; Kim et al., 2022). For each possible segmentation to produce more accurate OD segmentation, the MCDM considers confidence to choose segments with a higher probability and area score to compare each segment to a threshold range. The confidence scores help evaluate the validity of each segment, and the area-based approach ensures the identified segment region is neither too large nor too small. To reflect the anatomical size range of the OD, these area scores were determined empirically through validation experiments on the training dataset. This integrated criterion generates a superior and well-conformed OD segmentation.

As seen in Figure 5(2), determining boundaries can be challenging due to the presence of blood vessels and poor contrast. Thus, this work uses ellipse fitting (Pallawala et al., 2004) along with the MCDM-SAM technique to streamline the detection process and capture the essence of the OD boundary. To find the best-fitting ellipse, the algorithm uses the least squares techniques. It determines the ellipse that best fits a collection of 2D points, such as the edges of an object. It returns a rotated rectangle that completely contains the fitted ellipse. This rectangle is useful because it shows the position and orientation of the ellipse in relation to the points and ensures that the fitting is done effectively. It helps minimise the difference between the actual points and the ellipse. Finally, the ellipse-fitting algorithm can accurately place an ellipse around the object's edges. This Enhanced MCDM-SAM ensures more representation of actual object shapes. The general steps for the enhanced MCDM-SAM segmentation algorithm are as follows:

Input: Preprocessed image (obtained using Equation 1) for segmentation, points, and label for mask prediction, and area constraints for mask selection.

Output: Optic disc segmented image.

Step-by-step Process:

Step 1: Input the V-channel image and apply the median blur to reduce noise. The median blur operation can be represented in the following equation.

$$I' = \text{medianBlur}\left(I_{v,}\ k = 15\right) \quad [2]$$

where I_v represents the v-channel input image, I′ denotes the smoothed output image, and k indicates the kernel value with size 15.

Step 2: Load the segmentation predictor model and set the input image in the predictor.

$$sam = \text{sam\_model\_registry}[model\_type]\ \ (\text{checkpoint} = path) \quad [3]$$

$$Pred = \text{SamPredictor}(sam) \quad [4]$$

Step 3: Generate segmentation masks using points and labels of an image.

$$m, s, l = \text{Pred}(sam).\text{predict}(I_{Pt}, I_l, mm\_O = True) \quad [5]$$

where, m- masked images, s-score of the masked images, l-logits values, I_{pt}- Points of an input image, I_l- Labels of an input image, mm_O -Multi-mask output mode.

Step 4: Compute the pixel area on the generated segmentation mask image with its confidence score.

$$A_i = \sum_{x,y} M_i(x, y) \quad [6]$$

Where M_i- masked images, A_i- Area of the i-th mask

Step 5:

a. Sort the masked images based on their confidence scores.

$$S = \text{argsort}(s)[::-1] \quad [7]$$

Here, S contains indices sorted by descending score.

b. Apply criteria with area constraints for selection. If the first condition is met, it selects the highest-scoring mask as the final one. If the first condition is not satisfied,

it selects the second-high score mask. Otherwise, it selects the third-smallest-area mask, or else the second-highest-score mask as the final one.

$$m_{final} = \begin{cases} m_{idx[0]}, & \text{if } A_{min} \leq A_{idx[0]} \leq A_{max[1]} \quad [8] \\ m_{idx[1]}, & \text{if } A_{idx[1]} \leq A_{max[2]} \quad [9] \\ m_{idx[2]}, & \text{otherwise (fallback)} \quad [10] \end{cases}$$

Here, A_{min} represents the minimum area threshold, $A_{max[1]}$ is the second maximum area for the high-score mask, and $A_{max[2]}$ represents the maximum area for other masks. From this, we get the final selected mask image that satisfies the conditions.

Step 6: Apply the ellipse fitting on the selected mask image. The ellipse-fitting approximates the ellipse parameters to minimise the least-squares distance between the ellipse (E) and the set of contour points (P). This fitting process can be expressed as follows:

$$E = \text{fitEllipse}(P) \quad [11]$$

As depicted in Figure 6, the following gives a comprehensive explanation of each step of the OD segmentation process:

- The original image (Figure 6a) is assigned to the image encoder to transform it into an image embedding after being transformed into the desired colour channel (Figure 6b).
- To obtain the desired region, the prompt encoder is provided with the coordinate point that was previously found (Figure 6c).
- The predicted probability of masks (Figure 6d, Figure 6e, Figure 6f) is obtained.
- Based on the multiple criteria, the enhanced optic disc segmentation result is extracted (Figure 6g, Figure 6h).
- The optic disc boundary is enhanced using the ellipse fitting (Figure 6i) on the extracted OD image.

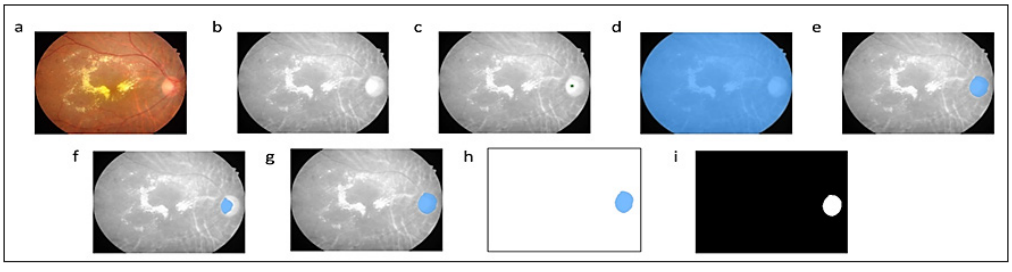

Figure 6. Results of each process of OD segmentation: (a) input image; (b) V-channel image; (c) positional information; (d-f) different mask probabilities; (g-h) final masks using MCDM-SAM; and (i) enhanced OD boundary

RESULTS AND DISCUSSION

Performance Measurement

The efficiency of the proposed OD segmentation approach is assessed using several evaluation metrics, such as precision, recall, sensitivity, F1 score, IoU (Jaccard score), and dice coefficient (DICE), relative to the ground truth. (Sarvamangala & Kulkarni, 2022). Accuracy is the number of accurate predictions among all predictions, used to assess overall correctness. Recall measures the number of actual positives among all actual positives, indicating how well the model detects real positive occurrences. Precision measures the accuracy of positive predictions by dividing the number of actual positives by the total number of predicted positives. The F1-score is perfect for unbalanced data since it balances recall and precision as their harmonic mean. This scale provides a single measurement of false negatives and positives. The Intersection over Union (IoU) is an important metric in segmentation work because it quantifies the overlap between the predicted and ground-truth areas. In the medical field, both IoU and Dice coefficient (DICE) are widely used to determine the degree of specificity of the segmentation methods (Furtado, 2020).

$$\text{Accuracy} = \frac{T_P + T_N}{T} \quad [12]$$

$$\text{Recall}(T_{PR}) = \frac{T_P}{F_N + T_P} \quad [13]$$

$$\text{Precision } (P_{Pv}) = \frac{T_P}{F_P + T_P} \quad [14]$$

$$\text{F1 score} = 2 \; X \; \frac{P_{PV} \; X \; T_{PR}}{P_{PV} + T_{PR}} \quad [15]$$

$$\text{IoU} = \frac{T_P}{T_P + F_P + F_N} \quad [16]$$

Here, T_P and T_N represent correctly classified positive and negative results, respectively. Similarly, F_P and F_N represent false-positive and false-negative events. T denotes the total number of pixel-level predictions evaluated, T_{PR} represents the true positive rate, and P_{PV} represents the positive predictive value.

Experimental Analysis using a Deep Learning Segmentation Approach

In the experiment, the effectiveness of the proposed method is compared with the current state-of-the-art (Rayed et al., 2024) deep learning-based models, such as a fully convolutional network (FCN) (Shelhamer et al., 2014), U-Net (Ronneberger et al., 2015), and DeepLab (Chen et al., 2018). The parameter settings used in this experiment are given in Table 1.

Although FCN and DeepLab achieved adequate precision and recall, their low segment quality is reflected in their F1 Scores and IoUs. The proposed enhanced MCDM-SAM method performs better than U-Net in all aspects of recall, F1-score, and IoU. This result shows that the proposed method is superior to all others in terms of overlap and segmentation accuracy. SAM and MCDM-SAM are similar in terms of precision and recall. The proposed Enhanced MCDM-SAM approach offers the highest level of accuracy and reliability for segmentation tasks, achieving the best possible recall, F1 score, and IoU. The results are presented in Table 2.

According to Table 2, the proposed approach consistently outperforms other approaches across all key measures. It achieves the best balance between precision and recall, resulting in the highest F1 score and IoU. FCN gives high accuracy, but extremely low recall due to class imbalance at the pixel level.

Table 1
Implementation setup details of baseline models

Model	Framework	Model Availability?	Pre-trained	Max iteration	Batch size
FCN	TensorFlow	Yes	ImageNet	25	1
U-Net	TensorFlow	Yes	ImageNet	25	1
DeepLab	TensorFlow	Yes	ImageNet	25	1

Table 2
Comparative analysis with the state-of-the-art deep learning-based segmentation approach

Methods	Accuracy	Precision	Recall (Sensitivity)	Specificity	F1 score/ DICE	IoU/ Jaccard
FCN	0.869	0.984	0.055	0.999	0.130	0.072
DeepLab	0.972	0.429	0.763	0.977	0.314	0.205
U-Net	0.995	0.927	0.857	0.998	0.696	0.544
SAM	0.965	0.948	0.838	0.968	0.853	0.789
MCDM-SAM	0.983	0.948	0.890	0.985	0.895	0.84
Proposed enhanced MCDM-SAM	0.981	0.942	0.916	0.982	0.907	0.862

It emphasises the importance of balanced evaluation metrics, such as DICE/ F1 and IoU. The results demonstrate that FCN is primarily accurate at predicting negatives but fails to identify most actual lesions, as evidenced by low IoU and F1 Scores, indicating subpar performance in OD pixel-level segmentation. DeepLab achieves good accuracy and recall; however, low precision results in a substantial number of false positives. A modest IoU suggests that the segmentation does not closely match the ground truth. The U-Net network delivers high accuracy, good precision, and high recall. The IoU scores indicate that the segmentation quality is acceptable, although not outstanding. The SAM consistently demonstrates high performance across all evaluation metrics. Good image segmentation quality can be inferred from high precision, a high recall, and a strong IoU. The MCDM-SAM enhances SAM by increasing its recall rate while sustaining a high level of precision. An improved IoU indicates high-quality segmentation results. But further enhancement in MCDM-SAM facilitates the best recall for identifying the most lesions. It maintains high precision, so few false positives. Moreover, the best F1 score and IoU indicate the most accurate segmentation overall. The IDRiD dataset images utilised in experimental analysis display a range of clinical conditions, such as blood vessel anomalies, comparable intensity to exudates, and other anomalies. The overlap results for various pathological images obtained using the proposed enhanced MCDM-SAM approach and alternative methods are illustrated in Figure 7. In terms of OD segmentation, the Enhanced MCDM-SAM approach outperforms the other models overall.

Experimental Analysis using Image Processing Approach

To compare the OD segmentation results with traditional image processing, k-means clustering with a template matching algorithm from another paper was reproduced. Table 3 shows the segmentation results assessed using the IoU.

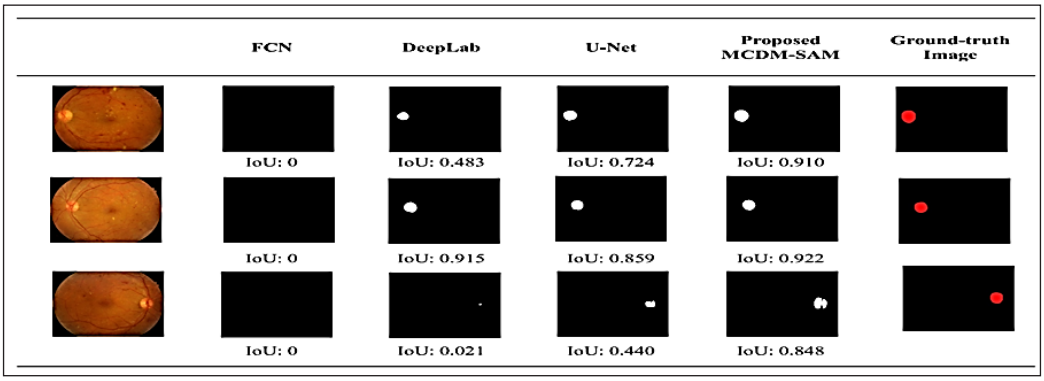

Figure 7. Comparison of overlaps in a few example cases

The table findings indicate that the K-means clustering with template matching method performs moderately across all metrics, with acceptable Recall but lower Precision and overlap metrics (IoU, Dice coefficient). This is because, as Figure 8 illustrates, template matching relies on exact matches to identify the optic disc; it has difficulty in handling variations such as varying lighting conditions or partial occlusions. Conversely, the proposed method significantly outperforms the K-means approach across all metrics. The Enhanced MCDM-SAM is more precise and efficient at segmenting regions of interest, as evidenced by its significantly higher Precision, Recall, F1 Score, IoU, and Dice coefficient. The study's results show that the proposed method performs better in OD segmentation.

To eliminate inter-study variability and assess the efficacy of the proposed OD method, controlled comparisons were conducted using FCN, U-Net, DeepLab, and template matching. This approach has been represented from the early days of FCN through popular U-Net, which is currently well-represented in clinical segmentation, to advanced semantic segmentation (DeepLab), and to foundational models (SAM).

Table 3
Evaluation result with traditional image processing methods

Metrics	Traditional Method (Basu et al., 2021)	Enhanced MCDM-SAM
Precision	0.49	0.94
Recall	0.84	0.92
F1 Score	0.62	0.91
IoU	0.47	0.86
Dice coefficient	0.62	0.91

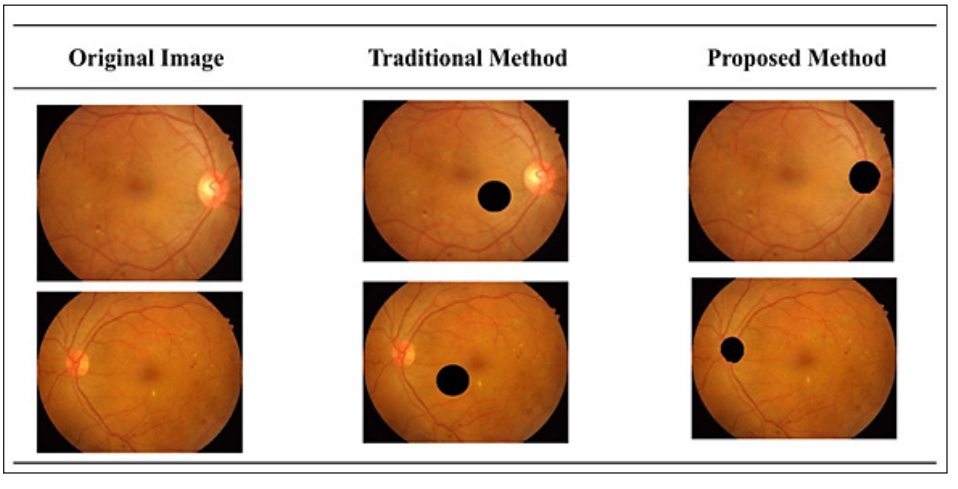

Figure 8. Comparison result of the segmentation

To ensure a fair comparison, the optimised hyperparameters for each model were obtained using the same validation set. Performance was evaluated using precision, recall, F1/DICE score, and IoU on a test dataset consisting of 27 fundus images with expert-verified OD annotations.

Analysis of Detection Results: With and Without OD Segmentation

The impact of OD segmentation on exudate detection was evaluated by utilising retinal images with and without OD segmentation. For this experiment, 54 high-resolution fundus images for training and 27 for validation, each with annotated hard exudate lesions, are used. In the preprocessing stage, the image was converted into the LAB colour space, and the initial enhancement of each image was performed using Contrast Limited Adaptive Histogram Equalisation applied to the L-channel within the LAB colour space. To enhance the model's robustness, the albumentations library was employed for data augmentation, including random scaling, flipping, rotation, and adjustments to brightness and contrast. The detection model was trained for 100 epochs using a YOLO-based framework, with an input image size of 1024×1024 pixels, a batch size of 8, and no model-specific hyperparameter tuning.

The results shown in Table 4 demonstrated improvements in hard exudate detection, specifically in precision from 0.468 to 0.476, recall from 0.314 to 0.344, and mAP50 from 0.310 to 0.332. Excluding the optic disc area appears to reduce the likelihood of false positives, mainly because its appearance and intensity are comparable to those of hard exudates. The overall fitness score increased from 0.157 to 0.165, which further supports the benefits of OD segmentation in DR screening at the lesion level.

From the above-mentioned experiments, the proposed framework was more successful at optic disc segmentation than the traditional segmentation model. In fully supervised deep learning models such as FCN, DeepLab, and U-Net, overfitting often occurs when the dataset has few training images. However, the proposed work utilises a pre-trained zero-shot SAM without task-specific retraining, and MCDM-based selection and ellipse fitting are rule-based post-processing steps. Hence, the risk of overfitting is reduced compared to fully supervised deep learning models. However, this study also has some limitations.

Table 4
Hard exudate test comparison results

Metric	Before OD Segmentation	After OD Segmentation
Precision	0.47	0.48
Recall	0.31	0.34
mAP@0.5	0.31	0.33
Images	130	130

Multi-centre dataset validation is necessary to confirm the model's generalisability, as the proposed work was conducted on a single dataset. Additionally, cross-dataset validation is required to evaluate the OD localisation approach, despite the V-channel being less sensitive to moderate illumination variations.

CONCLUSION

This study proposes an enhanced MCDM-SAM method to segment OD in fundus images while simultaneously addressing the problem of a lack of annotated training data. It uses a predetermined criterion with a wide range of possibilities to accurately determine the optic disc. Elliptical fitting has been used to smooth the OD boundary. The study's results show that the proposed method performs better than other approaches. The proposed OD segmentation approach will help prevent false positives and increase the accuracy of true hard exudate lesion detection. In this study, optic disc segmentation was performed using the proposed framework, and its impact on the detection accuracy of hard exudate lesions was investigated. No other lesions were examined. Therefore, investigation of other lesions would be helpful in grading interpretation. Future work will involve designing an experiment across all lesion types, testing it on larger, more diverse datasets, and evaluating the impact on grading accuracy for diabetic retinopathy. This work is significant for improving optic disc detection and segmentation, thereby supporting more reliable exudate detection.

ACKNOWLEDGEMENT

This work was supported by the Centre for Machine Learning and Intelligence (CMLI), an ISO-certified centre (ISO/IEC 20000-1:2018), funded by the Department of Science and Technology (DST-CURIE).

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF COMPETING INTEREST

The authors report there are no competing interests to declare.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used QuillBot and Grammarly to improve language and readability. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the publication.

DATA AVAILABILITY

The data supporting the findings of this study are openly available in the Indian Diabetic Retinopathy Image Dataset (IDRiD) at https://dx.doi.org/10.21227/H25W98.

CODE AVAILABILITY

The code supporting the findings of this study is available from the corresponding author upon reasonable request.

REFERENCES

Abdullah, A. S., Rahebi, J., Özok, Y. E., & Aljanabi, M. (2020). A new and effective method for human retina optic disc segmentation with fuzzy clustering method based on active contour model. *Medical and Biological Engineering and Computing, 58*(1), 25-37. https://doi.org/10.1007/s11517-019-02032-8

Ali, A. M., & Broumi, S. (2024). Machine learning with a multi-criteria decision-making model for thyroid disease prediction and analysis. *Multicriteria Algorithms with Applications, 2*, 80-88. https://doi.org/10.61356/j.mawa.2024.26961

Antonetti, D. A., Silva, P. S., & Stitt, A. W. (2021). Current understanding of the molecular and cellular pathology of diabetic retinopathy. *Nature Reviews Endocrinology, 17*(4), 195-206. https://doi.org/10.1038/s41574-020-00451-4

Barretta, R., Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia, 3*(1), 77-87. https://doi.org/10.3390/encyclopedia3010006

Basu, S., Mukherjee, S., Bhattacharya, A., & Sen, A. (2021). Segmentation of blood vessels, optic disc localisation, detection of exudates, and diabetic retinopathy diagnosis from digital fundus images. In *Proceedings of the International Conference on Intelligent Computing and Communication (ICICC 2020)* (pp. 173-184). Springer. https://doi.org/10.1007/978-981-16-1543-6_16

Bengani, S., Angel Arul Jothi, J., & Vadivel, S. (2020). Automatic segmentation of optic disc in retinal fundus images using semi-supervised deep learning. *Multimedia Tools and Applications, 80*(3), 3443-3468. https://doi.org/10.1007/s11042-020-09778-6

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, Y., Bai, Y., & Zhang, Y. (2024). Optic disc and cup segmentation for glaucoma detection using attention U-Net incorporating residual mechanism. *PeerJ Computer Science, 10*, Article e1941. https://doi.org/10.7717/peerj-cs.1941

Chincholi, F., & Koestler, H. (2023). Detectron2 for lesion detection in diabetic retinopathy. *Algorithms, 16*(3), Article 147. https://doi.org/10.3390/a16030147

Cleland, C. (2023). Comparing the International Clinical Diabetic Retinopathy (ICDR) severity scale. *Community Eye Health, 36*(119), 10. https://pmc.ncbi.nlm.nih.gov/articles/PMC10436766/

Dharmawan, D. A., Ng, B. P., & Rahardja, S. (2020). A new optic disc segmentation method using a modified Dolph-Chebyshev matched filter. *Biomedical Signal Processing and Control, 59*, Article 101932. https://doi.org/10.1016/j.bspc.2020.101932

Dietter, J., Haq, W., Ivanov, I. V., Norrenberg, L. A., Völker, M., Dynowski, M., Röck, D., Ziemssen, F., Leitritz, M. A., & Ueffing, M. (2019). Optic disc detection in the presence of strong technical artefacts. *Biomedical Signal Processing and Control, 53*, Article 101535. https://doi.org/10.1016/j.bspc.2019.04.012

Fu, Y., Zhang, G., Lu, X., Wu, H., & Zhang, D. (2023). RMCA U-Net: Hard exudates segmentation for retinal fundus images. *Expert Systems with Applications, 234*, Article 120987. https://doi.org/10.1016/j.eswa.2023.120987

Furtado, P. (2020). Segmentation of diabetic retinopathy lesions by deep learning: Achievements and limitations. In *Proceedings of the 7th International Conference on Bioimaging (BIOIMAGING 2020)* (pp. 95-101). https://doi.org/10.5220/0008881100002513

Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.-A., & Dou, Q. (2023). 3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable medical image segmentation. *arXiv*. https://doi.org/10.1016/j.media.2024.103324

Guan, G., Zhao, C., Li, Z., Zhang, P., Chen, Y., Ye, P., Wong, M. K., Chan, L. Y., Yan, H., Tang, C., & Zhao, Z. (2025). EmbSAM: Cell boundary localisation and Segment Anything Model for fast images of developing embryos. *Communications Biology, 9*(1), Article 8. https://doi.org/10.1038/s42003-025-09220-3

Halder, N., Roy, D., Banerjee, R., Roy, P., Sarkar, P. P., & Bandyopadhyay, S. (2020). Automatic detection and segmentation of optic disc (ADSO) of retinal fundus images based on mathematical morphology. In *Proceedings of the 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA 2020)*. https://doi.org/10.1109/NCETSTEA48365.2020.9119931

Hasan, M. K., Alam, M. A., Elahi, M. T. E., Roy, S., & Martí, R. (2021). DRNet: Segmentation and localisation of optic disc and fovea from diabetic retinopathy image. *Artificial Intelligence in Medicine, 111*, Article 102001. https://doi.org/10.1016/j.artmed.2020.102001

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 16000-16009). https://doi.org/10.1109/CVPR52688.2022.01553

Joshi, S., & Karule, P. T. (2018). A review on exudates detection methods for diabetic retinopathy. *Biomedicine and Pharmacotherapy, 97*, 1454-1460. https://doi.org/10.1016/j.biopha.2017.11.009

Kamble, R., Kokare, M., Deshmukh, G., Hussin, F. A., & Mériaudeau, F. (2017). Localisation of optic disc and fovea in retinal images using intensity-based line scanning analysis. *Computers in Biology and Medicine, 87*, 382-396. https://doi.org/10.1016/j.compbiomed.2017.04.016

Kim, B. S., Shah, B., He, T., & Kim, K. I. (2022). A survey on analytical models for dynamic resource management in wireless body area networks. *Ad Hoc Networks, 135*, Article 102936. https://doi.org/10.1016/j.adhoc.2022.102936

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3992-4003). https://doi.org/10.1109/ICCV51070.2023.00371

Kumar, S., Adarsh, A., Kumar, B., & Singh, A. K. (2020). An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation. *Optics & Laser Technology, 121*, Article 105815. https://doi.org/10.1016/j.optlastec.2019.105815

Liu, S., Jiang, X., Zhang, J., & Zou, W. (2025). Hard exudates segmentation for retinal fundus images based on longitudinal multi-scale fusion network. *Medical and Biological Engineering and Computing, 63*(12), 3761-3775. https://doi.org/10.1007/s11517-025-03426-7

Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., & Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study. *Medical Image Analysis, 89*, Article 102918. https://doi.org/10.1016/j.media.2023.102918

Misera, L., Nebelung, S., Carrero, Z. I., Bressem, K., Ligero, M., Kühn, J.-P., Hoffmann, R.-T., Truhn, D., & Kather, J. N. (2026). Reducing manual workload in CT and MRI annotation with the Segment Anything Model 2. *BMC Medical Imaging, 26*(1), Article 75. https://doi.org/10.1186/s12880-025-02075-4

Motevali, S., Khanal, A., & Estrada, R. (2021). Optic disc segmentation using disk-centred patch augmentation. *arXiv*. https://doi.org/10.48550/arXiv.2110.00512

Nagpal, D., Pattanaik, P. A., Madaan, V., & Agrawal, P. (2024). Enhancement of retinal images through modified anisotropic diffusion. *Journal of Medical Artificial Intelligence, 7*, Article 80. https://doi.org/10.21037/jmai-23-80

Pallawala, P. M. D. S., Hsu, W., Lee, M. L., & Eong, K. G. A. (2004). Automated optic disc localisation and contour detection using ellipse fitting and wavelet transform. In *Lecture Notes in Computer Science* (Vol. 3022, pp. 139-151). https://doi.org/10.1007/978-3-540-24671-8_11

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD). *IEEE Dataport*. https://dx.doi.org/10.21227/H25W98

Ramani, R. G., & Shanthamalar, J. J. (2020). Improved image processing techniques for optic disc segmentation in retinal fundus images. *Biomedical Signal Processing and Control, 58*, Article 101832. https://doi.org/10.1016/j.bspc.2019.101832

Rayed, M. E., Islam, S. M. S., Niha, S. I., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2024). Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked, 47*, Article 101504. https://doi.org/10.1016/j.imu.2024.101504

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (Vol. 9351, pp. 234-241). https://doi.org/10.1007/978-3-319-24574-4_28

Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence, 15*(1), 1-22. https://doi.org/10.1007/s12065-020-00540-3

Septiarini, A., Hamdani, H., Setyaningsih, E., Junirianto, E., & Utaminingrum, F. (2023). Automatic method for optic disc segmentation using deep learning on retinal fundus images. *Healthcare Informatics Research, 29*(2), 145-152. https://doi.org/10.4258/hir.2023.29.2.145

Sevastopolsky, A. (2017). Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis, 27*(3), 618-624. https://doi.org/10.1134/S1054661817030269

Shelhamer, E., Long, J., & Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 640-651. https://doi.org/10.1109/TPAMI.2016.2572683

Stojmenovic, M., & Nayak, A. (2007). Direct ellipse fitting and measuring based on shape boundaries. In *Lecture Notes in Computer Science* (Vol. 4872, pp. 221-235). https://doi.org/10.1007/978-3-540-77129-6_22

Teoh, C. S., Wong, K. H., Xiao, D., Wong, H. C., Zhao, P., Chan, H. W., Yuen, Y. S., Naing, T., Yogesan, K., & Koh, V. T. C. (2023). Variability in grading diabetic retinopathy using retinal photography and its comparison with an automated deep learning diabetic retinopathy screening software. *Healthcare, 11*(12), Article 1697. https://doi.org/10.3390/healthcare11121697

Theera-Umpon, N., Poonkasem, I., Auephanwiriyakul, S., & Patikulsila, D. (2019). Hard exudate detection in retinal fundus images using supervised learning. *Neural Computing and Applications, 32*, 179-192. https://doi.org/10.1007/s00521-019-04402-7

Usman Akram, M., Khalid, S., Tariq, A., Khan, S. A., & Azam, F. (2014). Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine, 45*(1), 161-171. https://doi.org/10.1016/j.compbiomed.2013.11.014

Valverde, C., Garcia, M., Hornero, R., & Lopez-Galvez, M. (2016). Automated detection of diabetic retinopathy in retinal images. *Indian Journal of Ophthalmology, 64*(1), 26-32. https://doi.org/10.4103/0301-4738.178140

Xue, X., Wang, L., Du, W., Fujiwara, Y., & Peng, Y. (2022). Multiple preprocessing hybrid level set model for optic disc segmentation in fundus images. *Sensors, 22*(18), Article 6899. https://doi.org/10.3390/s22186899

Yin, P., Xu, Y., Zhu, J., Liu, J., Yi, C., Huang, H., & Wu, Q. (2021). Deep level set learning for optic disc and cup segmentation. *Neurocomputing, 464*, 330-341. https://doi.org/10.1016/j.neucom.2021.08.102

Zhang, Y., Cai, X., Zhang, Y., Kang, H., Ji, X., & Yuan, X. (2021). TAU: Transferable attention U-Net for optic disc and cup segmentation. *Knowledge-Based Systems, 213*, Article 106668. https://doi.org/10.1016/j.knosys.2020.106668