Indian Journal of Health Care, Medical & Pharmacy Practice Vol 4; Issue 1, Jan-June 2023, ISSN 2583-2069

Original Article

Use of Segment Anything Model (SAM) and MedSAM in the optic disc Segmentation of colour retinal fundus images: Experimental Finding

Ravi Bhushan Bhardwaj [1*], Dr. Anum Haneef [2]

1 Data Scientist, Department of Data Science, Kaali Al Labs, Gorakhpur, Uttar Pradesh, India.

2 Consultant Ophthalmologist, Department of Surgery Ophthalmology, Tehsil Hospital, Lahore, Pakistan.

Corresponding Author*: Ravi Bhushan Bhardwaj, Data Scientist, Department of Data Science, Kaali Al Labs, Gorakhpur, Uttar Pradesh, India.

Email ID: ravibhushan@gmail.com
Doi: https://doi.org/10.59551/IJHMP/2023.4.9

Received: 17 May, 2023, Decision for Acceptance: 5 June, 2023

Abstract:

Detection of Optic Disc segmentation in retinal fundus images is important step in identification of various abnormal conditions like diabetic retinopathy, Glaucoma etc. and is an important part of eye that is routinely examined. We have used Segment Anything model (SAM) by Meta AI and based fine-tuned model MedSAM for segmentation of optic disc in retinal fundus images. We have used Indian Diabetic Retinopathy dataset (IDRiD) segmentation part. It consists of 81 original colour fundus images in jpg files split into train and test set. Ground-truth images for the lesions (Microaneurysms, Haemorrhages, Hard Exudates and Soft Exudates divided into train and test set - TIF Files) and Optic Disc (divided into train and test set - TIF Files). This dataset is feed into the deep learning neural network model; Segment Anything Model (SAM) and MedSAM for the image segmentation task. The Dice Similarity Coefficient (DSC) is used in the experiment for observing the model performance. We have used 80% for training and 20% for testing of Indian Diabetic Retinopathy Dataset (IDRiD) in both models. After 100 epochs with 32 batch size it is observed that average Dice Similarity Coefficient (DSC) of SAM and MedSAM models in the optic disc segmentation task are, 85.97 % and 90.15 %. Also, in finding it is observed that SAM and MedSAM both have very low DSC where the fundus images are having very high brightness. With the great success of the Segment Anything Model (SAM) in natural image segmentation, MedSAM model is especially fine-tuned for the medical images analysis by the developers available as open source can be used as a revolutionary tool in medical image analysis. Because SAM is not trained for medical image segmentation in spite of it DSC score is quite good but MedSAM results are very promising even its medical data set is only trained with 200,000 masks across

11 different modalities. SAM can be use as the annotating tool and foundation base model for other machine learning models.

Keywords: Deep Learning, Medical Image analysis, Optic Disc Segmentation, Diabetic Retinopathy, Segment Anything Model (SAM), MedSAM.

1. Introduction:

In an eye there is a round circular disc type spot formed by the axons and the retinal ganglion cells. These are responsible for transmitting signals from the photoreceptors of the eye to optic nerve which allows us to see this beautiful world and our loved ones. Optic disc is also called nerve head, a circular area on the back of the eye where the optic nerve exists and enters the brain. It is a light sensitive tissue usually found in the centre of the retina and lines the inner surface of the eye. It appears as a slightly raised yellowish white disc shaped area on the retina and is an important part of the eye that is routinely examine to detect various eye diseases and conditions. Detection of OD (Optic Disc) in retinal fundus images is an important step in identification of various abnormal conditions like; Diabetic Retinopathy, Glaucoma etc.

Renoh Johnson Chalakkal [10] used feature extraction from retinal images to automatically detects the optics disc (OD) based on histogram template that combines with the maximum sum of vessel information in the retinal images. Circular Hough transform used to segment the OD region. The proposed study uses fovea detection by uniformly dividing the retinal image into three horizontal strips and the strip including the detected OD is selected. They have used publicly available databases DRIVE [17], DIARETDB0 [12], DIARETB1 [13], CHASE [14], MESSIDOR [15], STARE [16] and IDRiD [7]. On IDRiD dataset available 81 images for OD segmentation of boundary on manually annotated masks, they observed an average overlap score and sensitivity of 0.856 and 0.966. Shuang Yu, Di Xiao, Shaun Frost, Yogesan Kanagasingam [3] has used deep learning based on U-Net architecture with backbone of ResNET – 34 and developed robust segemenatation model for the optic disc and cup segmentation for Glaucoma detection. They have use dice similarity coefficient and achieved an average dice value of 97.31% for disc segmentation and 87.61% for cup segmentation. There test results are based on the available DRISHTI-GS and RIM-ONE dataset [2]. Use of autoencoders for automatic optic disc segmentation is also popular. Shalee Bengani [11] used transfer learning method based on concepts of semi-supervised. They have used the DRISHTI-GS and RIM-ONE dataset [2] and through the deep learning neural network model tried to automatically segment the optic disc in retinal fundus images. A large number of unlabeled fundus images fed into the convolutional autoencoder from the dataset. After that the autoencoder learns the feature from the unlabeled fundus images and build a pre-trained neural network model. And with the help of transfer learning, network is trained with the ground-truth optic disc images and their respective retinal fundus images from the used data set. They have observed the experimental values based on DSC 0.967 and 0.902 test set of DRISHTI GS1 and RIM-ONE dataset.

It is worth mentioning that deep learning is data hungry so it requires more than thousands images in dataset for better accuracy. Using U-Net in medical [3] image segmentation gains a wide popularity since it is introduced in 2015. More than 2500 time's cited academically by researchers because of its good application in the upcoming future application in medical science.

Over traditional machine learning deep learning has gained popularity [4] for the prediction and analysis of medical images. Because the future is getting revolutionized with the innovations in the Artificial Intelligence. Introduction of Segment Anything Model [5] has changed everything and we can experience OpenAIGPT – 4 [1] of future in medical image analysis. Segment Anything Model (SAM) is released as an open source for the development and deep learning research by MetaAI. The dataset is trained with 1 billion mask on 11 million licensed and privacy respecting images. SAM is capable of segmenting user defined objects when prompts are given. SAM can learn any object and segment any object by user defined prompt. The subject images given to the SAM model are not introduced before the model and without any additional training, SAM is capable of segmenting object of interest in zero-shot learning. SAM is trained on (SA – 1B) dataset this dataset is developed in three stages. First stage includes the manual annotation by the humans on a set of images with the help of clicking and SAM generated masks are manually refined. In the second stage, the annotators segment the masks that were not accurately and confidently generated by SAM. In final step the mask are generated automatically by SAM with a set of point distributed across the image and selecting the confident masks. MA. Mazurowski [6] in his research paper explained about the various uses of the SAM in a medical image analysis. SAM can be used as the semi-automated annotation tool where human will provide prompts and masks are generated by SAM. SAM can be use with other models to generate masks on unlabelled images. Object detection based on bounding box can be use with the help of SAM for generating precise segmentation of mask. In experiment with the Segment Anything model (SAM) the research paper used MRI – Spine, MRI – Heart, MRI – Prostate, MRI – Brain, Xray – Chest, Xray- Hip, US-Breast, US-Kidney, CT-Kidney, CT-Liver, CT-Organ, PET-Liver dataset. The paper uses additional RITM for the comparison of the segmented results. SAM showed the better performance based on single prompt learning in comparison to the RITM model. But the SAM performance is found to be similar with the RITM when it comes to multiple larger number of prompts. Authors has observed based on the IoU performance the 0.1135 for spine MRI dataset to 0.8650 for a hip x-ray dataset. The research paper

[6] recommends the extensive research for the adaptation of SAM in Medical Image segmentation.

MedSAM by Jun Ma and BO Wang [8] is the extension in the research of the SAM for medical Image analysis. And it is the first implementation of SAM after fine tuning developing MedSAM for the medical image segmentation. MedSAM is available as an open source for developing and using the model for custom dataset in research. The objective of this research paper is to segment the OD (Optic Disc) based on the diabetic retinopathy disease with the help of SAM and MedSAM models. The dataset Indian Diabetic Retinopathy dataset (IDRiD) [7] used for the optic disc segmentation analysis. We will compare the results generated by the SAM and MedSAM on ground-truth and predicted masks for the optic disc segmentation using the Dice Similarity coefficient.

2. Methods

2.1 Understanding Segment Anything Model (SAM) and MedSAM

2.2 Segment Anything Model (SAM)

The Architecture of SAM is based on transformer architecture. This is very effective in natural language processing [8] and image analysis task. SAM vision transformer extract features from the images with the help of image encoder and prompt encoders allows users interactions. Mask decoder used to generate image segmentation results and confidence scores [5].

The Image encoder is pretrained with mask auto-encoder modelling, which allows to process high resolution images i.e.; 1024 x 1024. The obtained image embedding is 16x downscaled i.e.; 64 x 64. Points, boxes, texts and masks are the four prompts which SAM uses for any image segmentation task. Pre-trained text encoder in CLIP encodes the free form text. Point encoding is used for the bounding. Convolutional feature maps keeps the resolution of masks same as the input images. The mask decoder consisting two transformer layers are capable of mask prediction and Intersection over Union (IoU) score. This mask prediction is capable of generating three 4x downscaled masks, which address to the whole object, part and subpart of the subject [5].

Figure 1 Segment everything mode

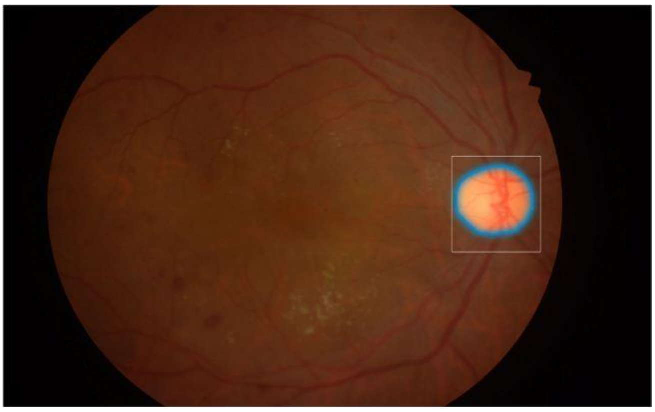

Figure 2 Bounding Box mode

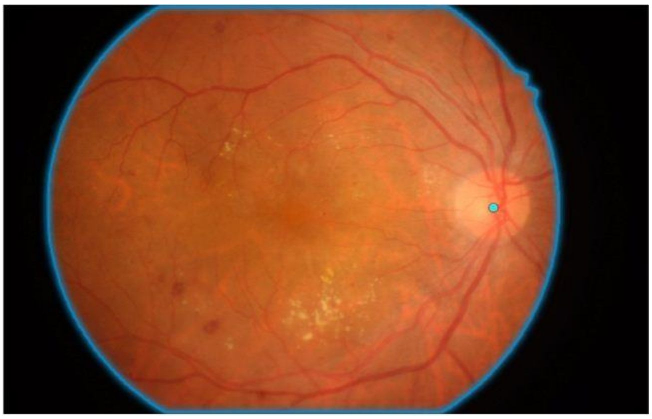

Figure 3 Point mode: 1st point

Figure 4 Point mode: 2nd point

We have checked out the SAM capability over the retinal fundus images for the segmentation task on online demo version provided by Meta AI. We applied the three mode of SAM i.e; Segment Everything mode, Bounding box and point mode. The Segment everything mode is fully automatic in *Figure 1* we can see that segment everything mode use the whole image and segment everything in the image based on the intensity of the image. But this mode segment the optic disc not correctly as well the mode does not work on the lesions segmentation. We can say that segment everything mode is not useful for the task. On the other hand, if we will see in *Figure 2* the bounding box method result is very promising and somewhere it is accurately segmented the particular region of the optic disc. Point mode of SAM is also not useful in our case because when we provided the retinal fundus image with a single point see *Figure 3* SAM segmented the whole eye region rather than the optic disc. We provided one more point as a background *Figure 4* the SAM was able to segment the optic disc but it was not accurate. Overall, we can see that the bounding box method of SAM is capable of segmentation of the optic disc for now only.

2.3 MedSAM: Dedicated model based on SAM for Medical Image Analysis

Segment Anything Model (SAM) is not suitable for medical image segmentation task but the model can be use after fine tuning in the medical image analysis so the MedSAM model is capable of doing this. MedSAM is trained with large – scale medical image dataset, ecompassing over 200,000 masks across 11 different modalities. The authors of MedSAM did experiment on 21 3D segmentation tasks and 9 2D segmentation task to compare the MedSAM and SAM over medical image segmentation task [8].

We have seen in above *Figure 2* that bounding box can be a good choice for the segmentation in the medical images. SAM model consist of Image encoder, prompt encoder and mask decoder. MedSAM model keep frozen the Image encoder to reduce the computational costs. Also with the prompt encoder is frozen [8] in order to eliminate the positional information of the bounding box because SAM uses pre-trained bounding box encoder and this will avoid reused of it. For all training images in the model the image embeddings are pre-computed this will avoid the replication of the image embeddings per prompt this will increase the training efficiency. The mask decoder in MedSAM generates only one mask because of the bounding box mode which specify the target segmentation task. See *Figure 5* to understand how MedSAM works.

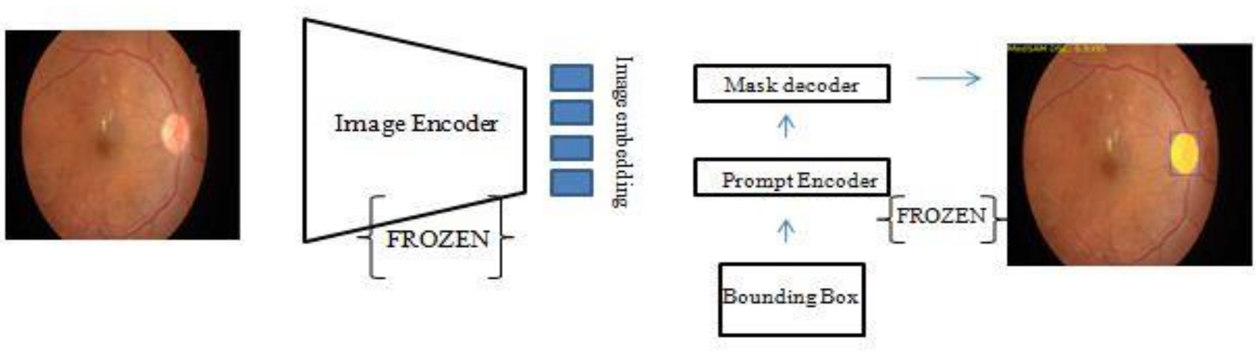

Figure 5 MedSAM: Working of fine-tuned SAM model as MedSAM

MedSAM uses the pre-trained ViT-Base model of the SAM as an image encoder where all the image embedding computed on the basis of the normalized images given to the image encoder. Here, the image encoder transforms the images to size of 3 x 1024 x 1024. The MedSAM models slices the 3D image into 2D image for the task along the plane dimension. The bounding box prompt was generated [8] from the ground-truth mask with the random perturbation of 0-20 pixels. The function used in MedSAM is the unweighted sum between Dice loss and cross entropy which are robust [9] in medical image segmentation task. In the network Adam optimizer with an initial learning rate of 1e-5 is used.

3. Results

3.1 Data Collection and Testing

The dataset Indian Diabetic Retinopathy Dataset (IDRiD) used in the model is split into 80% for the training and 20 % for the testing purpose (i.e. 65 images for the training and 16 images for testing with ground-truth masks). Also, the retinal images are patchify into 256 x 256 x 3. The intensity values are normalized to the range of [0, 255]. The 3D masks are changed into 2D by slicing along the plane. Number of epochs assigned for the segmentation task is 100 starting from initial 0 having batch size 32. The optic disc segmentation analysis is done with the google colab environment with the free allocated GPU and free available Ram. All other training rules are kept same provided by MedSAM model [8] as discussed above about MedSAM.

3.2 Evaluation of Results

We have use Dice Similarity Coefficient (DSC) as it is very popular among medical image analysis [9] and also the MedSAM uses the same DSC for the segmentation evaluation [8] for the ground-truth and image segmentation results. MedSAM performance was overall high with an average DSC score of 85.97 % and for SAM model and 90.15 % for MedSAM model. In the *Figure 6* and *Figure 7* on the left side image with ground-truth mask is shown with the optic disc region. In the middle and right side the predicted masks generated by SAM and MedSAM is shown. In *Figure 6* DSC score for SAM and MedSAM is observed 0.8802 and 0.9109. Also in *Figure 7* the DSC scores are 0.6937 and 0.7995 respectively, also we can see that how the brightness have effect on the accuracy of the both SAM and MedSAM model. DSC scores of 16 results of image segmentation on retinal fundus images are given in the Table 1 for the comparison with SAM and MedSAM model.

Our segmentation is based on the 2D images so the SAM and MedSAM both are performing their decent jobs. Because SAM is basically designed for the 2D segmentation and appropriate bounding box method helps SAM to predict masks worth comparable to MedSAM. The checkpoint used in the SAM model is (ViT – B) for comparing with MedSAM model. We have observed loss in training 0.071 which is quite good see *Figure 8*. The dice loss plus cross entropy loss graph is shown in *Figure 9* a loss value of 0.250 with 100 epochs is observed which is a good sign because the dataset was not huge only having 81 fundus images for training and testing purpose we can say that the model is performing well. This could be improved by feeding more training images in the dataset.

Figure 6 Ground-truth mask (on left), Prediction by SAM (Middle), Prediction by MedSAM (on right)

Figure 7 Ground-truth mask (on left), Prediction by SAM (Middle), Prediction by MedSAM (on right)

```
EPOCH: 98, Loss: 0.06862227618694305
100%|██████████| 2/2 [00:00<00:00,  2.47it/s]
EPOCH: 99, Loss: 0.0717271976172924
```

Figure 8 Number of Epochs and loss during training

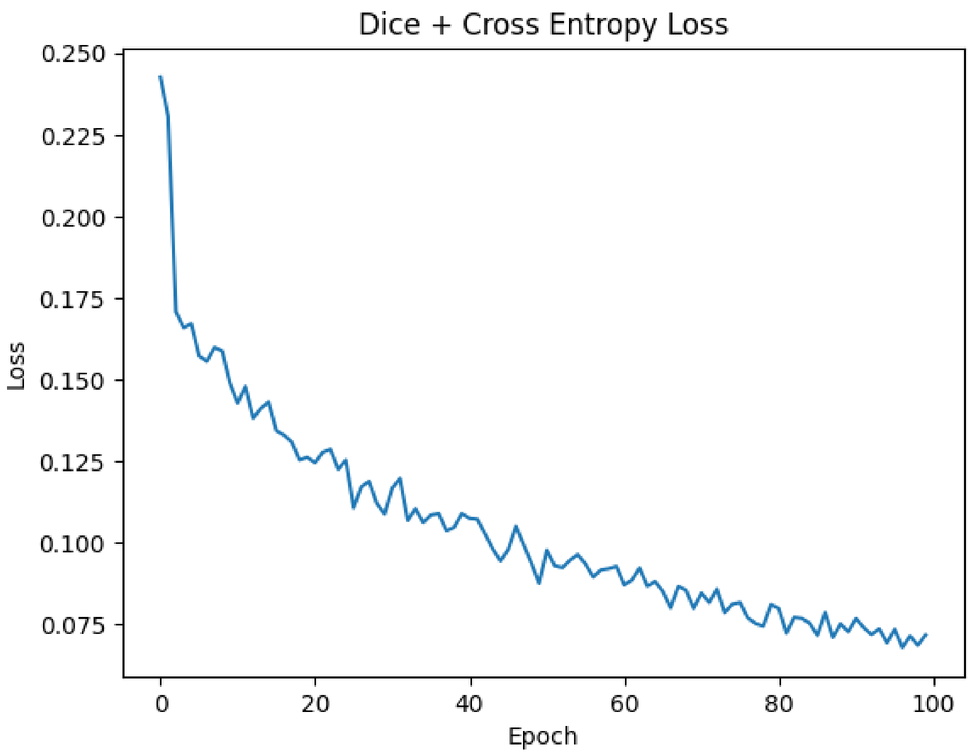

Figure 9 Dice plus Cross Entropy Loss

Table 1 Performance of SAM and MedSAM on Optic Disc Image Segmentation Task

Segmentation	Target Modality	DSC %	
		MedSAM	SAM
Optic Disc (OD)	**Retinal Fundus Images**	93.95	**94.71**
		79.95	69.37
		93.87	**94.99**
		90.85	89.95
		87.24	82.13
		91.09	88.02
		91.68	87.67
		90.85	89.95
		93.41	74.99
		91.27	90.35
		89.76	82.52
		75.67	68.34
		91.72	87.94
		95.78	93.85
		94.30	92.76
		91.06	87.98
Average		**90.15**	85.97

4. Discussion

4.1 Key Findings

We have observed a very good improvement in the Optic Disc Segmentation on Diabetic Retinopathy by using MedSAM over SAM results. The DSC scores clearly shows that SAM has only average DSC score of 85.97 % where MedSAM shows 90.15 % which is 4.18 % of improvement in the optic disc segmentation results between SAM and

MedSAM models. This is worth notice that the dataset has only few 81 images for both the models this will be a good idea to use pre-trained models for medical image segmentation results with zero-shot learning in future of medical image analysis. SAM and MedSAM both can be use as a tool for unlabelled medical images for the segmentation task.

4.2 Strengths and limitations

The SAM as well as MedSAM both performs well on the 2D segmentation tasks but when it comes to 3D segmentation task MedSAM performance is higher [8]. But for now both the models are not able to segment the weak boundaries as well lesions in the images. The observation on multi-task segmentation results were found to be very less accurate by the MedSAM and SAM in their research paper [8].

4.3 Explanations of findings

The experimental results observed on the segmentation of the Optic Disc in retinal fundus images using dataset Indian Diabetic Retinopathy dataset (IDRiD) is compared on the basis of Dice Similarity Coefficient. The Dice loss and cross entropy loss is used as a combination to evaluate the performance of model on the segmentation. DSC scores of prediction are good on high contrast images but low at the high brighter images. But the MedSAM model based on SAM after fine tuning have performed well and can be use in the medical image analysis even on unlabelled masks.

5. Conclusions

On retinal fundus images obtained from the Indian Diabetic Retinopathy dataset (IDRiD) for the Optic Disc segmentation task we have observed the Dice Similarity Coefficient (DSC) for SAM and MedSAM, 85.97 % and 90.15 % respectively. On retinal fundus images having higher brightness both models have similar average DSC scores. This can be conclude from the obtained results that MedSAM can be used for medical image analysis after training with more dataset because currently it has only 200,000 masks across 11 modalities. Appropriate more fine tuning can make the model to achieve higher accuracy. Implementation of deep learning models like SAM and MedSAM can increase accuracy and less human intervention. SAM can be use as annotating tool and foundation for segmentation models.

Conflict of Interest: The authors declares no conflicts of interest.

References

[1]. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, et al.. Sparks of Artificial General Intelligence: Early experiments with GPT-4. 2023;. Available from: http://arxiv.org/abs/2303.12712

[2]. Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. 2014; 53–6. Available from: https://typeset.io/papers/drishti-gs-retinal-image-dataset-for-optic-nerve-head-onh-120vmynnen

[3]. Yin XX, Sun L, Fu Y, Lu R, Zhang Y. U-Net-Based Medical Image Segmentation. 2022May10;2022. Available from: https://www.hindawi.com/journals/jhe/2022/4189781/ A Review on Traditional Machine Learning and Deep Learning Models for WBCs Classification in Blood Smear Images. 2020;9:10657–73.

[4]. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al.. Segment Anything. 2023;. Available from: http://arxiv.org/abs/2304.02643

[5]. Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. Segment Anything Model for Medical Image Analysis: an Experimental Study. 2023;. Available from: http://arxiv.org/abs/2304.10517

[6]. Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Fabrice Meriaudeau. Indian Diabetic Retinopathy Image Dataset (IDRiD) [Internet]. IEEE Dataport; 2018. Available from : https://dx.doi.org/10.21227/H25W98

[7]. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al.. Attention Is All You Need. 2017Jun12;. Available from: http://arxiv.org/abs/1706.03762

[8]. Ma J, Wang B. Segment Anything in Medical Images. 2023;. Available from: http://arxiv.org/abs/2304.12306

[9]. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. 2021May9;95:102026. Available from: https://www.sciencedirect.com/science/article/pii/S0895611121001750

[10]. Renoh Chalakkal, Faizal Hafiz, Waleed Abdulla, Akshya Swain, An efficient framework for automated screening of clinically Significant Macular Edema, Computers in Biology and Medicine, 10.1016/j.compbiomed.2020.104128, **130,** (104128),2021.

[11]. Bengani,S.,J.,A.A.J. & S., Automatic segmentation of optic disc in retinal fundus images using semi-supervised deep learning. Multimed Tools Appl 80, 3442-3468 (2021). https://doi.org/10.1007/s11042-020-09778-6

[12]. Staal J., Abramoff M.D., and Niemeijer M. et al: 'Ridge-based vessel segmentation in color images of the retina', IEEE Trans. Med. Imaging, 2004, 23, (4), pp. 501-509

[13]. Kauppi T., Kalesnykiene V., and Kamarainen J.K. et al; 'DIARETB0: evaluation database and methodology for diabetic retinopathy algorithms'. Technical report

[14]. Kauppi T., Kalesnykiene V., and Kamarainen J.-K. et al: 'The DIARETDB1 diabetic retinopathy database and evaluation protocol',. Proc. BMVC, Warwick, UK, 2007, pp. 15.1 – 15.10

[15].Messidor Database:Available at:http://www.adcis.net/DownloadThirdParty/Messidor.html, accessed 15 June 2017

[16]. Hoover A.D., Kouznetsova V., and Goldbaum M.:'Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response', IEEE Trans. Med.Imaging, 2000, 19, (3), pp.203-210.

[17]. Staal J., Abramoff M.D., and Niemeijer M. et al: 'Ridge-based vessel segmentation in color images of the retina', IEEE Trans. Med. Imaging, 2004, 23, (4), pp. 501 - 509

Cite this article Bhardwaj Bhushan R *et al,* Use of Segment Anything Model (SAM) and MedSAM in the optic disc Segmentation of colour retinal fundus images: Experimental Finding. Indian Journal of Health Care, Medical & Pharmacy Practice.2023; 4(1) 82-93.