[1,2,3]Nanditha Krishna

[4]Nagamani K.

A Deep Learning Model for Neural Network Optimization for Glaucoma Classification Using Fundus and Oct Feature Fusion

Journal of Electrical Systems

Abstract: - Irreversible vision loss is a common consequence of glaucoma, demands accurate and timely diagnosis for effective management. This research aims to enhance glaucoma classification accuracy by fusing information from two distinct imaging modalities like using deep learning explores the fusion of these modalities through an innovative neural network architecture with optimization. This approach combines Deep Stochastic Variational Autoencoder Convolution Neural Networks (DSVAECNN) and Adam optimization techniques to enable robust and accurate classification of glaucoma. A multi-path architecture is designed to accommodate both imaging as optical coherence tomography (OCT) radiomics features and fundus morphological features simultaneously. To ensure the effectiveness of the model, this study investigates the implementation of advanced optimization algorithm such as Adam, to expedite convergence and alleviate the risk of over fitting. The resulting model demonstrates improved generalization capabilities, critically for accurate diagnosis across diverse patient populations. A comprehensive OCT and fundus images dataset is used to evaluate the proposed approach from a representative cohort of glaucoma patients and healthy individuals. Quantitative various metrics including sensitivity, accuracy and specificity are employed toward appraise the accomplishment of the fusion-based classification model. Comparisons with current techniques show the dominance of the proposed move toward in accurately detecting glaucoma cases. This research provides the advancement of glaucoma diagnosis by effectively harnessing the synergy between fundus image and OCT scans and findings helpful information for clinics, ultimately facilitating early detection and personalized management of glaucoma, thus preserving vision for affected individuals.

Keywords: Classification, Features, Fundus, Image, Glaucoma, OCT, Optimization.

I. INTRODUCTION

The progressive optic neuropathy known as glaucoma; this condition causes irreversible vision loss globally. Timely and accurate classification of glaucoma stages is essential for effective patient management. The evolution of glaucoma classifying methods spanning from traditional methods to the recent advances in deep learning approaches gains research work extensively. Detecting glaucoma is a critical aspect of eye health, It is one of the leading causes of blindness in the world. There are various methods and techniques used for glaucoma detection, including both traditional and advanced approaches. An excess of fluid builds up in the anterior part of the eye, leading to retinal disease glaucoma, causing damage to the optic nerve. Among the types of glaucoma there are Angle-closure glaucoma (ACG) and Primary open-angle glaucoma (POAG) [1]. A common cause of glaucoma, POAG is caused by the eye being unable to drain fluid, causing the pressure to build up and damaging the optic nerve. Pressure in the eye rises when the drainage angle is blocked. There are a number of symptoms, including severe eye pain, blurry vision, headaches, etc. As a result, regular eye examinations are helpful in both cases [2], and early treatment can prevent blindness. It is predicted that by 2040, there will be 111 million people affected by this disease compared to 75 million in 2020. Asia and Africa will be the most affected, followed by South America. Having glaucoma in one's immediate relatives increases the likelihood that one will develop the disease compared to having an adverse family history [3]. In each of the 100 subjects (40 healthy and 60 glaucoma), a horizontal B-scan was acquired using OCT [4]. Adaptive compensation was used to enhance all images. Based on compensated images, a system developed and trained a deep learning network to digitally stain each layer of the optic nerve head (ONH). The intersection over union (IU) is a factor in addition to sensitivity, specificity and accuracy, the dice coefficient was used to assess the accuracy of the algorithm (in comparison to manual segmentations). Using glaucoma and healthy subjects as comparisons, the method looked at the effect of compensation and training images to achieved good results. The feasibility of automated glaucoma detection using CNNs trained from non-medical data [5]. In order to

[1 1] Research Scholar, Department of Telecommunication, RVCE, Bengaluru-560059, Karnataka, India.

[2] Research Scholar, Visvesvaraya Technological University,Belagavi-590018, Karnataka, India.

[3]Assistant Professor, Department of Medical Electronics Engineering, DSCE, Bangalore-560078, Karnataka, India.

[4]Professor, Department of Telecommunication, RVCE, Bangalore-560059, Karnataka, India.

*Corresponding author: nanditha13@gmail.com

generate feature vectors from fundus images, two CNNs were applied, namely OverFeat and VGG-S. Based on the analysis of the Drishti-GS1 dataset, the method approach is viable and provides significant evidence that when there is not enough data to fine-tune the network, in order for transfer learning to take place, it is essential to select a well-designed image preprocessor. Detection of glaucoma using the CNN U-net [6]. By using the U-net model, it can identify precisely where glaucoma develops in eye. In comparison to CNN model, U-net uses a simpler algorithm while being more efficient. In this U-net, images are segmented, features are extracted and classified using preprocessing, segmentation, and feature extraction. Method uses hybrid layers of ResNet-34 encoding and classical U-Net decoding for robust disc and cup segmentation [7]. U-Net and ResNet are combined to produce a pre trained model, thus enabling it to train faster with fewer epochs, and hence avoid over-fitting. A density-based probability function and local binary patterns are used in optical domain segmentation [8]. In order to prevent segmentation mistakes, the red channel was segmented first, followed by equalizing the histograms of all images before segmentation begins. Through the use of fundus image, an algorithm is developed in order to be used to diagnose glaucoma [9]. There are 1,542 images in the dataset and these are divided into 754 training images, 324 validation images, and 464 test images. Used Tensorflow to construct convolutional neural networks and simple logistic classifications from the datasets. Pretrained GoogleNet Inception v3 model was fine tuned using the same datasets. 82.9% of the training data accuracy, 79.9% of the validation data accuracy, and 77.2% of the test data accuracy and achieved a 92.2% AUROC in the convolutional neural network.By using a hybrid feature set, the technique detects glaucoma in digital fundus images [10]. For improving automated glaucoma diagnosis accuracy, a novel method applied of structural features, such as cup to disc ratio, is combined with non-structural features, such as texture and intensity. When structural and non-structural features conflict, a suspect class is introduced in automated diagnosis. A suspect class was introduced to ensure that the proposed system is highly sensitive to glaucoma cases from rural areas. In general, the proposed system has a sensitivity of 96 and a specificity of 87 %. Several aspects of the detection systems are examined and analyzed for mobile use in order to determine the most effective factors for Glaucoma detection systems based on CNNs[11]. According to MobileNet, training time, dataset variability and dataset size are important mobile considerations. Fundus images were utilized to detect ophthalmologic diseases [12]. The NCAR algorithm and the SVM algorithm were used to extract deep features with high representative details from the R-CNN+LSTM-based model. To make fast predictions, the model requires powerful hardware resources because to its high number of learnable parameters. By using five neural network models fully connected, a DL model was developed for the early detection of glaucoma[13]. Based on the results, ResNet50 appears to be the most accurate. The model used only one dataset for testing and training, and there was no comparison of how the models performed in actual clinical examinations by ophthalmologists. Through deep learning-based feature extraction, glaucoma can be detected at an early stage [14]. From the images containing optic cups (OC), Hybrid features descriptors are used to extract features from the optic disc (OD), among them are histograms orientate gradient (HOG), local binary patterns (LBP), and CNNs and speeded up robust features (SURF). A low-level feature extraction method is provided by the HOG descriptor, while a texture extraction method is provided by the LBP and SURF descriptors. Additionally, CNN is used to compute high-level features and also selected the most representative features using the MR-MR method, a technique that involves selecting and ranking features. In the final step, several multi-class classifiers, such as SVM, RF, and KNN, are used to classify fundus images as healthy or diseased. There have been good results obtained using models based on the RF algorithm combined with HOG, CNN, LBP, and SURF feature descriptors, when using k-fold cross validation to detect early glaucoma. Utilizing maps of retinal nerve fibre thickness is evaluated the diagnostic effectiveness of machine learning models for glaucoma diagnosis [15]. The model outperforms conventional RNFL thickness measurements in terms of accuracy. According to current literature, the limitations of current glaucoma detection methods must be overcome by expanding novel methods and testing them on a larger range of datasets. Flexible and reliable evaluation metrics is also crucial to ensuring the effectiveness of the techniques. Multimodal imaging techniques, including alternative imaging techniques, can enhance detection accuracy by exploring multiple modalities. In addition, novel algorithms must be developed in order to improve the accuracy of feature extraction and segmentation. The incorporation of clinical data in DL models is another way to make them more applicable in clinical settings. It is anticipated to promote research work in this field by employing DL and considering some factors such as regularizing network parameters will greatly improve automatic glaucoma diagnosis' precision and efficiency. Conducted a retinal fundus glaucoma challenge dataset and suggested a multitask deep learning model to diagnose glaucoma [16]. The model's application is limited by the availability and cost of expert annotations. In order to improve the classification accuracy of healthy and glaucoma samples, this work developed a deep learning model with neural network

optimization by fusing fundus and OCT features. A novel DSVAECNN and Adam optimization technique is combined in the proposed work to provide more effective and robust glaucoma classification.

II. PROPOSED METHODOLOGY

A new techniques of Deep Learning convolution Neural Networks (DLCNNs) and optimization techniques are developed to diagnose the retinal diseases using OCT and fundus images in this work. Radiomics features [17-18] i.e image texture features are extracted from OCT images and a Nadam optimizer [19] coupled with a U-net segmentation model [20] is used to extract morphological features from fundus images. As a final step, the DSVAECNN model [21] is used to classify retinal image as glaucoma or healthy based on features fusion of OCT and fundus. Figure 1 is an illustration of the system model block diagram.

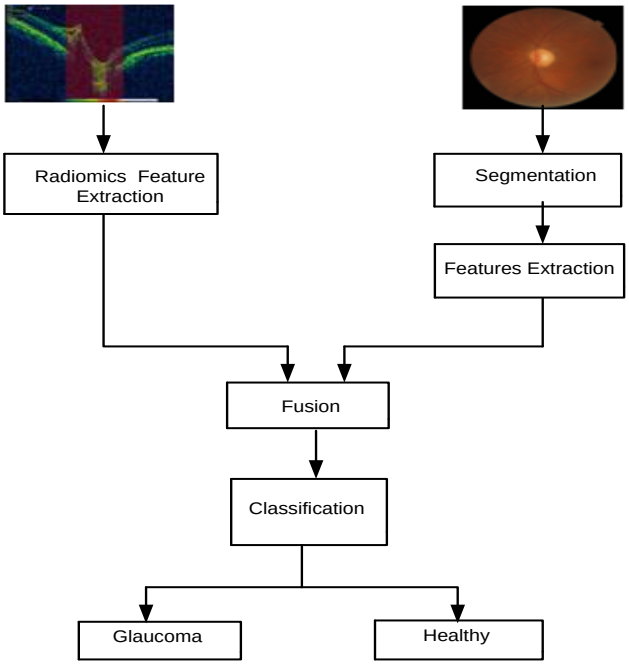


Figure 1: System model block diagram

2.1 OCT Image Feature Extraction

A total of 94 features were extracted from OCT images through six different radiomics which are as follows.

(i) **First order-19 features:** Voxel intensities within an image region are described by first-order statistics. They provide essential statistical information about an image without considering spatial relationships between pixels. These features are mainly useful for image characterization.

(ii) **Gray level co-occurrences Matrices (GLCM)-24 features:** Image texture analysis using GLCM goes beyond first-order texture features by considering the spatial relationships between pixels. It quantifies the occurrence frequencies of pairs of pixel intensities at specified relative positions in an image, they make available further detailed and descriptive texture information.

(iii) **Gray Level Dependence Matrix (GLDM) -14 features:** This provides unique texture features that complement the texture exploration techniques like GLCM by its directional and rotational invariance, local texture characterization, and also measures contrast and dissimilarity.

(iv) **Gray Level Run Length Matrix (GLRLM) -16 features:** It characterizes the runs of consecutive pixels with the same intensity value along specified directions; provides set of features so as to illustrate the circulation of these run lengths of texture patterns and structures present in an image.

(v) **Gray Level Size Zone Matrix (GLSZM) -16 features:** It quantifies the distribution of connected regions of pixels with the same intensity value in an image and provides a unique set of texture characteristics that provides information concerning the size and spatial arrangement.

(vi) **Neighboring Gray Tone Difference Matrix (NGTDM) -5 features:** It focuses on quantifying the differences between neighboring pixels in an image and provides a set of features that describe the allocation of these local intensity differences, offering fine-scale variations in texture patterns.

2.2 Fundus image feature extraction

The texture feature extraction from fundus image comprises a retinal segmentation that segments optic cup and disc using U-Net with Nadam optimization. Further an extraction of morphological features is performed from segmented image.

2.2.1 Segmentation

Segmentation of fundus images using U-Net is a popular and effective approach, particularly for tasks like segmenting retina image. The cup and disc are crucial anatomical structures within the optic nerve head, and their accurate segmentation is essential for detecting and monitoring glaucoma, here are the details of U-Net architecture:

- **Encoder-Decoder Structure**: The U-Net architecture is a unique encoder-decoder structure, the encoder captures contextual and hierarchical feature information from the image, while the decoder reconstructs the segmented masks. It encapsulates the intricate details and boundaries of the cup and disc regions.
- **Skip Connections:** With U-Net, the encoder and decoder have skip connections between their corresponding layers. These connections help in retaining fine-grained spatial information, enabling the network to recover accurate segmentation boundaries. For cup and disc segmentation, where precise boundaries are critical, skip connections are beneficial in maintaining spatial fidelity.
- **Limited Training Data**: The U-Net architecture ability for the purpose of learning on or after a smaller set of data and effectively capture relevant features makes it a suitable for accurate training.
- **Custom Loss Functions**: U-Net can easily accommodate custom loss functions for cup and disc segmentation, specialized loss functions can emphasize accurate boundary detection and prohibit misclassification near the optic nerve head.
- **Fine-Tuning and Transfer Learning**: U-Net architecture allows for fine-tuning on smaller datasets, which can help to improve segmentation accuracy using Nadam optimization. However, it depends on the network structural design and hyper parameters. The equations for the Nadam optimizer in U-Net are as follows:

When discussing weight initialization techniques and optimization algorithms, computing each parameter first moment m as in Eq. (1) and second moment v as in Eq. (2). These moments usually refer to statistics concerning a neural network's weights. An adaptive optimization algorithm and batch normalization algorithms rely heavily on these moments.

$$m_t = \beta_1 \times m_{t-1} + (1-\beta_1) \times g_t \tag{1}$$

$$v_t = \beta_2 \times v_{t-1} + (1-\beta_2) \times g_t^2 \tag{2}$$

Where g_t measures the gradient loss in relation to the parameters at a given current time step t, the m_t and v_t are the moments of the parameter at current time step, β_1 and β_2 are the exponential decay rates of first and second moments as in Eq. (3) and Eq.(4), typical values are nearly to one.

Compute the bias-corrected moments:

$$m_t^{\,h} = \frac{m_t}{(1-\beta_1^t)} \tag{3}$$

$$v_t^h = \frac{v_t}{(1-\beta_2^t)} \tag{4}$$

Update the parameters using the look ahead momentum as in Eq. (5) and Eq. (6):

$$v_t^b = (1-\beta_2) \times g_t^2 + \beta_2 \times v_t^h \tag{5}$$

$$m_t^b = (1-\beta_1) \times g_t + \beta_1 \times m_t^h \tag{6}$$

$$\theta_t = \theta_t - \alpha \times \frac{m_t^b}{(\sqrt{(v_t^b)} + \grave{o})} \quad (7)$$

Where model parameters are often represented by θ_t at time step t in Eq. (7), it describes how neurons in a network are connected and how they are biased, the α is learning rate, controlling the step size in the parameter update. The $\grave{o}$ important one which prevents division by zero by being the smallest value.

1.2.2 Fundus image feature extraction

As part of this study, cup disc ratio (CDR) [22], disc damage likelihood scale (DDLS) [23-24] and ISNT rule [25-26], segmented fundus images are used to calculate these features.

CDR: In the fundus image, a circular region is known as the optic disc where the optic nerve exists in the eye, and it appears as a lighter-colored circular area. As in figure 2, the cup is the central depression or hollow within the optic disc, and it appears as a darker circular area.

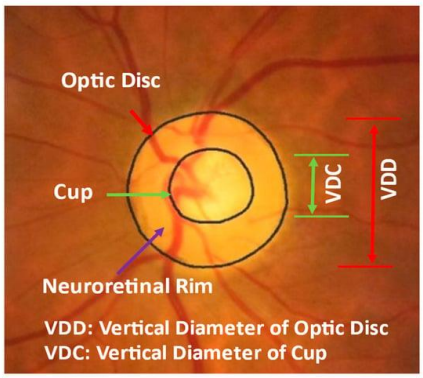


Figure 2: Optic cup and disc structure

The cup is formed because the retinal nerve fibers converge to form the optic nerve. The cup-to-disc ratio is typically expressed as a number between 0 and 1. A CDR value zero indicates that no cupping, and the entire optic disc is uniform. A CDR value one indicates that cup occupies the entire optic disc, indicating significant cupping. In glaucoma, there is progressive damage to the retinal nerve fibers, leading high CDR as the cup becomes larger due to neural loss. A higher cup-to-disc ratio can be indicative of glaucoma or other optic nerve-related abnormalities.

DDLS: It is not affected by the disc size change. A major factor in Glaucoma assessment is neuro retinal rim deterioration. It is possible to assess the initial disc region for DDLS. In neuro retinal rim, binary presentation is achieved by sectors. In subsampling images of configurations nearby optic discs, minimal width of rim region can be calculated. Cup edge points and image edge points can be measured using Euclidean distance.

Segmentation and feature extraction algorithm as follows**:**

Step 1: Image Preprocessing: Preprocess the fundus image includes cropping, resizing, and adjusting brightness/contrast to improve the contrast of blood vessels in the optic nerve head region and other structures using adaptive histogram equalization.

Step 2: Optic Disc Segmentation: To categorize the optic disc province in fundus image, segmentation achieved using U-Net machine learning-based approach, which results a binary mask indicating the boundaries of the optic disc.

Step 3: Optic Disc Features Extraction: Extract relevant features from the segmented optic disc region, features include area, perimeter, circularity, and aspects related to blood vessel distribution and cup-to-disc ratio.

Step 4: Cup-to-Disc Ratio Calculation: Calculate the cup-to-disc ratio (CDR) from the extracted features. It defined as the ratio of cup (depression in the optic disc) to optic disc size. High CDR values are associated with glaucoma and increased risk of vision loss.

Step 5: Optic Nerve Head (ONH) Assessment: Extracted features are combining with CDR, to assess the optic nerve head's condition. Assign a score evolved from the severity of the damage. A higher CDR and specific feature values may indicate a higher likelihood of damage.

Step 6: Disc Damage Likelihood Scale (DDLS) Classification: Use the assessment results to classify the optic nerve head (ONH) condition according to the DDLS scale typically ranges from zero to five, with zero demonstrating no harm and five representing severe harm.

Step 7: **Interpretation and Decision:** Interpret the DDLS classification to determine the likelihood of optic nerve damage due to glaucoma. A glaucoma or healthy classification is made based on this result.

ISNT rule: Describes the typical arrangement of the retinal nerve fiber layer (RNFL) thickness in the optic disc as seen in fundus images. The letters ISNT stands for: I – Inferior, S – Superior, N – Nasal, T – Temporal. ISNT rule suggests that in a healthy eye, the RNFL thickness is usually thickest in the inferior region of the optic disc, followed by superior, nasal, and temporal regions, in that order, this pattern is observed in fundus images. In glaucoma, there is progressive damage to the retinal nerve fibers, leading to characteristic changes in the RNFL thickness. Deviations from the normal ISNT pattern in fundus images can be an indicator of potential glaucoma or other optic nerve abnormalities.

2.3 Fusion techniques

The fusion of fundus image features (CDR, DDLS, ISNT rule) with OCT image radiomics features leverages the strengths of both imaging modalities, enhancing an accuracy and reliability of glaucoma. It enables a more comprehensive assessment of the retinal health by capturing dissimilar characteristics of the disease process and complementing the limitations of each imaging technique. Fusion of radiomics features from OCT and fundus is performed using concatenation operation.

III. GLAUCOMA IMAGE CLASSIFICATION

The Deep Stochastic Variational Autoencoder (DSVAE) represents that a deep learning and probabilistic modeling are combined in the network architectures. It provides significant classification due to its ability to learn hierarchical features, capture uncertainty, perform data augmentation and operate on scenarios to improving the performance and reliable classification data with limited labeling. As follows are the steps in DSVAE process:

- Hierarchical Feature Learning: DSVAE is built upon the Variational Autoencoder (VAE) framework, which is designed to learn compact and meaningful representations of data. A hierarchical feature can be learned automatically from the input images features, extracting both low-level and high-level features that are relevant for distinguishing between different classes.
- Uncertainty Estimation: VAE-based models' advantage is its ability to estimate uncertainty in the learned representations.
- Transfer Learning: DSVAE can ability to capture powerful features like complex and structured patterns in data, these features are used in downstream process. Once the DSVAE is trained, the learned features can be fine-tuned on smaller to improve the classification accuracy.
- Outlier Detection: The probabilistic nature of DSVAE allows it to assign likelihood scores to data points. It can help to identify images which are significantly deviate from the learned distribution, potentially indicating anomalies or rare conditions.

DSVAE consists of an encoder and a decoder, and it aims to learn a generative model that can efficiently encode and decode data while also creating a structured latent space. Encoder takes an input data point x and maps to represent it as a latent space z ,typically outputs the parameters of the distribution over the latent two variables the mean (μ) and logarithmic variance ($\log(\sigma^2)$) of Gaussian distribution over the latent variables z. These parameters will be used to sample the latent code during the training process. The encoder output is given by: μ , $log(\sigma^2) = Encoder(x)$. To obtain the actual latent code z, sample the Gaussian function with mean μ and variance σ^2. This sampling step introduces stochasticity into the model and enables us to back propagate through the sampling process during training. The reparameterization is used to sample from the Gaussian distribution, which involves sampling from a standard Gaussian $ò \sim N(0,1)$ and then transforming it using the mean and

variance: $z = \mu + \sigma \times \grave{o}$. The decoder takes the sampled latent code z and reconstructs the original data x'.The decoder output is given by: $x' = Decoder(z)$. The output of a decoder suffers from reconstruction loss and it depends on the type of data being reconstructed i.e., Mean Squared Error (MSE) loss is often used to measure loss as mentioned Eq.(8)

$$ReconstructionLoss = MSE(x, x') \tag{8}$$

The regularization loss encourages to follow a specific distribution, usually a unit Gaussian function. The Kullback-Leibler (KL) divergence as in Eq.(9) is used to measure the dissimilarity among the learned latent distribution and the target Gaussian distribution. Based on the encoder Gaussian distributions and target Gaussian distributions, the KL divergence is given by:

$$KL\ Divergence = 0.5 \times \sum\left(\sigma^2 + e^{(2 \times log(\sigma^2))} - \mu^2 - 1\right) \tag{9}$$

The total loss mentioned in Eq. (10) is the sum of the reconstruction loss and the regularization loss, it is usually weighted by a hyperparameter β, The regularization of the latent space versus the accuracy of the reconstructions is controlled by this parameter.

$$Total\ Loss = Reconstruction\ Loss + \beta \times KL\ Divergence \tag{10}$$

3.1 Variational Autoencoders with Adam optimizer with equations

To train Variational Autoencoders (VAEs) [27-28] using the Adam optimizer, compute gradients in relation to model parameters and update those parameters accordingly during each training iteration. The Adam optimizer is an adaptive learning rate optimization algorithm that combines the benefits of both RMS and momentum methods. Updating the parameters of a VAE using the Adam optimizer: Initialize the model parameters θ, first and second moment variables, m and v to zero.

Set the hyper parameters: learning rate α, β_1 are the rate at which the first moment exponentially decays, β_2 is rate at which the second moment exponentially decays and $\grave{o}$ is a small constant for numerical stability. For each training iteration (t): Sample a batch of data points, X_{batch}.Compute the reconstruction loss, $R(X_{batch}; \theta)$, and the regularization loss, $KL(X_{batch}; \theta)$. Compute the total loss, $L(X_{batch}; \theta)$ in Eq. (11) is the sum of reconstruction and regularization losses weighted by a hyperparameter β :

$$L(X_{batch}; \theta) = R(X_{batch}; \theta) + \beta \times KL(X_{batch}; \theta) \tag{11}$$

Compute the gradients of the total loss in relation to model parameters: $\nabla\theta(\ L(X_{batch}; \theta))$

Update m and v for each parameter mentioned in Eq. (12) and Eq. (13):

$$m_t = \beta_1 \times m_{t-1} + (1 - \beta_1) \times \nabla\theta\left(L(X_{batch}; \theta)\right) \tag{12}$$

$$v_t = \beta_2 \times v_{t-1} + (1 - \beta_2) \times \left(\nabla\theta\left(L(X_{batch}; \theta)\right)\right)^2 \tag{13}$$

Compute the bias-corrected moments in Eq. (14) and Eq. (15):

$$m_t^h = \frac{m_t}{\left(1 - \beta_1^t\right)} \tag{14}$$

$$v_t^h = \frac{v_t}{\left(1 - \beta_2^t\right)} \tag{15}$$

Apply the Adam update rule to the model parameters in Eq. (16):

$$\theta_t = \theta_{t-1} - \alpha \frac{m_t^h}{\left(\sqrt{(v_t^h)} + \grave{o}\right)} \tag{16}$$

Repeat the process with the next batch until the entire dataset is processed. Continue training the VAE for multiple epochs until convergence or a predefined number of iterations. In the above equations, $\nabla\theta$ denotes the gradient and α, β_1, β_2 and $\grave{o}$ are hyperparameters of the Adam optimizer. The α controls the step size in the parameter updation, and β_1 and β_2 control the both the moments exponential decay rates, respectively. The $\grave{o}$ term is added to ensure numerical stability and avoid division by zero. Typically, hyperparameters are $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$, and $\grave{o} = 1 \times e^{-8}$.

IV. DESCRIPTION OF IMAGE DATASET

The image dataset of fundus and OCT are collected from public Mendeley data base [29], which consists of healthy fundus and OCT of 31 each (31+31=62 in total) and Glaucoma fundus and OCT of 48 each (48+48=96 in total).The total samples are used for experimental analysis are 62+96=158.The fundus image is resized to 256×256 and OCT image to 64×64, which are JPG format color images. Since proposed work is considered feature fusion of both fundus OCT, in this view the one sample consists of two images, so the total sample data consists of 79 =48+31 (48- Glaucoma and 31- Healthy). Experimentally Table 1 illustrates the percentage of training and testing image data utilized for this work.

Table 1: Image Data Folding set

Total number of samples 79				
Data folding	**Train samples**		**Test samples**	
	Glaucoma	**Healthy**	**Glaucoma**	**Healthy**
20%	39	24	9	7
	39+24= 63		9+7=16	
30%	33	22	15	9
	33+22= 55		15+9=24	
40%	30	17	18	14
	30+17=47		18+14= 32	
50%	27	12	21	19
	27+12=39		21+19=40	

2.4 PERFORMANCE MEASURES

The confusion matrix provides valuable insights into the model performance by quantifying different types of predictions. From the confusion matrix, calculate various evaluation metrics, such as accuracy, precision, recall, and F1 score [30], which help assess the model effectiveness in different aspects. As shown in Table 2, a general confusion between two classes.

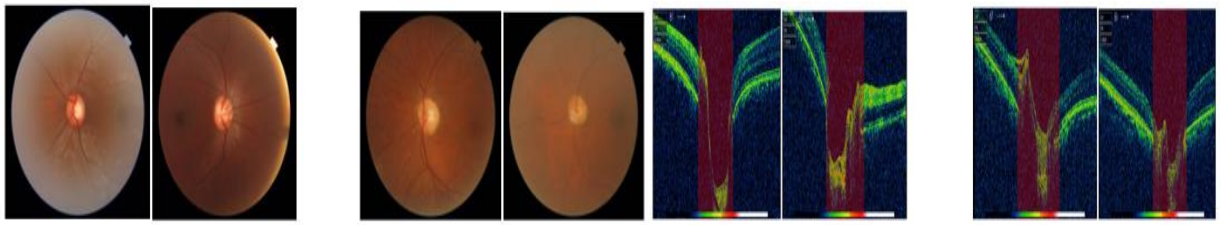
Table 2: Confusion Matrix

	Predicted true positive	Predicted true negative
Actual true positive	True positive (TP)	False positive (FP)
Actual true negative	False negative (FN)	True negative (TN)

V. RESULTS AND ANALYSIS

Fundus images provide a view of the back of the eye, highlighting structures like blood vessels and the optic disc, while OCT images use advanced light-based techniques to create cross-sectional images of retinal layers and other tissues. Both fundus and OCT images are crucial tools in ophthalmology for diagnosing and managing various eye conditions. Figure 3 shows the healthy and glaucoma sample images of fundus and OCT respectively. By observing

the images from both the modalities, it was difficult to distinguish between healthy and glaucoma condition. Thus, computer aided diagnosis is used in this work for an accurate analysis.



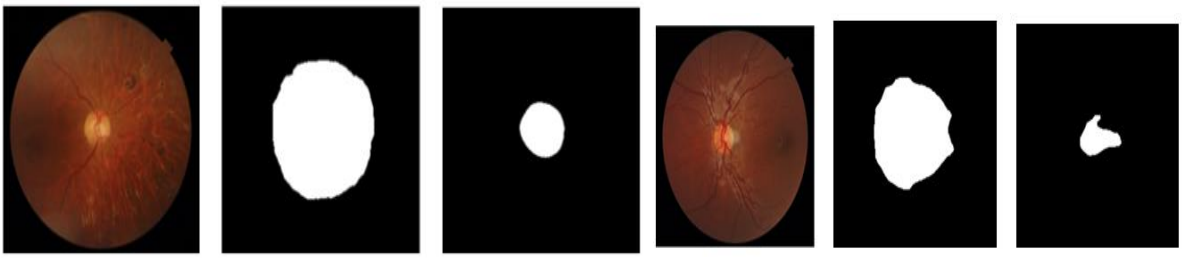
(a) Healthy (b) Glaucoma (c) Healthy (d) Glaucoma

Figure 3: (a-b) Sample set of fundus images, (c-d) OCT images

5.1 Segmentation of fundus

Cup and disc segmentation in fundus images is a critical task in ophthalmology that involves identifying and delineating two key structures within the optic nerve head: the optic disc and the cup. The optic disc is the area on the retina where the optic nerve exits the eye and carries visual information to the brain. It appears as a circular or slightly oval region and is characterized by its pale coloration due to the absence of photoreceptor cells. The optic disc is rich in blood vessels, which enter and exit the eye through the optic nerve. The cup is the central, depressed region within the optic disc. It is lighter in color than the surrounding neuro-retinal rim (the area around the cup) because it lacks the nerve fiber layer and other retinal layers present in the neuro-retinal rim. The cup is a key indicator in diagnosing and monitoring conditions like glaucoma, a disease that involves optic nerve damage is intensified and can lead to vision loss if not managed appropriately.

Deep learning approaches, such as U-Net architectures, have shown promising results due to their ability to capture complex image features. Once the cup and disc are segmented, various quantitative measurements can be extracted, such as the CDR, rim area, cup volume. These measurements provide valuable diagnostic information about the health of the optic nerve and the potential presence of glaucoma. Figure 4 shows that the segmentation of disc and cup for (a) Healthy (b) Glaucoma. In healthy images disc and cup shape are regular where as in glaucoma a deformation is noticed.



(a) Healthy (b) Glaucoma

Figure 4: Segmentation disc and cup of fundus (a) Healthy (b) Glaucoma

Cup segmentation performance using a U-Net architecture can be evaluated using several metrics, including loss, accuracy, and Intersection over Union (IoU). Each of these metrics provides different insights into how well the U-Net model is performing the segmentation task. Metric loss monitors the loss value during training to ensure its decreasing; a lower loss designates as model is converging towards better segmentations. Metric accuracy is a complementary metric if the foreground area is small relative to the background, a high accuracy might not necessarily indicate good segmentation. Metric IoU provides a robust measure of segmentation quality, accounting for true positive and false positive cases, a higher IoU values indicate better performance and closer alignment between predicted and true segmentations. Experimentally the U-Net architecture with Nadam's adaptive learning provides good training and validation performance because both the curves coincide with each other in all the three metrics as shown in Figure 5 and Figure 6 for disc and cup respectively [31]. Since Nadam's adaptive leads to a faster convergence and potentially better results, but the final performance depends on θ_t and α factors, by evaluating the metrics optimizers impact the model ability to accurately segment the optic disc region in fundus images.

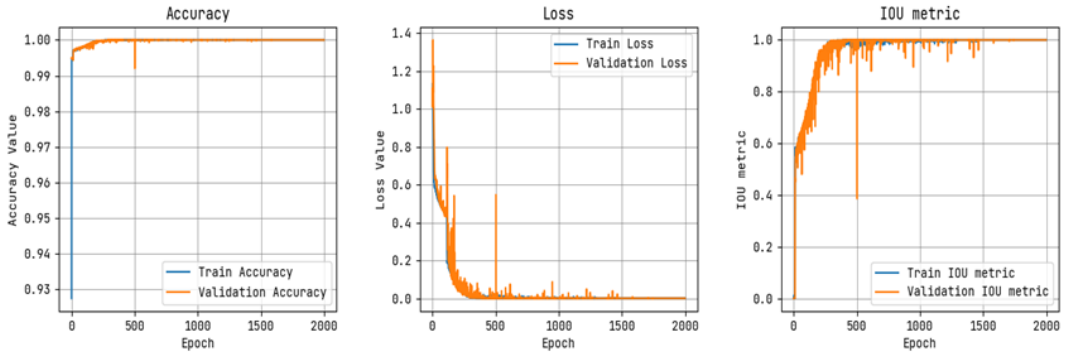


Figure 5: Disc Segmentation Performance

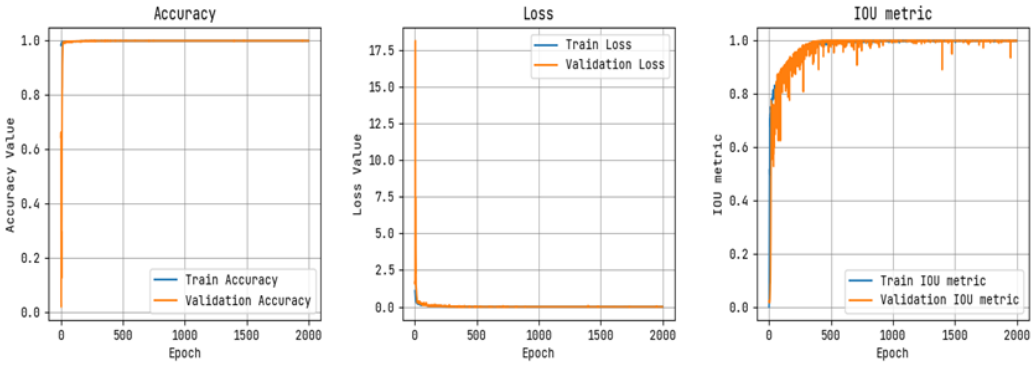


Figure 6: Cup Segmentation Performance

5.2 Features of fundus and OCT

The CDR measures the ratio of the size of the cup (the hollow center of the ONH) to the size of the disc (the entire ONH). As long as the eye is healthy, the cup is relatively small compared to the disc. An increased CDR can be indicative of optic nerve damage, which is a key sign of glaucoma. A CDR greater than 0.6 is often considered suspicious for glaucoma. A higher CDR indicates a greater likelihood of glaucoma or glaucoma progression. The DDLS is a classification system used to grade the severity of optic nerve damage in glaucoma. It assigns a score based on the appearance of the optic nerve head and surrounding structures, such as neuro-retinal rim thinning and notching. A higher DDLS score indicates a greater likelihood of glaucoma damage. The ISNT rule is a guideline for assessing the distribution of neuro retinal rim thickness within the optic nerve head. It is used in conjunction with other assessments to identify asymmetry in neuro-retinal rim thickness, which may be an early sign of glaucoma.

5.3 CLASSIFICATION ANALYSIS

5.3.1 OCT and Fundus training model analysis

OCT radiomics features are extracted from the volumetric data obtained through OCT imaging. These features capture a wide range of information about tissue microstructure and can include texture patterns, intensity distributions, and shape characteristics. These radiomics features are used as input for DSVAE algorithms to predict the input class. The accuracy of the system depends on the quality and quantity of labeled data used for training and validation. Proper model selection and evaluation methodologies, such as cross-validation are essential to ensure that the model generalizes well to new, unseen data. Training a DSVAE model for predict the input class based on features such as Fundus CDR, DDLS, and adherence to the ISNT rule from fundus images is a complex task. Train the model in batches on the training dataset over multiple epochs (iterations through the entire dataset). Monitor the loss and accuracy on both the training and validation datasets during each epoch. Make sure the model isn't underfitting and overfitting on the validation dataset at every iteration after the previous one.

Initially, training accuracy tends to increase steadily as the model learns from the data. Over time, the training accuracy may plateau, indicating that the model is learning less from each epoch. Validation accuracy typically

follows a similar pattern to training accuracy but may be slightly lower. If the model over fitting (i.e., it starts to perform worse on the validation set than on the training set), the validation accuracy curve will diverge from the training accuracy curve. Training loss usually decreases during training as the model gets better at minimizing the defined loss function. In the initial epochs, the loss reduction is often substantial, and it may slow down as training progresses. Validation loss tends to follow the training loss closely initially but may begin to increase if the model over fitting. Early stopping is a technique used to prevent over fitting by monitoring the validation loss. If the validation loss increases, training is stopped to prevent the model from fitting noise in the data. Monitoring training and validation curves are essential to assess the model has sufficiently learned from the data and to prevent overfitting. Optimize hyper parameters like batch size, learning rate and regularization to improve model performance using U-Net architecture with Nadam optimizer. The system providing good stability for 200 epochs for both individual OCT and Fundus features.

5.3.2 Feature fusion Model analysis for training

Monitor both training and validation loss to ensure that the model is effectively learning to encode and reconstruct the input data. Decreasing or stable loss values are indicators of progress. Accuracy might be relevant if you're using the DSVAE for a classification task within the latent space. Keep in mind that DSVAEs are particularly useful for unsupervised learning, data compression, and generative tasks, so the primary focus is often on the quality of data reconstruction and the distribution of latent space representations. Visualizing the training and validation loss curves over epochs are make availability of the model insight convergence with generalization. If both losses decrease or stabilize, it suggests that the DSVAE is successfully learning meaningful representations of the data. However, it's important to consider the specific task and purpose of the DSVAE to interpret these metrics appropriately. Figure 7 represents a system for 100 epochs that achieves same amount of training loss.

Figure 7: Epoch=100 system remains achieves same amount of training loss

5.4 Model Analysis For Testing

5.4.1 Fundus features model

A well-designed system DSVAE accurate algorithms for CDR, DDLS, and adherence to the ISNT rule. The accuracy of the system can be evaluated using various metrics, including sensitivity, specificity, precision, and the receiver operating characteristic (ROC) curve. These metrics provide insights into how well the system can correctly identify individuals with glaucoma or at risk of glaucoma and rule out those who are healthy. The accuracy of these systems varies, but they can be highly sensitive and specific when properly trained and validated. The accuracy can range from 67% to 78%, depending on the dataset folding. Table 3 shows Confusion matrix for different folding and Table4 shows system performance for different data folding in percentage and Figure8 represents confusion matrix of 40% folding for fundus images.

Table 3: Confusion matrix for different folding

Total number of samples 79				
Data Folding	**TP**	**FP**	**TN**	**FN**
20%	6	3	5	2
30%	12	3	5	4
40%	15	3	10	4
50%	15	6	12	7

Table 4: System performance for different data folding in percentage

Total number of samples 79					
Data Folding	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AU-ROC**
20	68.75	66.66	75.00	70.58	71.56
30	70.83	80.00	75.00	77.41	76.89
40	78.12	83.33	78.94	81.07	82.78
50	67.50	71.42	68.18	69.76	69.34

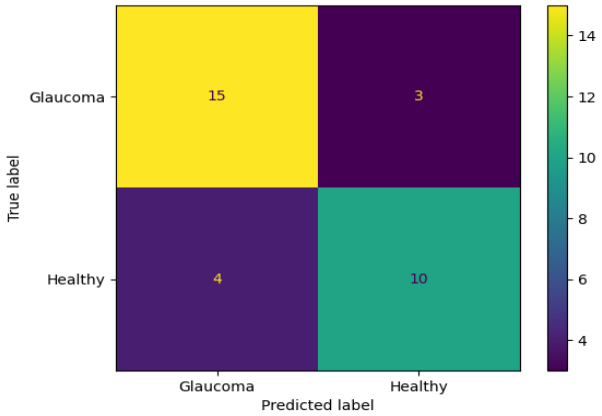


Figure 8: Confusion matrix of 40% folding for fundus images.

5.4.2 OCT features model

Using DSVAE to learn meaningful representations of radiomics features and then applying a classification head to classify OCT images based on these features. Accuracy measures overall correctness, while the F1 score endow with a sense of balance among precision and recall. In medical image analysis, which is useful where FP and FN can have critical implications. Table 5 shows Confusion matrix for different folding; Table 6 shows system performance for different data folding in percentage and Figure 9 represents confusion matrix of 40% folding for OCT images.

Table 5: Confusion matrix for different folding

Total number of samples 79				
Data Folding	**TP**	**FP**	**TN**	**FN**
20%	6	3	6	1
30%	13	2	7	2
40%	16	2	11	3
50%	18	3	15	4

Table 6: System performance for different data folding in percentage

Total number of samples 79					
Data Folding	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AU-ROC**
20	75.00	66.66	85.71	74.99	80.12
30	83.33	86.66	86.66	86.66	86.45
40	84.37	88.88	84.21	86.48	86.56
50	82.50	85.71	81.81	83.71	83.67

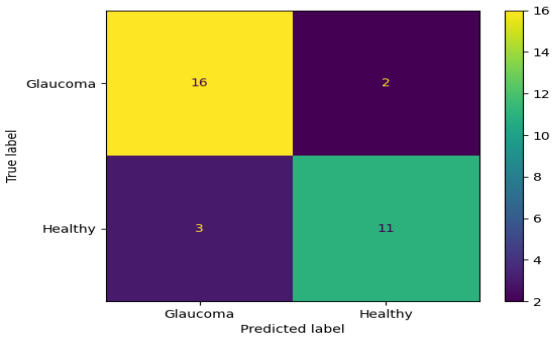


Figure 9: Confusion Matrix of 40% Folding for OCT Images

5.4.3 Fusion of OCT and Fundus features model

Testing accuracy in the context of a DSVAE, testing accuracy might be relevant if you've trained the model for a specific classification task using the latent space representations. The proposed work used the DSVAE to learn features that help classify certain classes and compute the testing accuracy by comparing the predicted classifications based on the latent space representations to the true labels. Cross-validation is a technique used to assess the model performance on multiple subsets of the data to ensure reliable generalization, which is a cross-validation using k-fold subsets or folds is a common approach. Each fold serves as the validation set, and the model is trained and evaluated on k-fold dataset, the training set consists of the remaining folds. Table 7 shows Confusion matrix for different folding and Table 8 shows system performance for different data folding in percentage for fusion.

Table 7: Confusion matrix for different folding

Total number of samples 79				
Data Folding	**TP**	**FP**	**TN**	**FN**
20%	9	0	6	1
30%	15	0	8	1
40%	18	0	13	1
50%	21	0	17	2

Table 8: System performance for different data folding in percentage

Total number of samples 79					
Data Folding	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AU-ROC**
20	93.75	95.00	02.86	95.32	94.44
30	95.83	96.88	95.24	95.45	95.93
40	96.88	97.37	96.43	96.80	97.22
50	95.00	95.65	94.74	94.95	96.99

Using the Adam optimizer, cross-validation, and confusion matrices together provide a comprehensive assessment of a DSVAE's classification performance across different subsets of the data. This approach helps you understand the model's generalization capabilities and identify potential areas for improvement. Figure 10 represents confusion matrix of 40% folding for fusion

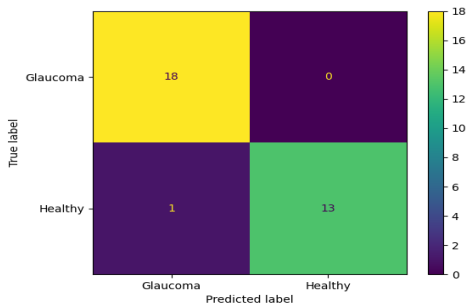


Figure 10: confusion matrix of 40% folding for fusion images

The confusion matrix provides insights into how well the DSVAE's predictions align with the true labels for each class. As determined by the confusion matrix, recall, precision and F1-score, metrics are calculated for individual classes, which give a more detailed understanding of the model performance across different classes. Overall Performance of the system by considering all metrics together to form a comprehensive understanding of the model's classification performance. A well-rounded model should have high accuracy, F1-score, and recall while maintaining a reasonable balance between precision and recall. Evaluating a DSVAE's performance with different data folds using the Adam optimizer requires considering multiple metrics like accuracy, F1-score, recall, and sensitivity. These metrics provide insights into how well the model is classifying instances, especially when classes are imbalanced or when false positives and false negatives have varying consequences. Cross-validation helps ensure that the model performance is consistent across various subsets of the data.

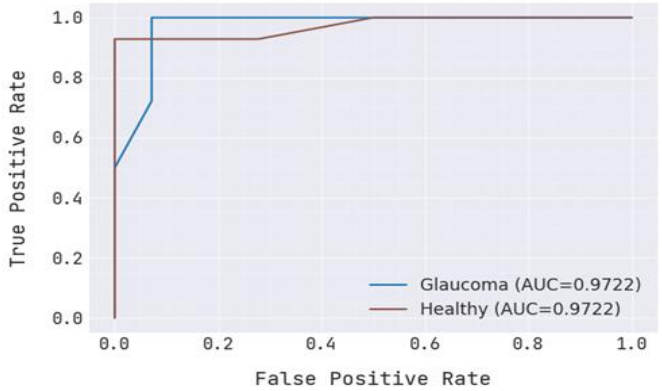


Figure 11: AU-ROC Curve

AU-ROC (Area Under the Receiver Operating Characteristic Curve) is a performance metric commonly used in binary classification problems, especially in the context of imbalanced datasets. The ROC curve is a graphical representation of the trade-off between the true positive rate (sensitivity) and the false positive rate (1-specificity) as the classification threshold varies. The AU-ROC value summarizes the overall performance of the classifier regardless of the chosen threshold. In the context of a DSVAE with an AU-ROC curve, evaluating the performance of the model in a binary classification task. The DSVAE might be used to generate representations of data, and these representations are then used as input features for a binary classifier. The AU-ROC curve would depict how well the classifier, trained on these representations, can discriminate between the two classes is as shown in Figure 11. A higher AU-ROC value indicates better discrimination performance.

VI. CONCLUSION

In this study, set to enhance the accuracy and reliability of glaucoma classification by leveraging the synergistic potential of fundus photographs and Optical Coherence Tomography (OCT) scans through the application of optimized deep learning neural networks. Our findings demonstrate the successful fusion of these distinct modalities, leading to a significant improvement in glaucoma detection performance. The development of a novel multi-path neural network architecture allowed us to effectively extract complementary features from both fundus and OCT images. This fusion of structural and textural information enabled the model to capture intricate patterns, facilitating a more comprehensive understanding of the disease. A pivotal aspect of our approach was the integration of advanced optimization techniques, notably the Adam optimizer, into the neural network training process. By refining the network's parameters and accelerating convergence, these techniques contributed to the model's enhanced generalization capability. Overfitting was mitigated, and the resultant classifier demonstrated robust performance across diverse patient cohorts, bolstering its clinical applicability. In the evaluation phase, the proposed fusion-based classification model exhibited impressive accuracy, sensitivity, and specificity. The utilization of quantitative metrics, including the Area Under the Receiver Operating Characteristic Curve (AU-ROC), reaffirmed the model's capability to discriminate between glaucoma cases and healthy individuals with precision. Notably, this approach outperformed existing methods, underscoring its potential to revolutionize glaucoma diagnosis. Looking forward, this work paves the way for further research, innovation, and collaboration to continually refine and advance the boundaries of glaucoma diagnosis and patient care.

ACKNOWLEDGEMENTS

We thank all the authors who contributed to the research on glaucoma diagnosis and classification.

ETHICS DECLARATIONS

CONFLICT OF INTEREST:

On behalf of all the authors, the corresponding author states that there is no conflict of interest.

FUNDING:

There are no funding agencies related to this work.

DATA AVAILABILITY

Public Mendeley Database for study on Glaucoma, Data shall be provided upon reasonable request.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

[1] Maheshwari, Devendra, Swathi Kanduri, Mohideen A. Kadar, Rengappa Ramakrishnan, and Madhavi R. Pillai. "Midterm outcome of mitomycin C augmented trabeculectomy in open angle glaucoma versus angle closure glaucoma." Indian Journal of Ophthalmology, (2019): 1080.

[2] Tham, Yih-Chung, Xiang Li, Tien Y. Wong, Harry A. Quigley, Tin Aung, and Ching-Yu Cheng. "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis." Ophthalmology 121, no. 11 (2014): 2081-2090.

[3] McMonnies CW. Glaucoma history and risk factors. Journal of optometry. 2017 Apr 1;10(2):71-8.

[4] Devalla, Sripad Krishna, Khai Sing Chin, Jean-Martial Mari, Tin A. Tun, Nicholas G. Strouthidis, Tin Aung, Alexandre H. Thiéry, and Michaël JA Girard. "A deep learning approach to digitally stain optical coherence tomography images of the optic nerve head." Investigative ophthalmology & visual science 59, no. 1 (2018): 63-74.

[5] Orlando, J.I., Prokofyeva, E., del Fresno, M. and Blaschko, M.B., 2017, January. Convolutional neural network transfer for automated glaucoma identification. In 12th international symposium on medical information processing and analysis (Vol. 10160, pp. 241-250). SPIE.

[6] Kaveya.S, Syedhusain.S, Revathi.T, Sathiya Priya.S, Subiksha.S, Sushma.M. (2020). Glaucoma Detection Using UNET Model. International Journal of Advanced Science and Technology, 29(3), 9089 – 9095.

[7] Yu, Shuang, Di Xiao, Shaun Frost, and Yogesan Kanagasingam. "Robust optic disc and cup segmentation with deep learning for glaucoma detection." Computerized Medical Imaging and Graphics 74 (2019): 61-71.

[8] M. U. Muthmainah, H. A. Nugroho, and B. Winduratna, “Glaucoma classification based on texture and morphological features,” in Proceedings of the 2019 5th International Conference on Science and Technology (ICST), IEEE, Yogyakarta, Indonesia, 30-31 July 2019.

[9] Ahn, Jin Mo, Sangsoo Kim, Kwang-Sung Ahn, Sung-Hoon Cho, Kwan Bok Lee, and Ungsoo Samuel Kim. "A deep learning model for the detection of both advanced and early glaucoma using fundus photography." PloS one 13, no. 11 (2018): e0207982.

[10] Salam, A.A., Khalil, T., Akram, M.U., Jameel, A. and Basit, I., 2016. Automated detection of glaucoma using structural and non-structural features. Springer plus, 5, pp.1-21.

[11] Esengönül, Meltem, and António Cunha. " Glaucoma Detection using Convolutional Neural Networks for Mobile Use." Procedia Computer Science 219 (2023): 1153-1160.

[12] F.Demir and B. Taşcı, " An Effective and Robust Approach Based on R-CNN+LSTM Model and NCAR Feature Selection for Ophthalmological Disease Detection from Fundus Images," Journal of Personalized Medicine, vol. 11, no. 12, p. 1276, 2021.

[13] T.I.Chayan, "Decipherable Classification of Glaucoma using Deep Neural Network Leveraging XAI," Doctoral dissertation, Brac University, 2022.

[14] Mahum, Rabbia, Saeed Ur Rehman, Ofonime Dominic Okon, Amerah Alabrah, Talha Meraj, and Hafiz Tayyab Rauf. "A novel hybrid approach based on deep CNN to detect glaucoma using fundus imaging." Electronics 11, no. 1 (2021): 26.

[15] P.Wang, J. Shen, R. Chang, M. Moloney, M. Torres, B. Burkemper, et al., "Machine learning models for diagnosing glaucoma from retinal nerve fiber layer thickness maps," Ophthalmology Glaucoma, vol. 2, no. 6, pp. 422-428, 2019.

[16] L.Pascal, O. J. Perdomo, X. Bost, B. Huet, S. Otálora, and M. A. Zuluaga, "Multi-task deep learning for glaucoma detection from color fundus images," Scientific Reports, vol. 12, no. 1, pp. 1-10, 2022.

[17] https://pyradiomics.readthedocs.io/en/latest/features.html PyRadiomics extracted various radiomic features.

[18] J.Qian, M. G. Herman, D. H. Brinkmann, N. N. Laack, B. J. Kemp, C. H. Hunt, et al., "Prediction of MGMT status for glioblastoma patients using radiomics feature extraction from 18F-DOPA-PET imaging," International Journal of Radiation Oncology* Biology* Physics, vol. 108, no. 5, pp. 1339-1346, 2020
[19] Dozat, T. "Incorporating Nesterov momentum into ADAM, in proceedings of the 4th International Conference on Learning Representations (ICLR 2016)." San Juan, Puerto Rico (2016): 2-4.
[20] Zunair, Hasib, and A. Ben Hamza. "Sharp U-Net: Depthwise convolutional network for biomedical image segmentation." Computers in Biology and Medicine 136 (2021): 104699.
[21] García-Ordás, M. T., Benítez-Andrades, J. A., García-Rodríguez, I., Benavides, C., & Alaiz-Moretón, H. (2020). Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. Sensors, 20(4), 1214.
[22] Kumar, E. Sudheer, and C. Shoba Bindu. "Two-stage framework for optic disc segmentation and estimation of cup-to-disc ratio using deep learning technique." Journal of Ambient Intelligence and Humanized Computing (2021): 1-13.
[23] Mohammed, F. G., S. D. Athab, and S. G. Mohammed. "Disc damage likelihood scale recognition for Glaucoma detection." Journal of Physics: Conference Series. Vol. 2114. No. 1. IOP Publishing, 2021.
[24] R. U. Singh and S. Gujral, "Assessment of DDLS (DDLS) for automated glaucoma diagnosis," Procedia Computer Science, vol. 36, pp. 490-497, 2014.
[25] Lee, Yoon Pyo, et al. "ISNT rule satisfaction in Korean non-glaucomatous subjects." European Journal of Ophthalmology 31.1 (2021): 125-129.
[26] K.T.M.Han,P. Boonsieng, W. Kongprawechnon, P. Vejjanugraha, W. Ruengkitpinyo, and T. Kondo, "An Automated Framework for Screening of Glaucoma using Cup-to-Disc Ratio and ISNT Rule with a Support Vector Machine," Trends in Sciences, vol. 19, no. 9, pp. 3971-3971, 2022.
[27] Li, Peizheng, et al. "Variational Autoencoder Assisted Neural Network Likelihood RSRP Prediction Model." 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2022.
[28] Wei, Ruoqi, Cesar Garcia, Ahmed El-Sayed, Viyaleta Peterson, and Ausif Mahmood. "Variations in variational autoencoders-a comparative evaluation." IEEE Access 8 (2020): 153651-153670.
[29] https://data.mendeley.com/datasets/trghs22fpg
[30] https://towardsdatascience.com/understanding-confusion-matrix-precision-recall-and-f1-score-8061c9270011
[31] Nanditha Krishna, K. Nagamani " Multi-Modal Imaging-Based Feature Fusion for Accurate Glaucoma Diagnosis with Deep Learning" Research square, preprint https://www.researchgate.net/publication/370979431