Optimizing glaucoma diagnosis using fundus and optical coherence tomography image fusion based on multi-modal convolutional neural network approach

Nanditha Krishna[1,2], Nagamani Kenchappa[3]

[1]Department of Electronics and Telecommunication Engineering, RV College of Engineering, affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Department of Medical Electronics Engineering, Dayananda Sagar College of Engineering, Bengaluru, India
[3]Department of Electronics and Telecommunication Engineering, RV College of Engineering, Bengaluru, India

Article Info

Article history:

Received Feb 2, 2024
Revised Mar 20, 2024
Accepted Apr 2, 2024

Keywords:

Deep learning network
Fundus images
Glaucoma
Multi modal convolutional neural network
Optical coherence tomography images

ABSTRACT

A novel approach that combines segmented fundus images (FIs) and optical coherence tomography image (OCTIs) are presented here, by incorporating deep learning network (DLN) techniques, to address the imperative need for advanced diagnostic algorithms in detecting and classifying glaucoma. By combining these two images, glaucoma diagnoses are made to improve the accuracy with more reliability. Multi modal convolutional neural networks (MMCNNs) are proposed for automatically extracting discriminatory features from both segmented FIs and OCTIs, allowing for comprehensive ocular analysis. A significant improvement in glaucoma diagnosis is achieved through segmentation of both FIs and OCTIs, ensuring robustness generalization to diverse clinical scenarios, DLN models are trained on datasets encompassing a wide range of glaucoma cases. The integrated approach outperforms individual modalities in terms of early detection of glaucoma and accurate classification. This method demonstrates promising potential in early glaucoma detection due to its effectiveness. By combining segmented features from both FIs and OCTIs through MMCNNs, improved efficiency in diagnosing predominant ocular glaucoma disorder is achieved compared to existing methods. Within the scope of this research, GoogLeNet (GN) is applied to independently classify glaucoma (uni-modal) in segmented FIs and OCTIs, providing a basis for comparison with the evaluation of MMCNNs.

Corresponding Author:

Nanditha Krishna
Department of Electronics and Telecommunication Engineering, RV College of Engineering, Bengaluru, affiliated to Visvesvaraya Technological University
Belagavi, Karnataka, India
Email: nanditha13@gmail.com

1. INTRODUCTION

Glaucoma, characterized by its progressive nature, poses a risk of permanent vision impairment if not identified and managed promptly. Recognizing and preventing glaucoma early is essential in addressing this severe eye ailment, which can lead to vision loss and irreversible blindness. Traditional diagnostic approaches for glaucoma are time-consuming, subject to human error, and inefficient, emphasizing the need for automated diagnosis to improve precision and streamline the process. Multiple research articles have been published, each proposing different approaches to detect glaucoma. Introduces deep learning network (DLN) technology for early glaucoma detection, utilizing U-Net combined with transfer learning frameworks for optic cup segmentation [1]. DenseNet-201 is employed for feature extraction from retinal fundus images (FIs). The proposed model achieves high accuracy rates, with 98.82% accuracy rate (AR) during training. A comparison is made with convolutional neural network (CNN) based deep learning (DL) approaches, highlighting the efficacy of the proposed method in identifying glaucoma before symptoms manifest, potentially improving patient diagnosis and management. The transfer learning technique offers versatility across diverse fields like manufacturing, transportation engineering, and industrial imaging. Further it can explore the integration of fuzzy and semi-supervised techniques to enhance its applicability and effectiveness. CNN's offer a robust method for automated glaucoma diagnosis, leveraging a dataset of 1113 FIs [2]. A 13-layer CNN algorithm implemented via Google Colab facilitated the process. The dataset was partitioned into 70% for training, 20% for validation, and 10% for testing to evaluate the algorithm's performance. The model achieved impressive metrics, including sensitivity rate (SER) of 85.42%, specificity rate (SPR) of 100%, and precision rate (PR) of 100%, demonstrating its effectiveness in diagnosing glaucoma accurately. Enhancing performance involves integrating advanced transfer learning technology to refine classification outcomes. By adopting upgraded approaches, better classification results can be achieved. Various machine learning algorithms, such as support vector machines (SVMs), neural networks (NNs), and adaptive neuro fuzzy inference system (ANFISs), are utilized for glaucoma detection based on FIs [3]. The ANFIS demonstrates strong performance, achieving a PR of 97.56%, RR of 97.81% in glaucoma detection. Disadvantages of ANFIS machine learning classifiers include their complexity in model design and optimization, which can demand substantial expertise and computational resources. ANFIS models may lack interpretability compared to simpler algorithms, making it challenging to understand their decision-making process. DCNN and fusion classifiers are developed for fundus image stage classification, utilizing pre-trained models like ResNet50 and DenseNet-201 [4]. The model is evaluated across four public datasets: ACRIMA, RIM-ONE, HVD, and Drishti. Classifier fusion via maximum voting is employed to combine CNN models. Results show promising performance with 97.57% AR on the ACRIMA dataset, 97% AUC on HVD. To enhance the glaucoma diagnosis model by incorporating larger, more diverse datasets to improve generalizability. Integration of clinical data such as patient age, sex, and medical history could enrich diagnostic accuracy and relevance.

The three pre-trained CNNs-ResNet, VGGNet, and GoogleNet-for glaucoma FIs categorization [5]. Performance evaluation on five datasets demonstrates the model. The ensemble architecture achieves 91.11% AR, 85.555% SER, and 95.20% SPR on PSGIMSR, while achieving 95.63% AR on DRIONS-DB, 98.67% AR on HRF, 95.64% AR on DRISHTI-GS, and an overall AR of 88.96% on combined datasets. The deep architecture of these networks may pose challenges in interpretability, making it difficult to understand how specific features contribute to classification decisions. Fine-tuning pre-trained CNNs for new tasks may be intensive and require expertise in hyperparameter tuning. Utilizing optical coherence tomography image (OCTIs), glaucoma diagnostics are enhanced through a novel cross-sectional i.e., segmented optic nerve head (ONH) feature. DL techniques are employed, focusing on the four most significant features identified through statistical analysis [6]. Automated glaucoma detection using DL yields optimal results, achieving a highly accurate AUC of 0.98 and 98.6% AR on test data. The accuracy of the segmented ONH cup area is critical; any inaccuracies in this process could lead to erroneous diagnoses and compromise the reliability of the model. Utilizing FIs, an automatic glaucoma diagnosis system is developed based on an adapted version of GoogleNet [7]. The system employs a sliding-window approach in conjunction with the network, incorporating manually extracted FI structures and region of interest (ROI) sub-images for training. Following training, the algorithm demonstrates good accuracy even when using images from data augmentation or databases of poor quality. Implementing pre-processing steps for database images adds computational overhead, potentially slowing down the overall workflow and increasing resource requirements. The model evaluates various CNN schemes to assess the impact of dataset size, architecture, and transfer learning on glaucoma classification performance [8]. It investigates the influence of both FIs and clinical history data from patients. Employing transfer learning with VGG19, the model achieves notable metrics, including an AUC of 0.94, SER of 93%, and SPR of 93.18%, highlighting the effectiveness of the approach in enhancing glaucoma detection accuracy. Assessments of architectural approaches are crucial to address evolving challenges and improve the performance of computer-aided diagnosis (CAD) systems for FIs analysis in glaucoma detection.

Utilizing CNN in conjunction with pre-trained efficientNet-b0 [9], glaucoma detection is performed on FIs with or without demographics information. The model trained on 1539 FIs achieves notable metrics including a 0.98 AUC, 86% AR, 91% PR, 86% SER, and 0.86 F1-score without demographic information. Incorporating demographics information, the model achieves a 0.98 AUC, 87% AR, 87% PR, 87% SER, and 0.87 F1-score. Lack of comprehensive comparison between deep learning systems (DLS) using red-free color FIs, potentially overlooking important differences in performance and diagnostic accuracy. Advanced classification, segmentation, and detection methods for glaucoma utilizing FIs have been developed [10], notably employing deep learning convolutional neural networks (DLCNNs) and their variants. The architecture for glaucoma classification incorporates ImageNet pre-trained Deep CNN architectures, including InceptionResNetV2 and NasNet-Large. NasNet-Large exhibits superior performance with metrics such as 98.1% SER, 98.4% SPR, 98.3% AR, and an AUC of 0.97, showcasing its effectiveness in glaucoma detection. Training more robust glaucoma classifiers using deep features may exacerbate the risk of overfitting, particularly if the models are not properly regularized or validated on diverse datasets. A CNN-based architecture is employed to differentiate between glaucoma and non-glaucoma patterns in FIs [11]. The hierarchical structure of FIs is utilized within the CNN framework, consisting of six layers with dropout mechanisms. Evaluation on the SCES and ORIGA datasets yields high accuracy rates of 95.67% and 96%, respectively, demonstrating the effectiveness of the proposed approach in glaucoma pattern classification. Deep CNNs are susceptible to overfitting, especially with small datasets, which may lead to poor generalization performance on unseen data. Glaucoma classification into mild or severe types using FIs is investigated [12], employing 3,200 images from the MESSIDOR dataset. Six deep learning models, including VGG16, VGG19, InceptionV3, InceptionResNetV2, ResNet50, and DenseNet169, are utilized for classification. Among these models, DenseNet169 achieves an accuracy rate (AR) of 85.19%, demonstrating its effectiveness in distinguishing between mild and severe glaucoma. DenseNet169 is a deep neural network with a large number of parameters, making it computationally intensive to train and requiring significant computational resources.

In a study, 940 FIs along with clinical notes and demographic data [13] were collected from an Eye Hospital, focusing on glaucomatous optic neuropathy (GON). VGGNet is employed to detect GON in FIs, with a SVM integrated for final classification when the CNN provides low confidence ratings. Advanced scoring of GON is utilized to minimize missed instances, and the CNN classifier's performance is evaluated using datasets from TVGH and Drishti-GS. A multi-task DL model is developed based on the similarities between eye-fundus tasks and measurements used in glaucoma diagnosis [14]. This model is designed to simultaneously learn various segmentation and classification tasks, leveraging the shared characteristics among them. Trained on 1,200 images from diverse sources within the retinal fundus glaucoma dataset, the model outperforms other multi-task learning approaches, demonstrating improved effectiveness compared to training each task independently. With multiple tasks being learned simultaneously, there is an increased risk of overfitting, particularly if the model is not properly regularized or if tasks have varying levels of data availability or complexity. DLCNN and machine learning (ML) methods [15] are employed to develop three distinct approaches for glaucoma detection. Features such as cup-to-disc ratio (CDR) and rim-to-disc ratio (RDR) are computed from FIs and classified using SVMs and the k-nearest neighbor strategy. Notably, the VGG-16 DLCNN achieves an impressive accuracy rate, showcasing its effectiveness in glaucoma detection compared to traditional ML methods. SVMs and k-nearest neighbor strategies rely on hyperparameter tuning for optimal performance, requiring careful selection and validation to prevent overfitting or underfitting, which can be time-consuming and computationally intensive. Efficient Net is evaluated for glaucoma disease classification using transfer learning techniques [16]. Public datasets including RIM-ONE V2 and V3, ORIGA, DRISHTI-GS1, HRF, and ACRIMA are utilized for model comparison against frameworks such as VGG16, InceptionV3, and Xception. EfficientNetB4 emerges as the top-performing model, surpassing others with a remarkable 98.38% accuracy. Additionally, it achieves superior results for metrics like F1-score, Kappa score, and area under the curve (AUC) compared to alternative methods. These models may lack the adaptability to capture intricate patterns and variations in diverse datasets, potentially leading to suboptimal performance when faced with complex data distributions or novel classes. The study evaluates CAD frameworks with DLN for unbiased analysis of glaucoma patients [17]. It discusses recent advancements in AI techniques, highlighting the superiority of DL over traditional machine learning in eliminating the need for manual feature extraction. CAD frameworks, aiding social professionals in glaucoma screenings, achieve a high accuracy rate of 96% in glaucoma classification, demonstrating their potential utility in examining various eye conditions during glaucoma testing. CAD frameworks with DLN require large and diverse datasets for effective training, which may be challenging to acquire and annotate, particularly for rare conditions or specific patient populations. A computationally lightweight CNN system is developed [18] for automated glaucoma detection, optimized for efficiency. Utilizing color fundus photographs, the CNN efficiently identifies glaucoma indicators, aiming to streamline diagnosis and enable earlier intervention, crucial for preventing irreversible blindness. Regular updates and refinement of the model is necessary to maintain its effectiveness. Utilize deep learning algorithms proficient with hybrid in recognizing the intricate features essential for classification tasks [19], which encompass microaneurysms, exudate, and retinal hemorrhage. Exploit the predictive capabilities [20] of artificial intelligence (AI), introduce two deep learning (DL) networks based on vision transformers (ViT). Initially, refine a spatiotemporal ViT to categorize the rate of glaucoma progression (GP) utilizing only three baseline visual fields (VFs). The investigation spans the threshold mean deviation (MD) rate of change from -0.3 to -1.5 dB/year, achieving an impressive 89% accuracy in detecting GP.

Glaucoma, a leading cause of irreversible blindness worldwide, necessitates early and accurate diagnosis for effective management. While FIs and OCTIs are routinely utilized for glaucoma diagnosis, each modality offers unique insights into ocular pathology. However, integrating information from both modalities to enhance diagnostic accuracy remains a challenge. To address these challenges, this paper focuses an optimized glaucoma diagnosis framework utilizing FIs and OCTIs fusion based on a multi-modal CNN approach. This paper presents designing a glaucoma detection system focused on early diagnosis, utilizing a fusion approach with both FIs and OCTIs. The research work highlights how combining data from different imaging modalities enhances diagnostic accuracy for glaucoma detection. The multi modal convolutional neural networks (MMCNNs) analysis is applied in this work to integrate the segmented FIs and OCTIs to intensify the precision of glaucoma diagnosis. Utilizing segmented images from both FIs and OCTIs substantially enhances diagnostic accuracy. The integration of multiple imaging modalities, allows for a more comprehensive analysis by capturing complementary information. This fusion of features enhances the model ability to detect subtle abnormalities associated with glaucoma. When segmented FIs and OCTIs are combined using MMCNNs, SER and SPR are significantly enhanced, enabling significantly more AR in detection and classification than when detected independently using GoogLeNet (GN).

2. DETAILED DESCRIPTION OF DATASETS

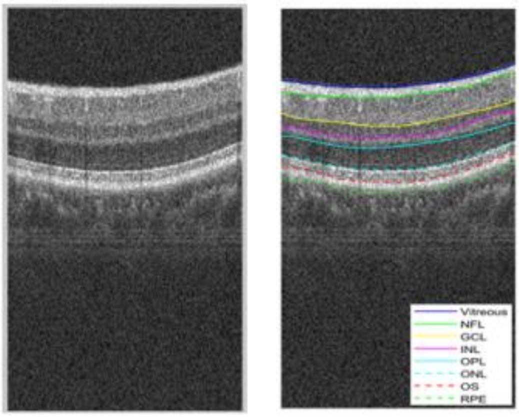
Datasets of FIs and OCTIs are essential for developing and testing algorithms for diagnosis. Kaggle publicly available datasets includes 1000 FIs and OCTIs normal and 1000 FIs and OCTIs glaucoma diseased eyes. The 2D images size 500×500. An optic nerve and retina are visible in fundus photography, while OCT provides detailed cross-sectional images of the retina. These datasets are valuable for studying vascular abnormalities and diseases. FIs involves capturing images of the back of the eye, including the retina, optic disc, and blood vessels, provides image texture feature information for various eye conditions such as glaucoma. OCTIs is a non-invasive imaging technique that captures high-resolution cross-sectional images with respect to retinal layers (RLs). The OCTIs are comprehensively used to diagnose and monitor RLs diseases, including glaucoma.

3. PROPOSED RESEARCH METHOD

Designing an efficient fusion method to integrate FIs and OCTIs information for enhanced glaucoma diagnostic accuracy. Developing a multi-modal CNN architecture capable of effectively perform fusion features for glaucoma diagnosis. Evaluating the performance of the proposed approach against existing methods using diverse datasets to demonstrate its effectiveness in terms of accuracy, sensitivity, specificity, and computational efficiency. The introduction of MMCNNs represents a significant leap in automated feature extraction from segmented FIs and OCTIs, enabling a holistic analysis of ocular health. By integrating features from both images remarkable advancement is achieved in the diagnosis of glaucoma, offering unparalleled accuracy and depth in ocular assessment. The robustness and adaptability of DLN models, trained on diverse glaucoma datasets, ensure that the proposed MMCNNs can effectively handle a wide range of clinical scenarios, fostering confidence in their diagnostic capabilities. With the integration of data from segmented FIs and OCTIs, the MMCNNs not only enhance the early detection of glaucoma but also provide a comprehensive understanding of ocular pathology, revolutionizing the field of ophthalmic diagnostics. Through the synergistic combination of advanced neural network architectures and multi-modal imaging techniques, the proposed MMCNNs pave the way for more accurate, efficient, and versatile glaucoma diagnosis, offering hope for improved outcomes and vision preservation. This paper presents a uni-model, utilizing GN architecture independently for FIs and OCTIs, to validate MMCNNs.

3.1. Retinal layers segmentation in OCTIs

Segmenting retinal layers (RLs) [21] in OCTIs are a critical step in the reasoning of inflammation in eye, such as glaucoma. Accurate segmentation allows extraction of essential information about the various RLs aiding in disease diagnosis. There is degradation in OCTIs scan quality and resolution. This work applied interference reduction strategy as a median filtering, to reduce the interference and to improve perceptibility using histogram equalization. Segmentation of RLs exploitation the technique as Sobel edge detection. Detect edges in the OCTIs and use them as boundaries for retinal layers and compute gradients to find boundaries between layers. Applied adaptive thresholding techniques to separate different layers based on intensity that evolve to the boundaries of the RLs with energy minimization. The layers are as shown in Figure 1(a) is the input image to system and Figure 1(b) OCTI is the segmented into 7-layers. The process involves input an OCT image into the system, as shown in Figure 1(a). This image is then subjected to segmentation, dividing it into seven distinct layers, as depicted in Figure 1(b). The segmentation allows for the detailed analysis and characterization of each layer, providing valuable insights into the tissue structure and pathology. This method enables precise identification and measurement of various features within the OCT image. The seven-layer OCT image typically reveals distinct anatomical features within different layers of tissue. The vitreous is a transparent gel-like substance that fills the central cavity of the eye between the lens and the retina. The layer features include the retinal nerve fiber layer (NFL), representing axons of retinal ganglion cells. The layer ganglion cell layer (GCL), contains cell bodies of ganglion cells. The nuclear layer (INL) is the cell bodies of bipolar, horizontal, and amacrine cells. The outer plexiform layer (OPL) contains synapses between photoreceptors and bipolar cells. The outer nuclear layer (ONL) contains the cell bodies of photoreceptors. The outer segment (OS) layer typically exhibits moderate reflectivity compared to other retinal layers due to the presence of the tightly packed photoreceptor outer segments. Finally, the retinal pigment epithelium (RPE) layer, the outermost layer, contains pigmented cells that support and nourish the retina. Analyzing these features aids in diagnosing of glaucoma in this research work.

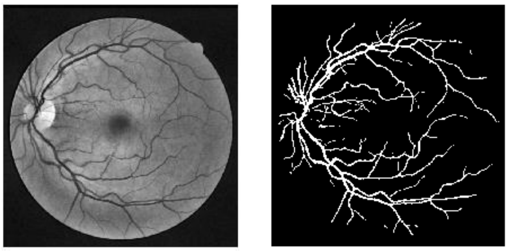


(a) (b)

Figure 1. Segmentation of OCTI (a) input image and (b) segmented image

3.2. Segmentation of FIs

Segmentation of retinal blood vessels in FIs are achieved using active contour techniques with matched filters using Hessian matrix-based features. Utilized a Hessian matrix [22] and matched filters to segment retinal vessels. In order to segment retinal vessels, used matched filters and Hessian matrices eigenvalues. The parameter tuning and fine-tuning of the algorithms are necessary to achieve accurate segmentation results with the influence of distinctive blood vessels of FIs. A matched filter is designed to enhance specific features in an image by convolving the image with a filter kernel that matches the desired feature. To extract vessel features from a FIs using the Hessian matrix, compute the eigenvalues of the Hessian matrix at each pixel location. The eigenvalues represent the curvature of the intensity surface around each pixel, and they used to enhance vessel-like structures. A fundus image captures the interior surface of the eye, including the retina, optic disc, and blood vessels. Retinal vessel segmentation involves extracting the vascular network from the fundus image, crucial for diagnosing conditions of glaucoma. Figures 2(a) depicts a fundus input image through Figure 2(b), the corresponding retinal vessel segmentation highlights the extracted blood vessel network, aiding in quantitative analysis and disease detection.



(a) (b)

Figure 2. Segmentation of FI (a) input image (b) segmented image

3.3. Classification techniques

Classifying medical images is crucial for accurate and timely diagnosis in medical diagnostics, enabling healthcare professionals to analyze medical images automatically. With a view to classify glaucoma and healthy, two unlike DLN techniques are employed in this work, for instance a GN in contemplation of segmented FIs and OCTIs separately (uni-modal) and a multi-modal (MM) CNNs with combination of segmented FIs and OCTIs. The utilization of two different image data within CNN architectures holds great promise for refining the classification of glaucoma, influencing the unique advantages of each modality to improve accuracy and clinical relevance.

3.3.1. GoogLeNet architecture

As part of the TensorFlow open-source machine learning library, Google has developed deep learning frameworks. While TensorFlow itself does not constitute a NN architecture but ML models and NNs, are deployed in flexible platform [23]. A paradigm for capturing features at multiple scales is also invoked by inception networks, a GN system that uses parallel convolutions of different filter sizes. Several advancements in GN architectures have occurred since the introduction of Inception, which have significantly influenced the exploitation of DL techniques for image classification. Using inception modules, it allows for simultaneous processing of multiple scales of features. The network comprises 144 layers in deep, including convolutional, pooling, and fully connected layers. This architecture addresses the phenomenon as gradient problem and promotes efficient feature learning.

3.3.2. Multi-modal CNN

With the intention of improve the system performance, robustness and nuanced understanding of the underlying data, multi-modal CNNs [24] integrate and leverage many different information sources. Images from multiple modalities and heterogeneous datasets make this method particularly useful. This article introduced a MMCNN in classifying glaucoma or healthy. By processing each modality and extracting its features, the architecture is designed to make accurate predictions by combining these features. CNN with multi-input is a neural network architecture as shown in Figure 3, comprising 8 unique layers for every input, with the fusion section incorporating 4 layers. This is designed to handle multiple input sources simultaneously such as segmented FIs and OCTIs. It involves sliding a small filter (as a kernel size 5×5) over the input data to extract features. In the context of MMCNNs, each stream of input source networks has their individual set of convolutional layers (CLs) filters. In pooling layers (PLs) part of the network, down-sampling operation that reduces the spatial dimensions for an image input data. This includes max pooling and average pooling is applied independently to each input source. The parameters of the MMCNN, including the convolutional filter weights and biases, are updated during the training process using an optimization algorithm like gradient descent. Back-propagation is used to compute gradients and update the parameters to minimize a defined loss function. In the forward pass, input data is passed through the network layer by layer. For each input source, the data goes through CLs, PLs, and other types of layers like fully connected layers. The information from output of each layer becomes the input for the next layer. After the forward pass-through individual branches of the network corresponding to each input source, the outputs are aggregated involves in this work is concatenation with fully connected classification layer (FCCL). The final aggregated output is compared to the ground truth (target) using a loss function, that measures the differences obtained from predicted output and the actual target. The loss function includes the cross-entropy for classification tasks.

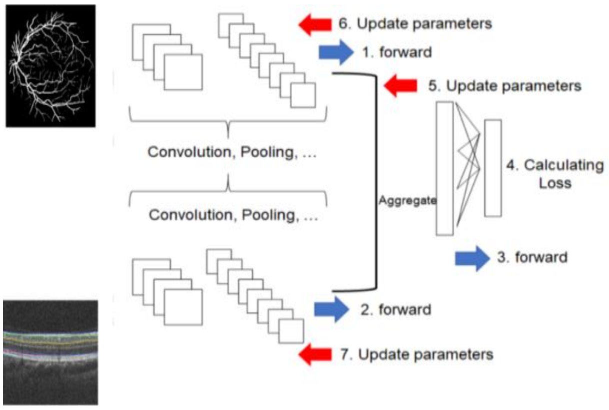


Figure 3. MMCNN layer structure

CNNs are a class of deep neural networks particularly effective in image analysis tasks. They consist of convolutional layers that automatically learn hierarchical representations of features from input images. The ability of CNNs to automatically extract relevant features makes them well-suited for complex tasks like medical image classification. The training process involves feeding labeled FIs and OCTIs into the CNN. The network learns to recognize patterns indicative of various eye conditions through iterations. Data augmentation techniques may be employed to increase the diversity of the training dataset and enhance the model's robustness. Exploration of multi modal approaches, combining information from FIs and OCTIs, may further enhance the accuracy of diagnostic models. The application of CNNs in the classification of FIs and OCTIs are seizing a good assurance for progressive ophthalmic diagnostics to significantly improve patient outcomes in the field of ophthalmology.

4. MEASUREMENT OF SYSTEM METRICS

When evaluating the intended results and performance of DLN models [25], several measures are commonly used. Some of the most fundamental performance measures include accuracy, F1-score, recall, precision, specificity and confusion matrix (CM). The CM summarizes the performance of the system with respect to classification algorithm. It compares the actual with predicted labels in the data sets. The CM relates attributes such as i) true negatives (TN): instances in which the outcome is negative (belonging to the negative class) and are correctly predicted as negative by the model; ii) false positives (FP): instances in which the outcome is negative but are incorrectly predicted as positive by the model; iii) true positives (TP): a true positive occurs when the model correctly predicts a positive instance as positive; and iv) false negatives (FN): in neural networks, a false negative occurs when the model incorrectly predicts a negative instance as positive. Accuracy: this measures the proportion based on the total number of instances, there were instances that were correctly classified as in (1).

$$Accuracy = \frac{Number\ of\ correctly\ classified\ insances}{Total\ number\ of\ instances} \tag{1}$$

Precision: as in (2) measures the proportion based on a total of all positive instances, the percentage of true positive predictions (correctly predicted positives) is calculated.

$$Precision = \frac{True\ positives}{True\ positives + False\ positives} \tag{2}$$

Recall (sensitivity): This measures the proportion of predictions of true positives based on all actual positives as in (3).

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ negatives} \tag{3}$$

F1-score: the F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall as in (4).

$$F1\ score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{4}$$

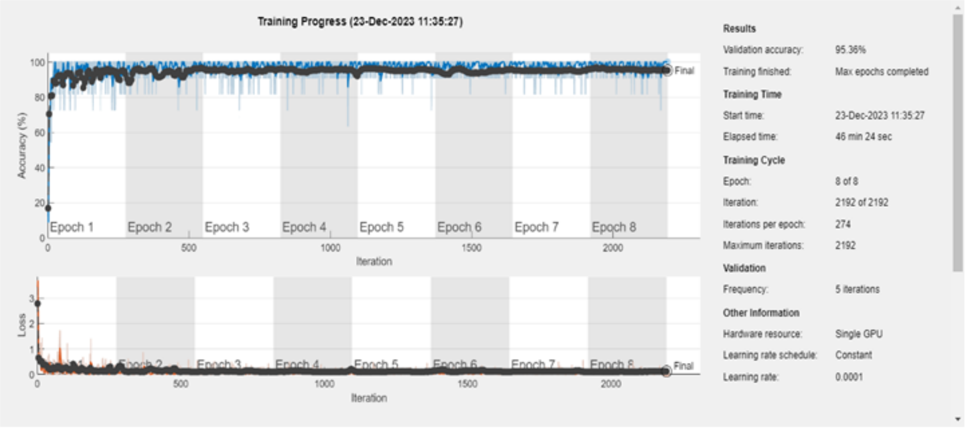
Specificity: also known as the true negative rate is a measure used in binary classification tasks. It quantifies the proportion of true negatives (correctly predicted negatives) out of all actual negatives in the data set. It is particularly useful when the cost of false positives is high as in (5).

$$Specificity = \frac{True\ negatives}{True\ negatives + False\ positives} \tag{5}$$

5. RESULTS AND DISCUSSION

5.1. Training process

The GN [26] architecture is created to address an issue in the training process by utilizing DLNs, which allows using filters of multiple sizes in parallel, capturing both fine and coarse features. Training a neural network like GN for tasks such as FIs and OCTIs classification involves several steps: feed segmented dataset of FIs and OCTIs to the network with labeled to the corresponding classes as normal and glaucoma. Fragmented the dataset into training, validation and test sets. Resize images to a consistent input size 224×224 suitable for the network to normalize pixel values to bring into a similar range. Augment the dataset with techniques like rotation to increase the diversity in the image texture and improved in significant feature generalization. The model architecture GN consists of multiple inception modules for classification of glaucoma or normal. Modify the output layer to match the number of classes in dataset. Define an appropriate loss function for the task, such as cross-entropy loss for classification problems. The optimizer as stochastic gradient descent (SGD) [27] to minimize the loss during training. Calculated the loss and back propagate the gradients to update the weights of the NNs using optimizer. Monitored the validation image sets to ensure the model generalizes well to unseen data and does not overfit. The training options such as initial learn rate 0.0001, mini batch size 6, max epoch 8, validation frequency is 5. Training NNs often benefits from particularized graphics processing unit (GPU) hardware for highly parallel processors capable of handling complex computational tasks with less time. The Figure 4 shows accuracy and loss of training and validation process for FIs input with 90% dataset is used to training gives validation accuracy 95.36%.



Results
Validation accuracy: 95.36%
Training finished: Max epochs completed
Training Time
Start time: 23-Dec-2023 11:35:27
Elapsed time: 46 min 24 sec
Training Cycle
Epoch: 8 of 8
Iteration: 2192 of 2192
Iterations per epoch: 274
Maximum iterations: 2192
Validation
Frequency: 5 iterations
Other Information
Hardware resource: Single GPU
Learning rate schedule: Constant
Learning rate: 0.0001

Figure 4. Accuracy and loss of training and validation process of GNN

5.2. Multimodal fusion of fundus and OCT training process

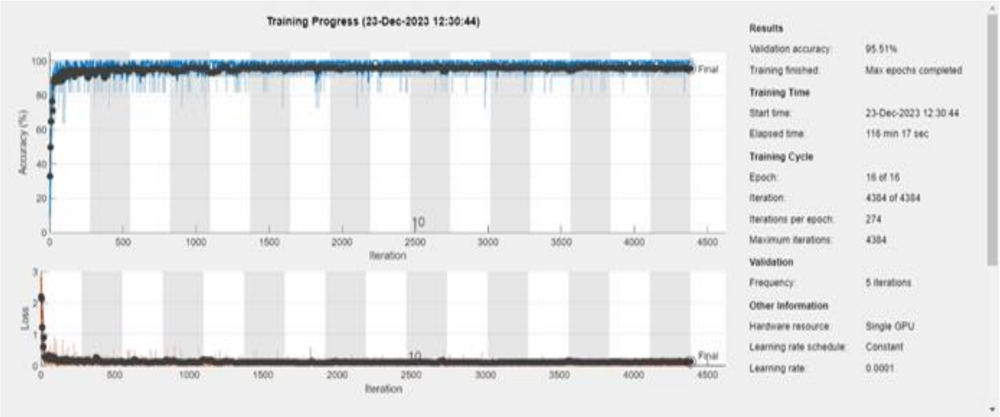
Training multi-modal CNNs [28] with segmented FIs and OCTIs involves by fusion of image texture features from both modalities to improve the model's performance on specific tasks. The training process includes paired Fundus and OCT images for the input to the network. Each pair to be associated with a specific label normal and glaucoma for classification. Preprocess the segmented Fundus and OCT images separately to ensure they have consistent dimensions and pixel values. Normalize pixel values and applied augmentation techniques to increase the heterogeneity features in the image data using rotation.

The architecture might include separate branches for each modality, with shared layers or connections at deeper layers to facilitate information fusion. Define an appropriate loss function for the chosen task. For classification tasks, cross-entropy loss [29] is commonly used. Select an optimizer as SGD to minimize the loss during training. Incorporated mechanisms for feature fusion to combine information from both modalities. This process involves as concatenation fusion strategy. The training options such as initial learn rate 0.0001, mini batch size 6, max epoch 8, validation frequency is 5. Figure 5 shows accuracy and loss of training and validation process with input with 90% dataset is used to training gives validation accuracy 95.51%.

5.3. GoogleNet testing process

The testing process involves evaluating its performance on a separate dataset that it has not seen during training. Here is a step-by-step explanation of the testing process for FIs and OCTIs classification using GoogleNet. Prepared the segmented test dataset by re-sizing images to the same dimensions used during training and normalizing pixel values. Load the pre-trained GoogleNet framework that was trained on the OCTIs and FIs dataset. Pass the segmented test images through the trained GoogleNet model. Obtain the model's predictions for each image. The uni-modal for FIs (500 images) results more efficient than the OCTIs (500 images). Table 1 demonstrates the split of image data set for experimental verification of uni-modal and multi-modal system. There is a 10%, 20%, and 30% percentage image data set allocation is dedicated to testing and a remaining allocation is considered for training.

The model performance for FIs without segmentation is shown in Table 2 for different percentage of test data. Using 10% allocation data for testing yields better results than using 20% or 30% of the test data, and Table 3 illustrates the performance of the system without segmentation. According to Table 4, CM can be applied to FIs with segmentation. In Table 5, we show system performance with segmented input images. This is because the images themselves contain significant texture information, which results in better results.



Results

Validation accuracy:	95.51%
Training finished:	Max epochs completed

Training Time

Start time:	23-Dec-2023 12:30:44
Elapsed time:	116 min 17 sec

Training Cycle

Epoch:	16 of 16
Iteration:	4384 of 4384
Iterations per epoch:	274
Maximum iterations:	4384

Validation

Frequency:	5 iterations

Other Information

Hardware resource:	Single GPU
Learning rate schedule:	Constant
Learning rate:	0.0001

Figure 5. Accuracy and loss of training and validation process of MMCNN

Table 1. Split of image data set

Total number of samples =1000 (500 FIs + 500 OCTIs)		
Image data	Train samples	Test samples
10%	900	100
20%	800	200
30%	700	300

Table 2. Classification results without segmentation of FIs

Image data	TP	FP	TN	FN
10%	41	9	42	8
20%	78	22	79	21
30%	115	35	118	32

Table 3. Evaluation metrics without segmentation of FIs

Image data	PR	SPR	SER	F1-score	AR	AUC
10%	82%	82%	84%	83%	83%	0.84
20%	78%	78%	79%	78%	79%	0.78
30%	77%	77%	78%	77%	78%	0.77

Table 4. Classification results with segmentation of FIs

Image data	TP	FP	TN	FN
10%	45	5	46	4
20%	87	13	83	17
30%	125	25	128	22

Table 5. Evaluation metrics with segmentation of FIs

Image data	PR	SPR	SER	F1-score	AR	AUC
10%	90%	90%	92%	91%	91%	0.91
20%	87%	86%	84%	85%	85%	0.85
30%	83%	83%	85%	84%	84%	0.83

5.4. Multimodal fusion of fundus and OCT testing process

Testing trained MMCNNs with segmented input of FIs and OCTIs involves evaluating its performance on a separate dataset that it has not seen during training. Here is a step-by-step explanation of the testing process. Segmented test dataset containing FIs and OCTIs pairs, normalize pixel values and ensure that the dimensions data match the input requirements of the trained multi-modal CNN. Load the pre-trained multi-modal CNN that accomplished high-grade validation accuracy during training. Pass the pairs of

FIs and OCTIs through the loaded multi-modal CNN model. Obtain the prediction classification labels of model for the specific datasets and compare the predicted labels with the target labels for each image pairs. As shown in Table 6, CM for multi-modal is demonstrated without segmentation and in Table 7, illustrated the system performance without segmentation. However, MMCNNs gives better results than uni-modal.

According to Table 8, classification results for MMCNNs are demonstrated for segmented inputs. In case, TP=49, meaning the model correctly identified 49 instances as positive out of the total 50. Here, TN=50, indicating that the model correctly identified 50 instances as negative out of the total 50. In scenario, FP=1, indicating that the model mistakenly classified 1 instance as positive when it was actually negative. In FN=0, indicating that the model did not mistakenly classify any positive instances as negative for allocation of 10% image data for testing. Nevertheless, the model outperforms the classification results achieved less in case of 20% and 30%.

Table 9 illustrates MMCNNs evaluation metrics with segmentation. However, over all MMCNNs with segmentation inputs gives better results than uni-modal for allocation of 10% image data for testing. With an AR of 99%, it means that 99% of the instances were correctly classified by the model. SER, also known as recall rate, with a sensitivity of 98%, it indicates that the model correctly identified 98% of the positive instances. A SPR of 98% means that the model correctly identified 98% of the negative instances. With PR of 98% indicates that 98% of the instances predicted as positive by the model were indeed true positives. An F1-score of 98% suggests a good balance between precision and recall, with high values indicating both high precision and high recall. Receiver operating characteristic (ROC) curve, which plots the TP rate against the FP rate. A higher AUC value, closer to 1, indicates better discrimination between positive and negative instances. An AUC of 0.99 suggests excellent performance in distinguishing between classes.

Classifier performance is visualized using ROC [30] curves, aiding in model comparing performance and understanding true versus false positive trade-offs. Figure 6 shows the ROC curve for 10% image test data with segmented inputs for class-1 (glaucoma) and class-2 (normal). A collection of ROC curves for each class provides insights into the classifier performance for each class during testing. An AUC value closer to 1 indicates better classification performance, while an AUC of 0.5 suggests random guessing. ROC 95% training accuracy indicates that the model achieved high discriminatory power in distinguishing between class-1 and class-2 instances during the training phase. This suggests that the model predictions were mostly correct when evaluated on the training data with reference to Figure 6(a). The ROC 95.51% for validation accuracy implies that the model performance remained consistently high when assessed on a separate validation dataset. It suggests that the model generalization ability is robust, as it can effectively discriminate between class-1 and class-2 instances on unseen dataset considering in Figure 6(b). ROC of 99% for testing accuracy indicates that the model maintained its strong discriminatory power when evaluated on an independent testing dataset as in the context of Figure 6(c). This suggests that the model performance is reliable and consistent across different datasets. For all ROC indicates that the model discriminatory power is consistently high across, this implies that the model predictions are consistently reliable, irrespective of the image dataset chosen for classification as observing Figure 6(d).

A graphical representation of CM in Figure 7(a) without segmentation and through Figure 7(b) with segmented input to MMCNNs summarizes the classification algorithm performance. A comparison is made between the True class of the dataset and the predicted class by the model. True classes are represented in rows, while predicted classes are represented in columns, TP=49 and FP=1 out of 50 total glaucoma instances, and TP=50 and TN=50 out of 50 total normal instances, with FN=0, when allocating 10% of image data for testing gives better results with respect to 20% and 30% allocation of image data for testing. Overall, Table 10 provides a comprehensive comparison of different methods for analyzing Fundus and OCT images to classify glaucoma or healthy, highlighting their respective testing accuracy rate. The proposed method appears to outperform the other methods in terms of accuracy.

Table 6. Classification results segmentation of FIs and OCTIs

Image data	TP	FP	TN	FN
10%	47	03	48	02
20%	91	09	90	10
30%	130	20	131	19

Table 7. Evaluation metrics without segmentation of FIs and OCTIs

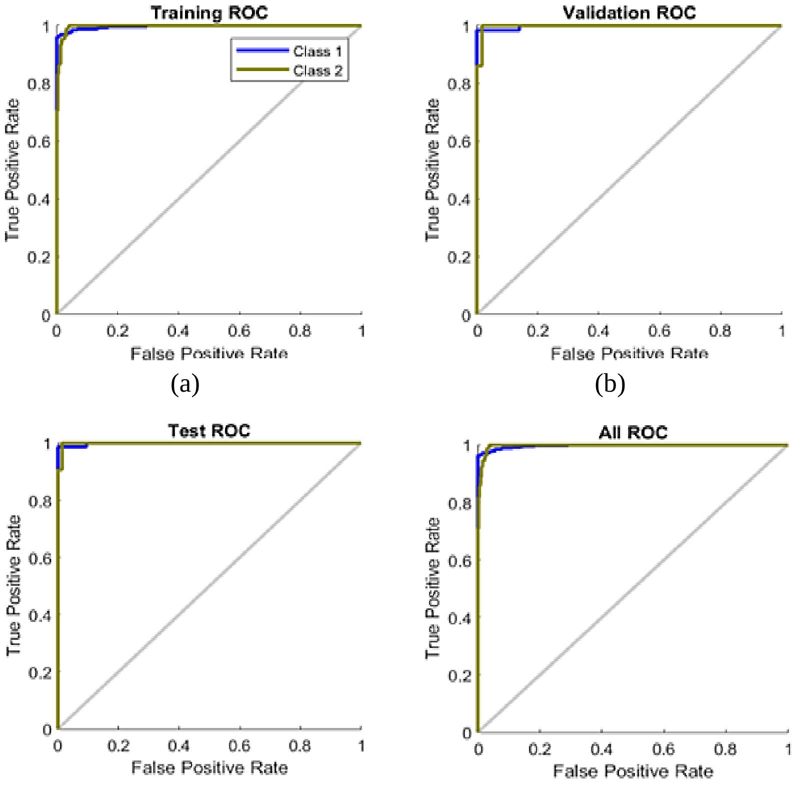
Image data	PR	SPR	SER	F1 Score	AR	AUC
10%	94%	94%	96%	95%	95%	0.95
20%	91%	91%	90%	90%	91%	0.90
30%	86%	87%	87%	86%	87%	0.86

Table 8. Classification results segmentation of FIs and OCTIs

Image data	TP	FP	TN	FN
10%	49	1	50	00
20%	97	03	96	04
30%	135	15	138	12

Table 9. Evaluation metrics with segmentation of FIs and OCTIs

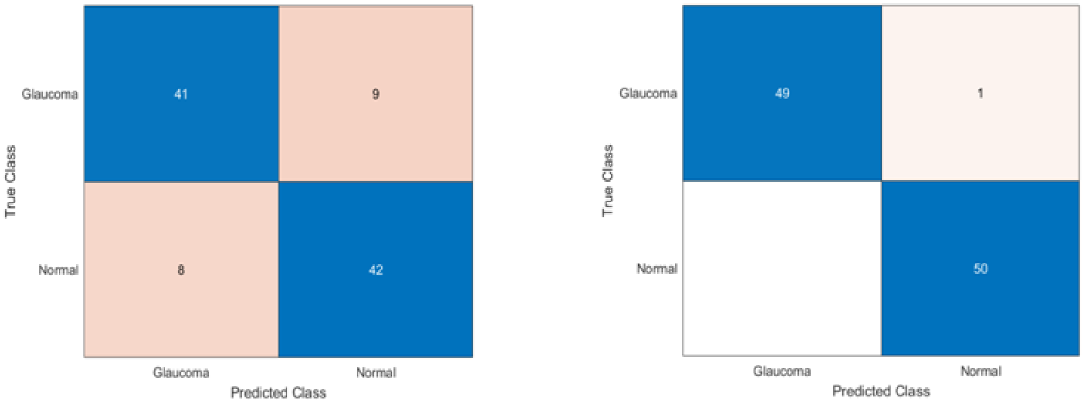
Image data	PR	SPR	SER	F1-score	AR	AUC
10%	98%	98%	98%	98%	99%	0.99
20%	97%	97%	97%	97%	97%	0.97
30%	90%	90%	90%	90%	91%	0.90



(a) (b)

(c) (d)

Figure 6. Input segmentation ROC for MMCNNs: (a) training ROC, (b) validation ROC, (c) test ROC, and (d) all ROC



(a) (b)

Figure 7. Graphical representation of CM for MMCNNs (a) without and (b) with segmented inputs

Table 10. Comparison of system accuracy across various methods

Ref.no.	Input image	Methods	Testing accuracy (%)
[1]	Fundus	DenseNet-201	96.90
[3]	Fundus	ANFIS	97.20
[4]	Fundus	CNN	97.57
[5]	OCT	ResNet	98.67
[15]	Fundus	VGG-16	98.6
Proposed work	Fusion of fundus and OCTs	MMCNNs	99

6. CONCLUSION

In this study, trained and evaluated the GN architecture comprising uni-modal for OCTIs and FIs for glaucoma disease classification. The GN demonstrated a commendable overall accuracy in predicting of the test set correctly. The model exhibited a robust sensitivity to relevant features for segmented FIs. A comprehensive work is rigorously tested the performance of MMCNNs on a diverse dataset containing paired segmented FIs and OCTIs. The incorporation of feature fusion mechanisms facilitated the utilization of information from both modalities. The MMCNNs highlighted specific strengths i.e., effective information fusion, robust performance in diverse scenarios, making it a promising algorithm for glaucoma detection and classification. The MMCNNs achieves better system performance, when dealing with segmented inputs of FIs and OCTIs, as well as when compared to uni-modal approaches. The fusion of information from distinct modalities is indeed advantage, but the effectiveness of the model may vary depending on the quality and variability of the input data.

Future research endeavors could focus on improving the proposed model, e.g., fine-tuning for specific pathologies, integration of additional modalities. The testing of MMCNNs on FIs and OCTIs unveils promising prospects for advancing diagnostic capabilities in specific medical contexts. The model adept fusion of information from distinct modalities paves the way for enhanced clinical decision support systems. Ongoing research and refinement are crucial to exploit the full potential of this multi-modal approach.

REFERENCES

[1] R. Kashyap, R. Nair, S. M. P. Gangadharan, M. Botto-Tobar, S. Farooq, and A. Rizwan, “Glaucoma detection and classification using improved U-Net deep learning model,” *Healthcare*, vol. 10, no. 12, Dec. 2022, doi: 10.3390/healthcare10122497.

[2] S. Ajitha, J. Akkara, and M. Judy, “Identification of glaucoma from fundus images using deep learning techniques,” *Indian Journal of Ophthalmology*, vol. 69, no. 10, 2021, doi: 10.4103/ijo.IJO_92_21.

[3] G. Latha and P. Aruna Priya, “Glaucoma retinal image detection and classification using machine learning algorithms,” *Journal of Physics: Conference Series*, vol. 2335, no. 1, Sep. 2022, doi: 10.1088/1742-6596/2335/1/012025.

[4] V. K. Velpula and L. D. Sharma, “Multi-stage glaucoma classification using pre-trained convolutional neural networks and voting-based classifier fusion,” *Frontiers in Physiology*, vol. 14, Jun. 2023, doi: 10.3389/fphys.2023.1175881.

[5] S. Joshi, B. Partibane, W. A. Hatamleh, H. Tarazi, C. S. Yadav, and D. Krah, “Glaucoma detection using image processing and supervised learning for classification,” *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, Mar. 2022, doi: 10.1155/2022/2988262.

[6] N. Akter, J. Fletcher, S. Perry, M. P. Simunovic, N. Briggs, and M. Roy, “Glaucoma diagnosis using multi-feature analysis and a deep learning technique,” *Scientific Reports*, vol. 12, no. 1, May 2022, doi: 10.1038/s41598-022-12147-y.

[7] A. Cerentinia, D. Welfera, M. C. D’Ornellasa, C. J. P. Haygertb, and G. N. Dottob, “Automatic identification of glaucoma using deep learning methods,” *Proceedings of the 16th World Congress on Medical and Health Informatics-Precision Healthcare through Informatics, (MEDINFO)*, Hangzhou, China, 2017, pp. 318-321, doi: 10.3233/978-1-61499-830-3-318.

[8] J. J. Gómez-Valverde *et al.*, “Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning,” *Biomedical Optics Express*, vol. 10, no. 2, Feb. 2019, doi: 10.1364/BOE.10.000892.

[9] K.-H. Hung *et al.*, “Application of a deep learning system in glaucoma screening and further classification with colour fundus photographs: a case control study,” *BMC Ophthalmology*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12886-022-02730-2.

[10] Aziz-ur-Rehman, I. A. Taj, M. Sajid, and K. S. Karimov, “An ensemble framework based on deep CNNs architecture for glaucoma classification using fundus photography,” *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 5321–5346, 2021, doi: 10.3934/mbe.2021270.

[11] M. Pandey and D. S. Solanki, “A review of glaucoma detection using machine learning,” *International Journal of Scientific Research and Engineering Trends*, vol. 7, no. 6, pp. 3507–3510, 2021.

[12] M. K. Shukla, “Classification of different stages of glaucoma using deep learning approaches,” MSc Research Project, School of Computing, National College of Ireland, 2020.

[13] J. Surendiran and M. Meena, “Deep learning-assisted glaucoma diagnosis and model design,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1, pp. 269–276, 2023.

[14] 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, Jul. 2022, doi: 10.1038/s41598-022-16262-8.

[15] S. Bhattacharya and Y. K. Rathore, “Review on restoration of vision in glaucoma through image processing using machine learning techniques,” *International journal of health sciences*, pp. 8218–8225, May 2022, doi: 10.53730/ijhs.v6nS2.7085.

[16] A. Geetha and N. B. Prakash, “Classification of glaucoma in retinal images using efficientnetB4 deep learning model,” *Computer Systems Science and Engineering*, vol. 43, no. 3, pp. 1041–1055, 2022, doi: 10.32604/csse.2022.023680.

[17] S. Karkuzhali, A. Mishra, M. S. Ajay, and S. Wilson Prakash, “Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models,” in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2020, pp. 1–9, doi: 10.1109/ICCCI48352.2020.9104157.

[18] S. Saha, J. Vignarajan, and S. Frost, "A fast and fully automated system for glaucoma detection using color fundus photographs," *Scientific Reports*, vol. 13, no. 1, Oct. 2023, doi: 10.1038/s41598-023-44473-0.

[19] G. V. Datta, S. R. Kishan, A. Kartik, G. B. Sai, and S. Gowtham, "Glaucoma disease detection using deep learning," in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Feb. 2023, pp. 1–6, doi: 10.1109/ICECCT56650.2023.10179802.

[20] Y. Tian, M. Zang, A. Sharma, S. Z. Gu, A. Leshno, and K. A. Thakoor, "Glaucoma progression detection and Humphrey visual field prediction using discriminative and generative vision transformers," in *International Workshop on Ophthalmic Medical Image Analysis*, 2023, pp. 62–71.

[21] B. I. Dodo, Y. Li, D. Kaba, and X. Liu, "Retinal layer segmentation in optical coherence tomography images," *IEEE Access*, vol. 7, pp. 152388–152398, 2019, doi: 10.1109/ACCESS.2019.2947761.

[22] J. Pumplin *et al.*, "Uncertainties of predictions from parton distribution functions. II. The Hessian method," *Physical Review D*, vol. 65, no. 1, Dec. 2001, doi: 10.1103/PhysRevD.65.014013.

[23] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018, doi: 10.1016/j.procs.2018.05.198.

[24] R. Rajasree, C. C. Columbus, and C. Shilaja, "Multiscale-based multimodal image classification of brain tumor using deep learning method," *Neural Computing and Applications*, vol. 33, no. 11, pp. 5543–5553, Jun. 2021, doi: 10.1007/s00521-020-05332-5.

[25] A. M. Carrington *et al.*, "Deep ROC analysis and AUC as balanced average accuracy to improve model selection, understanding and interpretation," *arXiv preprint arXiv:2103.11357*, 2021.

[26] S. Barman, M. R. Biswas, S. Marjan, N. Nahar, M. S. Hossain, and K. Andersson, "Transfer learning based skin cancer classification using GoogLeNet," in *International Conference on Machine Intelligence and Emerging Technologies*, 2023, pp. 238–252.

[27] A. Q. Albayati, S. A. J. Altaie, W. N. I. Al-Obaydy, and F. F. Alkhalid, "Performance analysis of optimization algorithms for convolutional neural network-based handwritten digit recognition," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 1, pp. 563–571, Mar. 2024, doi: 10.11591/ijai.v13.i1.pp563-571.

[28] S. Agarwal, K. V. Arya, and Y. K. Meena, "MultiFusionNet: multilayer multimodal Fusion of deep neural networks for Chest X-Ray image classification," *arXiv preprint arXiv:2401.00728*.

[29] E. Matsuyama, M. Nishiki, N. Takahashi, and H. Watanabe, "Using cross entropy as a performance metric for quantifying uncertainty in DNN image classifiers: an application to classification of lung cancer on CT images," *Journal of Biomedical Science and Engineering*, vol. 17, no. 01, pp. 1–12, 2024, doi: 10.4236/jbise.2024.171001.

[30] P. Krakowski, R. Karpiński, R. Maciejewski, and J. Jonak, "Evaluation of the diagnostic accuracy of MRI in detection of knee cartilage lesions using receiver operating characteristic curves," *Journal of Physics: Conference Series*, vol. 1736, no. 1, Jan. 2021, doi: 10.1088/1742-6596/1736/1/012028.

BIOGRAPHIES OF AUTHORS

Nanditha Krishna received her B. E degree in instrumentation technology from Visvesvaraya Technological University, Belagavi in 2009 and MTech degree in medical electronics from Visvesvaraya Technological University, Belagavi in 2011. She is currently pursuing her Ph.D. from R.V. College of Engineering under Visvesvaraya Technological University, Belagavi. Her area of interest includes image processing, biomedical signal processing, biomedical instrumentation, and digital design. She is presently working in Dayananda Sagar College of Engineering, Bengaluru. She has 13 years of teaching experience. She is a life member of ISTE professional society. She can be contacted at email: nanditha13@gmail.com.

Nagamani Kenchappa holds a doctorate degree in image processing from Visvesvaraya Technological University, Belagavi. She has received her M.Tech. in digital communication engineering from Visvesvaraya Technological University, Belagavi in 2006 and B.E in electronics and communication from BMSCE. She works as a professor in the Department of Electronics and Telecommunication, R.V. College of Engineering, Bengaluru and has 19 years of teaching experience. Her area of interest includes image compression, wireless communication, IoT. She is IEEE senior member, fellow IETE, ISTE life member, Execom member of IEEE Signal Processing Society, Execom member and Secretary of IEEE sensors Council Bangalore Chapter. She has published and reviewed many research papers in various journals and conferences. She has chaired many IEEE conference sessions and is also a BOE member in reputed autonomous universities, Bengaluru. She can be contacted at email: nagamanik@rvce.edu.in.