STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **12**, May 2024, pp 630–645.
Published online in International Academic Press (www.IAPress.org)

Multi-Dataset Convolutional Neural Network Model for Glaucoma Prediction in OCT Fundus Scans

Badr ELKARI [1], Loubna OURABAH [1], Hiba SEKKAT [1], Mohamed Mouad OUHASNI [1], Achraf RACHID [1], Chaymae KHANI [1], Karim EL MOUTAOUAKIL [2],*

[1] *EIDIA, Euromed Research Center, Euro-Med University(UEMF),Fez, Morocco*
[2] *Engineering Science Laboratory, Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University of Fez, Morocco*

Abstract Glaucoma, a major factor in permanent blindness across the globe, necessitates accurate and efficient diagnostic methods. This article presents a comprehensive approach to glaucoma prediction by combining three diverse datasets: ORIGA, ACRIMA, and REFUGE. A novel multi-Datasets CNN (MD-CNN) architecture is proposed, specifically tailored to effectively handle combined data comprising diverse image characteristics across multiple datasets. This innovative approach demonstrates remarkable robustness in accommodating variations in image attributes, including lighting, zoom levels, and other disparate features, thus showcasing its potency in addressing glaucoma prediction across different datasets. The approach demonstrates improved accuracy (96.88%), sensitivity (94.34%), specificity (97.20%), precision (94.34%), and AUC (99.02%) compared to individual dataset-based models, addressing challenges in glaucoma detection. This research showcases the potential of combining diverse datasets for more effective CNN-based glaucoma detection.

Keywords MD-CNN, glaucoma, dataset, fundus images, convolutional neural network

DOI: 10.19139/soic-2310-5070-1935

1. Introduction

The human eyes are the most frequently utilized sensory organs among the five senses. Visual processing requires a significant portion of the brain's capacity. Glaucoma, often resulting from elevated pressure within the eye [1]. Glaucoma is a chronic eye disease that, if not effectively managed, can lead to permanent vision loss [2]. According to a study conducted in 2013, approximately 64.3 million individuals between the ages of 40 and 80 were affected by glaucoma worldwide. However, projections suggest that this number could reach 111.8 million by 2040. These estimates highlight the potential growth and impact of glaucoma on global populations [3]. In Morocco, glaucoma is a significant contributor to blindness and ranks as the second leading cause, responsible for approximately 14.3% of blindness cases[4] .
Furthermore, chronic diseases are health conditions that persist over an extended period and typically necessitate continuous medical attention, which can substantially affect an individual's quality of life. Additionally, these diseases can impose a considerable financial burden on both individuals and healthcare systems [5] In the case of glaucoma, the annual direct medical expenses per patient in the United States vary depending on the severity of the disease, ranging from 623 USD to 2511 USD [6].
Moreover, the optic nerves are made up of the one million or so nerve fibers that make up the retina layer. Optic disk (OD) or optic nerve head (ONH) refers to the starting location of the optic nerves within the retina layer. This region appears circular and stands out prominently in retinal images. Glaucoma is known as "the quiet thief of

*Correspondence to: ELKARI Badr (Email: b.elkari@ueuromed.org). EIDIA, Euromed Research Center, Euro-Med University(UEMF),Fez, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

sight" since its symptoms frequently don't appear until the condition is already slightly advanced. Therefore, it is essential to identify glaucoma early and promptly [7].
With the prevalence of glaucoma on the rise and its potential to cause permanent vision loss, early and accurate diagnosis becomes paramount. Traditional methods of diagnosing glaucoma rely heavily on manual examination and interpretation of retinal images by trained professionals. However, these approaches are time-consuming, subjective, and can be prone to human error. Recently, there has been a boost in interest in utilizing machine learning and artificially intelligent technology (AI) to help diagnose and predict glaucoma.
Additionally, in numerous medical imaging applications, machine learning methods, particularly Convolutional Neural Networks (CNNs), have shown considerable potential, including the analysis of retinal images. By utilizing large datasets of OCT (Optical Coherence Tomography) fundus scans, CNNs can learn to extract intricate features and patterns that might not be readily discernible to the human eye. This enables them to identify subtle abnormalities and potential indicators of glaucoma with remarkable accuracy [8], [9].
The application of machine learning AI in glaucoma diagnosis holds immense potential for revolutionizing the field of ophthalmology. By automating and streamlining the diagnostic process, these technologies can assist healthcare professionals in making faster, more reliable assessments, leading to earlier interventions and improved patient outcomes [10]. Furthermore, the adoption of machine learning AI has the potential to alleviate the financial burden associated with glaucoma, making healthcare resources more efficient and accessible [11].
Consequently, we will explore the use of CNNs for predicting glaucoma in OCT fundus scans. By harnessing the power of machine learning and AI, we aim to develop a robust and efficient tool that can aid in the early detection and management of glaucoma, ultimately preserving and enhancing the vision of individuals affected by this debilitating disease [12].

2. Related work

Expanding on the potential of machine learning [13], artificial intelligence (AI), and deep learning in prediction[14], [15], significant research efforts have been devoted to utilizing these techniques for analyzing OCT (Optical Coherence Tomography) scans. OCT provides high-resolution images of the retinal layers, enabling the extraction of valuable features that can aid in the early detection and management of glaucoma. In the following section, we delve into a collection of influential articles that showcase advancements in the use of machine learning, AI, and deep learning for analyzing OCT scans and predicting glaucoma. These studies explore diverse methodologies, datasets, and evaluation approaches, shedding light on the progress made in this field and offering valuable insights into the potential of these techniques for enhancing glaucoma diagnosis.
In their study, Daneshvar et al. [16] conducted an analysis on the predictive capabilities of baseline OCT (Optical Coherence Tomography) measures for visual field (VF) progression in patients with suspected or diagnosed glaucoma. They compared the performance of these measures with semiquantitative optic disc measures. The findings indicated that baseline measurements of pRNFL (peripapillary retinal nerve fiber layer) and macular OCT parameters can be utilized to assess the risk of future glaucoma progression. Individuals with abnormal OCT findings require enhanced care to prevent the progression of functional damage.
In their study [17], The researchers tested 1542 fundus pictures from 786 healthy individuals and 756 patients with glaucoma. For logistic regression, these photos were converted into a 1 (240 240 3) one-dimensional array. The classification work was carried out with the help of the Google Net Inception v3 model and a customized classification layer. They employed the Adam optimizer for backpropagation and cross-entropy as the loss function. The performance evaluation was based on the curve of ROC, specifically measuring sensitivity and specificity. The final version of the model achieved an 87.9% as accuracy of and an AUC of 0.94 on the test data, 92.2% accuracy and 0.98 AUC on the training data, and 88.6% accuracy and 0.95 AUC on the validation data. These results demonstrate the effectiveness of the developed model in accurately classifying fundus images for glaucoma detection.
In their study, A.Guangzhou et al. [18] proposed a model for glaucoma detection using open-angle glaucoma based on 3D color image data. The model employed a CNN architecture with various fundus images as input. The output

from each CNN model was combined, and the random forest method was utilized for classifying the fundus images into healthy and glaucoma-infected eyes. The evaluation of the model resulted in an AUC (Area Under the Curve) value of 0.96, indicating a high level of performance in distinguishing between healthy and glaucoma-affected eyes.

In their work, Sengar et al. [19] highlighted the use of image processing techniques for glaucoma detection, specifically utilizing fundus images. They introduced a decisive parameter for the detection of glaucoma. The proposed mechanism demonstrated strong performance, achieving an accuracy of 93.57%. They developed a multi-level deep convolutional neural network (ML-DCNN) for glaucoma detection and classification in [20], using advanced machine deep-learning technique on retinal fundus images to diagnose glaucoma affected and normal images. The CNN framework is used on 1338 images to extract features through a multilayer from raw pixel images, and statistical measures such as Specificity (SP), Sensitivity (SE), Accuracy (ACC), and Precision (PRC) are used to achieve an average SE of 97.04%, SP of 98.99%, ACC of 99.39%, and PRC of 98.2%.

In [21], The authors concentrated on the segmentation and the localization of the optic disc head for glaucoma diagnosis utilizing 3D datasets, pixel-level analysis of glaucomatous alterations, and an artificial neural network (ANN) for identifying glaucoma progression.

In [22], a multilayer convolutional neural network (CNN) was employed for glaucoma detection. The network consisted of four convolutional layers followed by two fully connected layers, effectively utilizing deep learning techniques to identify glaucoma patterns and make accurate predictions. In their study, Ahn et al. [17] proposed a method for detecting glaucoma using fundus photography and deep learning techniques. The authors demonstrated that machine learning, combined with fundus images, could accurately identify both advanced and early stages of glaucoma. They utilized a dataset of 1,542 images, which was divided into training, validation, and test datasets. The newly developed model, trained using a convolutional neural network (CNN), proved to be more effective and accurate in detecting early-stage glaucoma.

In reference [23], the proposed design follows a similar structure to the original U-Net architecture, consisting of a contracting path on the left and an expansive path on the right. The contracting path is responsible for repeating the typical structure found in the convolutional part of a classification network. In the expansive path, information is gathered from corresponding layers of the contracting path with higher resolution and lower resolution layers of the expansive path.The study evaluates the results using publicly available datasets, namely Drions-db, Rim-One, and Drishti. The analysis results and visual comparisons demonstrate that automatic optic disc segmentation can be achieved with a level of quality comparable to that of human performance. This indicates the effectiveness and potential of the proposed method in accurately segmenting optic discs in fundus images.

In their study, Serener et al. [24] focused on Open-angle glaucoma, which is a common and progressive disease leading to gradual vision loss. Manual diagnosis of glaucoma by experts can be time-consuming and costly. To address this, the authors proposed a method for automated detection of both early and advanced stages of glaucoma. They trained and fine-tuned deep convolutional neural network (CNN) algorithms, specifically 'ResNet-50' and 'GoogLeNet', using transfer learning. The results indicated that the 'GoogLeNet' model outperformed the 'ResNet-50' model in detecting both early and advanced stages of glaucoma in patients' eyes.This research suggests that leveraging deep CNN algorithms, such as 'GoogLeNet', can improve the automated detection of glaucoma, providing a more efficient and effective approach for diagnosing the disease.

Having explored the advancements in glaucoma prediction using machine learning, AI, and deep learning in various studies, we now turn our attention to the methodology employed in this research. Building upon the insights gained from these influential articles, we propose a comprehensive approach for glaucoma prediction using a multi-dataset Convolutional Neural Network (CNN). Our methodology encompasses several key steps, including data analysis and collection, data preprocessing, the development of a specialized CNN architecture, model compilation and training, performance evaluation, fine-tuning and optimization, interpretability and visualization, and model deployment.

During the data analysis and collection phase, we gather diverse datasets to capture the wide range of glaucomatous pathology, incorporating information from ORIGA, ACRIMA, and REFUGE. The data is then subjected to rigorous preprocessing techniques, ensuring numerical stability and consistent scaling across the images.

Next, we develop a specialized CNN architecture that leverages the combined information from the diverse datasets

to extract relevant features from the OCT scans. The model is compiled and trained using the prepared dataset, and its performance is evaluated using appropriate metrics. Fine-tuning and optimization techniques are applied to enhance the model's predictive capabilities.

Convolutional Neural Networks (CNNs) demonstrate a remarkable capability in detecting subtle glaucoma indicators within OCT (Optical Coherence Tomography) fundus scans. These indicators encompass various alterations, including changes in the optic nerve head (ONH), such as shifts in cup-to-disc ratio or irregularities in contour, serving as markers for glaucomatous damage. Furthermore, CNNs can discern minute irregularities or thinning in the retinal nerve fiber layer and identify alterations within the ganglion cell complex (GCC), consisting of the ganglion cell layer, inner plexiform layer, and inner nuclear layer, often indicating early glaucoma signs. Additionally, these networks excel in recognizing slight variations in macular thickness, especially in regions linked to the fovea, and detecting localized structural changes in the optic nerve and its surrounding areas. Leveraging their ability to grasp intricate image patterns, CNNs play a crucial role in accurately identifying these subtle anomalies. This capability significantly contributes to early detection, ongoing disease monitoring, and the potential prediction of glaucoma onset before substantial visual impairment occurs.

Delayed or inaccurate diagnosis of glaucoma poses significant risks and far-reaching consequences for individuals' vision and well-being. Glaucoma, often insidious in its progression, can cause irreversible damage to the optic nerve, leading to permanent vision impairment if left untreated. The delayed identification of this condition can result in a critical loss of peripheral and central vision, impacting daily activities and significantly reducing one's independence and quality of life. Moreover, delayed diagnosis may necessitate more extensive treatments, surgeries, and heightened healthcare costs to manage the advanced stages of the disease.

Our study significantly advances current diagnostic methods for glaucoma by introducing a pioneering approach utilizing a multi dataset CNN architecture. This innovative model is specifically engineered to adeptly handle diverse datasets characterized by variations in lighting, zoom levels, and other nuanced features. By amalgamating multiple datasets, our method synergizes information from various sources, enhancing the model's robustness and precision in predicting glaucoma. The distinctive contribution lies in our multi dataset CNN's ability to efficiently learn and extract intricate features and patterns from these diverse datasets, enabling the identification of subtle glaucoma indicators often overlooked by traditional diagnostic methods or single-dataset models. Consequently, our research aims to amplify the sensitivity and specificity of glaucoma detection, facilitating earlier and more accurate disease identification.

Furthermore, we explore interpretability and visualization methods to gain insights into the model's decision-making process, which can aid in understanding the underlying factors contributing to glaucoma prediction. Finally, we focus on model deployment to make the developed CNN-based glaucoma prediction system accessible and usable in clinical settings.

By following this comprehensive methodology, we aim to achieve accurate and efficient glaucoma prediction, enabling early detection and intervention to mitigate vision loss associated with this debilitating disease.

3. Methodology

The present study aims to investigate the predictive capability of a multi-dataset Convolutional Neural Network (CNN) model trained on OCT fundus scans from three distinct datasets: ORIGA, ACRIMA, and REFUGE. The study utilizes a diverse dataset consisting of 1755 images obtained from these multiple sources. The dataset is carefully curated and encompasses a wide range of demographics and characteristics to ensure the robustness and generalizability of the model. The dataset is separated into three sections: training, validation, and testing. 70% of the total samples are in the training set., while the validation set and test set contain 20% and 10% of the samples, respectively.

3.1. Data Collection

To address potential biases in demographics, data collection methods, and image quality, we employed a multi-dataset approach. Three distinct datasets were utilized: ORIGA, ACRIMA, and REFUGE. This selection aimed to

eliminate bias and provide a more comprehensive representation of glaucoma cases.
One of the challenges in glaucoma detection is the limited availability of positive samples. For instance, the ORIGA dataset contained only 168 positive samples, and the REFUGE dataset had 40 positive samples. Such limited sample sizes can impact model robustness and generalizability. To address this issue, we incorporated the ACRIMA dataset, which provided an additional number of positive samples, with 396 samples. By combining datasets, we aimed to enhance the overall sample size and improve the performance of the models.
Class imbalance refers to situations where one class has significantly more samples than the other. In the ORIGA dataset, for example, there were 482 negative samples compared to 168 positive samples, resulting in a class imbalance. Class imbalance can pose challenges during model training and evaluation, particularly in effectively capturing the minority class. To mitigate this issue, the inclusion of the ACRIMA dataset helped balance the class distribution by providing a greater number of positive samples. This step aimed to improve the model's ability to learn from and accurately classify both classes.
To summarize, the ACRIMA dataset significantly aided in rectifying class imbalance issues, notably improving the model's capacity to classify glaucoma instances accurately. By offering a more balanced representation of both positive and negative cases, ACRIMA effectively mitigated biases favoring the majority class during model training.
Datasets collected from different sources or at different times may exhibit variations in image quality, imaging devices, lighting conditions, or annotation quality. These variabilities can introduce inconsistencies and impact model performance and consistency. To account for data collection variability, samples were extracted from the ORIGA, ACRIMA, and REFUGE datasets, each with its unique characteristics. This approach allowed us to train models on diverse data, incorporating different image shapes, luminosities, and data collection protocols.
To ensure consistency and enhance model performance, appropriate data preprocessing techniques were applied using Keras TensorFlow. Notably, data normalization was performed to standardize the input data across the datasets. This normalization step helped eliminate potential biases arising from variations in image intensities and facilitated fair model training and evaluation.
To summarize, the dataset analysis phase encompassed addressing dataset bias through the utilization of multiple datasets, mitigating the challenges of limited sample size and class imbalance by incorporating the ACRIMA dataset, accounting for data collection variability by including diverse datasets, and applying data preprocessing techniques, such as normalization, to enhance model performance and fairness. These considerations ensured a more comprehensive and robust analysis of glaucoma detection using image processing methodologies.

3.2. Data Preprocessing

In this study, the researchers employed TensorFlow's dataset API, which offers a convenient way to build data pipelines instead of loading the entire dataset into memory at once. By using this API, the researchers were able to efficiently handle and process large amounts of data.
To load the data into the pipeline, the researchers utilized Keras helpers, a set of utility functions provided by the Keras library. These helpers enabled them to load the data from their designated folder and organize it into a dataset. Additionally, the dataset was split into two classes, likely corresponding to the positive (glaucoma) and negative (non-glaucoma) samples.
To ensure randomness and avoid bias during training, the images in the dataset were shuffled. This step helps to break any inherent order or patterns in the data, making the training process more robust and preventing the model from learning biased associations based on the input sequence.
To facilitate efficient training, the researchers divided the entire dataset into batches of 32 images. This batch size can be adjusted based on the specific requirements of the model and the available computational resources.
Furthermore, the pixel values of the images were scaled down from the range of "0 to 255" to the range of "0 to 1". This scaling process serves two main purposes:
Numerical Stability: Working with pixel values in the range of 0 to 1 can enhance numerical stability during calculations or when applying various algorithms. It helps prevent issues like overflow or underflow that may arise when dealing with larger pixel values.
Normalization: Scaling the pixel values to the range of 0 to 1 is a form of normalization. Normalizing the pixel

values ensures that the data has a consistent scale across different images or channels, facilitating easier comparison and processing. This normalization step can help the model effectively learn from the data and make accurate predictions.

3.3. The Multi-Datasets CNN (MD-CNN) Architecture

It is introduced a Multi-Datasets Convolutional Neural Network. It is similar to a traditional neural network in that it combines different neuron layers, learning rates, and other characteristics. In this study, we used five distinct CNN methodologies for network selection (convolution Layer, Max Pooling Layer, Flatten, Fully Connected Layer, and Classification Layer), as shown in Figure 1. Figure 2 depicts the working procedure of the suggested A Multi-Datasets Convolutional Neural Network. The CNN that was utilized is discussed further below.

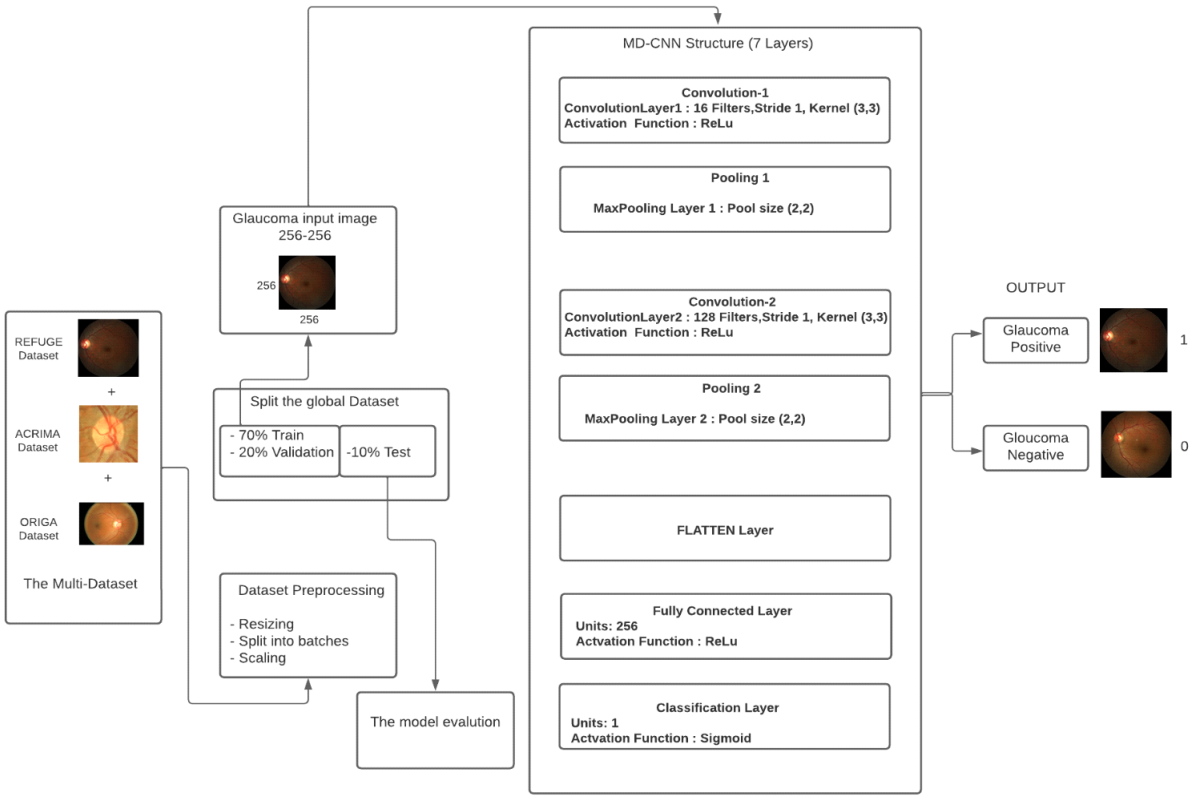

Figure 1. The illustration of the MD-CNN's operational procedure.

In this section, we employ a CNN architecture for glaucoma detection, as illustrated in Figure 2. The CNN, known as MD-CNN, comprises seven layers, including 2 convolutional layers, 2 max-pooling layers, a Flatten layer, and fully connected Layer and classification layers (refer to Table 1 for details), each serving a specific purpose in the detection process.

The convolution layer's primary role is feature extraction, achieved by applying a convolution function followed by an activation function. Multiple convolution layers are employed for this process.

The convolution operation uses a linear function known as the kernel or filter to extract features from the input image tensor I of dimensions $m_1 \times m_2 \times m_c$, where m_1, m_2, m_c represent the height, width, and number of channels, respectively.

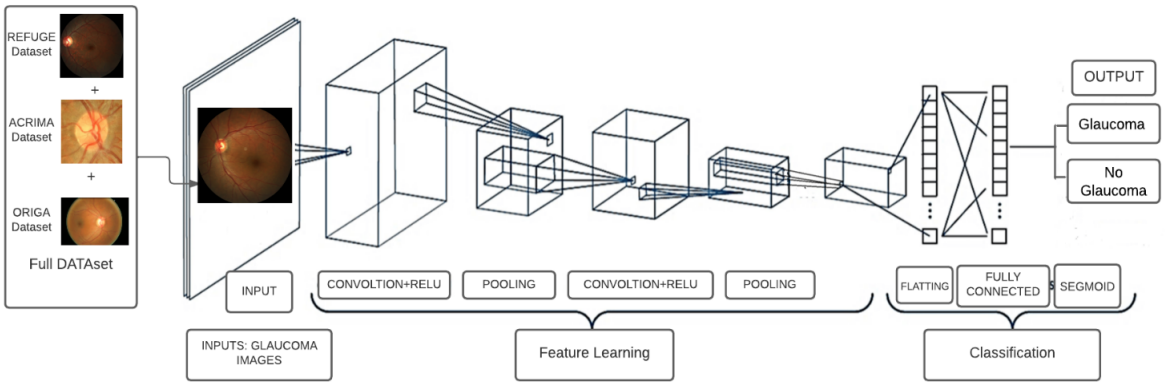

Figure 2. Multi-Dataset Deep Convolutional Neural Network Architecture (MD-CNN).

Table 1. Description of Layers of the multi-datasets CNN (MD-CNN).

CNN Layers Description with Learning Rate = 0.001 and Epochs = 40		
Layer	*Layers Description*	Parameters
0	Data Input	Input Shape: (256, 256, 3)
1	Convolution	Filters: 16 Kernel Size: (3, 3) Stride: 1 Activation: ReLU
2	MaxPooling	Pool Size: (2, 2) Strides: None
3	Convolution	Filters: 128 Kernel Size: (3, 3) Stride: 1 Activation: ReLU
4	MaxPooling	Pool Size: (2, 2) Strides: None
5	Flatten	-
6	Fully-Connected	Units: 256 Activation: ReLU
7	Classification	Units: 1 Activation: Sigmoid

A tensor of dimensions $(n_1 \times n_2 \times n_c)$ acts as a filter applied to the image (with the same number of channels). Where n_1, n_2, n_c represent the height, width, and number of channels, respectively. This filter moves across the image, performing element-wise multiplication between the corresponding parts of I and K, summing up these products. The stride determines the movement step size of the filter. The resulting tensor from this operation is of dimensions $(m_1 - n_1 + 1) \times (m_2 - n_2 + 1) \times 1$. So we can obtain:

$$Dim(I) = m_1 \times m_2 \times m_c \qquad (1)$$

$$Dim(K) = n_1 \times n_2 \times n_c \qquad (2)$$

$$Dim(F) = (m_1 \times m_2 \times m_c) \times (n_1 \times n_2 \times n_c) \times 1 \quad (3)$$

And,

$$F[i,j] = (I * K)_{[i,j]} \quad (4)$$

Where F represents the result of the kernel on the image. The i,j-th entry of the feature map is given as below:

$$F[i,j] = \sum_{x}^{m1} \sum_{y}^{m2} \sum_{z}^{mc} K_{[x,y,z]} I_{[i+x-1,j+y-1,z]} \quad (5)$$

The activation function is typically applied after adding a bias term 'b' to the convoluted part:

$$c = F + b \quad (6)$$
$$c = I * K + b \quad (7)$$

$$Conv(I,K) = \varphi_a(c) \quad (8)$$

Where a represents an activation function. Various activation functions, such as sigmoid, tangent, and hyperbolic tangent functions, exist. However, in our case, we employ ReLU, which effectively removes negative values:

$$R(x) = \max(0,x) \quad (9)$$

The pooling layer decreases the spatial dimensions of the convoluted features, capturing key image features. It applies a pooling function to the output from the convolution layer. Consider the following assumption:

$$Conv(I,K) = C \quad (10)$$
$$P = \varphi_p(C) \quad (11)$$

Where $\varphi_p(C)$ is an pooling function. The dimension of pooled part is given as :

$$Dim(p) = (\frac{m1 + 2p - n1}{s}) \times (\frac{m2 + 2p - n2}{s}) \times mc \quad (12)$$

The fully connected layer processes the flattened vector to generate another vector. To address class imbalances in machine learning models, balanced weights are used alongside the pooled section, introducing a bias term before applying the activation function. Here is the mathematical representation:

$$X = \sum_{i} w_i p_i + b \quad (13)$$

$$z = g(X) \quad (14)$$

Where g is an activation function for the fully connected layer. The architecture includes two convolutional layers responsible for extracting features from the input image by applying filters of size 3 × 3. The first convolutional layer operates on the initial input image of size 256 × 256, utilizing 16 filters and a stride of 1. Subsequently, the second convolutional layer reduces the image size to 125 × 125 while employing 128 filters, maintaining the same filter size and stride as before. Following the convolutional layers, two max-pooling layers are employed to down sample the feature maps, capturing the most prominent features. The max-pooling layers serve in reducing the spatial dimensions of feature maps while maintaining the most critical information. Next, a Flatten layer is utilized to reshape the pooled feature maps into a one-dimensional vector, preparing the data for the subsequent fully connected layer. The fully connected layer is responsible for capturing complex relationships among the features, enabling higher-level understanding and abstraction. Finally, the classification layer generates the glaucoma disease label based on the learned features and provides the final prediction of whether the input image exhibits signs of glaucoma or not. Overall, this CNN architecture, MD-CNN, effectively leverages the layers' functions, including convolution, pooling, flattening, and fully connected layers, to extract meaningful features from the input image and make accurate glaucoma disease predictions.

4. Results and Discussion

4.1. Data Analysis

The datasets we will use are ORIGA, ACRIMA, and REFUGE, each contributing to the study of glaucoma detection.
The ORIGA dataset comprises a total of 650 images, capturing the entire eye fundus. Among these images, 482 are classified as glaucoma negative, indicating the absence of glaucoma, while 168 images are classified as glaucoma positive. The images in the ORIGA dataset have a size of 2518x2048 pixels, and each image has an average file size of approximately 350KB. With its comprehensive view of the eye fundus, the ORIGA dataset provides valuable data for glaucoma detection research.
The ACRIMA dataset consists of 705 images that focus on capturing the optic disc of the eye. Out of these images, 309 are classified as glaucoma negative, and 396 are classified as glaucoma positive. The images in the ACRIMA dataset have a size of 349x349 pixels, and each image has an average file size of approximately 17KB. Unlike the ORIGA dataset, which provides a broader view of the eye fundus, ACRIMA images concentrate specifically on the optic disc, adding a different perspective to glaucoma detection studies.
The REFUGE dataset includes 400 images that capture the entire eye fundus. These images have a darker shade compared to those in the ORIGA dataset. Among the 400 images, 360 are classified as glaucoma negative, indicating the absence of glaucoma, while 40 images are classified as glaucoma positive. The images in the REFUGE dataset have a size of 2124x2056 pixels, and each image has an average file size of approximately 1.7MB. The REFUGE dataset provides a distinct set of images for glaucoma detection research.
By combining these diverse datasets, we can obtain a comprehensive and varied collection of images, enabling a more robust analysis of glaucomatous pathology. The different sizes and characteristics of the images contribute to a more comprehensive understanding of glaucoma and facilitate the development of effective detection models.
This table provides a convenient overview of the dataset characteristics discussed above:

Table 2. Summary of Dataset Characteristics

Dataset Characteristics						
Dataset	***Total Images***	***Glaucoma Negative***	***Glaucoma Positive***	***Image Size***	***Average Image***	***Description***
ORIGA	650	482	168	2518x2048	350KB	Entire eye fundus
ACRIMA	705	309	396	349x349	17KB	Only optic disc
REFUGE	400	360	40	2124x2056	1.7MB	Entire eye fundus (darker lighting)
Merged	1755	1151	604	-	523.93KB	Combined datasets

Note that the ”Merged” row represents the combined dataset after merging all three datasets together, providing the total number of images, the count for glaucoma negative and glaucoma positive cases and the average size of images in the dataset.
The following table shows a simplified version of the results we got after the preprocessing phase:

Table 3. Simplified Summary of Dataset Characteristics

Dataset Description	
Dataset	***Description***
ORIGA	Entire eye fundus, Large images
ACRIMA	Only optic disc, Small images
REFUGE	Entire eye fundus (darker lighting), Large images
Merged	Combined datasets

4.2. Data Splits

Having discussed the characteristics of the ORIGA, ACRIMA, and REFUGE datasets, we now shift our focus to the dataset split for training, validation, and testing purposes. It is crucial to appropriately partition the data to ensure robust model training, evaluation, and generalization. In this study, we adopt a commonly used split ratio, where 70% of the dataset is allocated for training, 20% for validation, and 10% for testing. This division allows us to effectively train the model on a large portion of the data while still reserving independent sets for validation and final evaluation. By following this data split strategy, we can assess the performance and generalizability of our proposed glaucoma prediction model accurately. This table provides a convenient overview of the data splits:

Table 4. Summary of Data Splits

Data Splits		
Split	***Number of Samples***	***Percentage***
Train	1228	70%
Validation	351	20%
Test	160	10%

4.3. Test and Evaluation

In the context of our glaucoma prediction model, we utilize various metrics to evaluate its performance: Precision, Accuracy, Sensitivity, AUC (Area Under the Curve), Specificity.
We now turn our attention to the confusion matrix, which provides a more detailed breakdown of the predictions made by the model. The confusion matrix lets us know how many true positives, true negatives, false positives, and false negatives we got from our classification procedure. The number of successfully anticipated positive instances is represented by true positives, whereas the number of accurately anticipated negative instances is represented by true negatives. On the other hand, false positives represent the number of negative cases that were mistakenly classed as positive, while false negatives represent the number of positive instances that were mistakenly classified as negative. By analyzing the confusion matrix, we gain a comprehensive understanding of the model's predictive capabilities and its performance in distinguishing between different classes. Let us now delve into the specifics of the confusion matrix and interpret the results obtained from our model on the test split.

We can now use the results from the confusion matrix and its breakdown of true positives, true negatives, false positives, and false negatives to construct the assessment metrics stated earlier. These metrics, including precision, accuracy, sensitivity, specificity, and AUC, rely on the information provided by the confusion matrix. By analyzing the true positives, true negatives, false positives, and false negatives, we can compute these metrics, which provide valuable insights into the performance of our model in glaucoma prediction. Let's proceed to calculate and analyze these metrics to gain a comprehensive understanding of the effectiveness of our model in accurately detecting and classifying glaucoma cases.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{15}$$

$$Accuracy = \frac{TruePositives + TrueNegatives}{TotalPredictions} \tag{16}$$

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{17}$$

$$Specificity = \frac{TrueNegatives}{TruePositives + FalseNegatives} \tag{18}$$

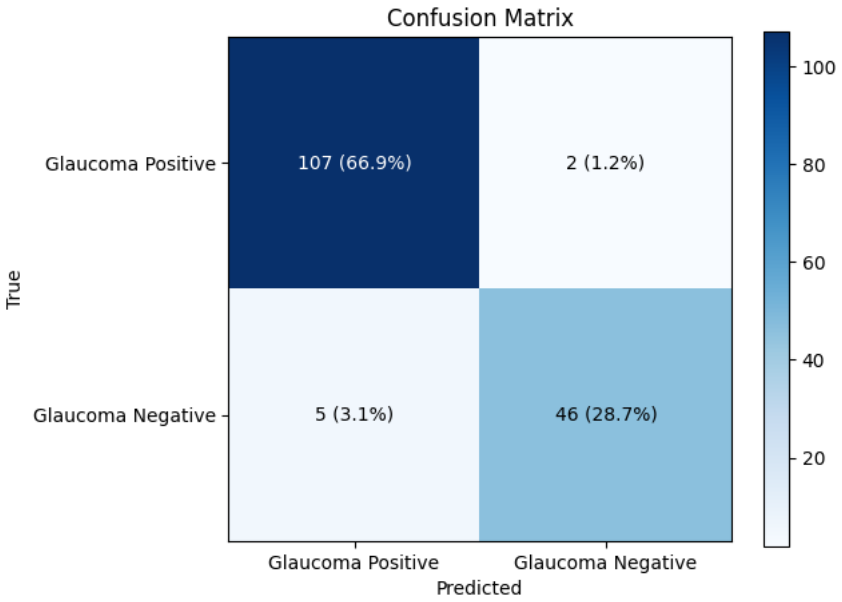

Figure 3. Confusion Matrix of Glaucoma prediction using MD-CNN.

AUC (Area Under the Curve) is calculated by plotting the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity) at various classification thresholds. The resulting curve is known as the Receiver Operating Characteristic (ROC) curve. The AUC is the area under the curve that shows the model's ability to differentiate between positive and negative cases.

To compute the AUC, the ROC curve is divided into small trapezoids, and the area of each trapezoid is calculated. These areas are then summed up to obtain the overall AUC value. The AUC ranges from 0 to 1, with 0.5 indicating random classification, and a value of 1 represents perfect classification. Based on the calculations utilizing the confusion matrix's true positives, true negatives, false positives, and false negatives, we have determined the following evaluation metrics for our model when used on the test split:

Table 5. Performance of the Proposed Algorithm.

Performance Metrics	
Metric	***Value (%)***
Precision	94.34
Accuracy	96.88
AUC (Area Under the Curve)	99.01
Sensitivity	94.34
Specificity	97.20

The evaluation metrics for our model indicate high performance in glaucoma prediction: Precision refers to the proportion of correctly identified positive cases (true positives) out of all predicted positive cases (true positives + false positives). With a precision of 94.34%, our model demonstrates a high level of accuracy in identifying actual glaucoma cases.

Accuracy is the overall correctness of the predictions and is derived by dividing the total number of forecasts by the sum of true positives and true negatives. With an accuracy of 96.88%, our model exhibits a strong ability to correctly classify both glaucoma-positive and glaucoma-negative cases.

AUC (Area Under the Curve) is a measure of the model's ability to discriminate between glaucoma-positive and glaucoma-negative cases. With an AUC of 99.01%, our model demonstrates excellent discriminatory power, indicating that it can effectively distinguish between the two classes. The proportion of genuine positive cases (true positives) accurately recognized by the model is measured by sensitivity, also known as the true positive rate. With a sensitivity of 94.34%, our model exhibits a high capability to correctly identify glaucoma-positive cases.

Specificity represents the true negative rate, which measures the proportion of actual negative cases (true negatives) correctly identified by the model. With a specificity of 97.20%, our model shows a strong ability to correctly identify glaucoma-negative cases.

These metrics collectively demonstrate the effectiveness of our model in accurately predicting glaucoma and highlight its potential for aiding in early detection and intervention to mitigate vision loss associated with this debilitating disease.

Having examined the metrics for glaucoma prediction, we now shift our focus to the analysis of the loss and accuracy graphs obtained during the training phase. These graphs provide valuable insights into the performance and progress of our model throughout the training process. By studying the trends and patterns exhibited by these graphs, we can gain a deeper understanding of the model's learning dynamics, identify potential issues or improvements, and make data-driven decisions for further optimization. Let us now delve into the analysis of these crucial visual representations.

The analysis of the provided accuracy and loss graphs reveals promising results for the model's training progress over the 40 epochs.

Firstly, the train accuracy reaching 98.93% indicates that the model has successfully learned the patterns and features present in the training data, achieving a high level of accuracy in classifying glaucoma cases. This high train accuracy suggests that the model has effectively captured the underlying information and can make accurate predictions on the training set.

Secondly, the validation accuracy reaching 92.05% indicates that the model has also generalized well to unseen data. This high validation accuracy suggests that the model's performance extends beyond the training set, achieving a commendable level of accuracy on previously unseen samples. This demonstrates the model's ability to generalize and make accurate predictions on new data.

Regarding the loss values, the train loss reaching 0.0328% indicates that the model has effectively minimized the errors during the training process. This low train loss suggests that the model has successfully converged towards an optimal set of weights, resulting in a minimal discrepancy between the predicted and actual values on the training set.

Similarly, the validation loss reaching 0.2043% indicates that the model's performance on the validation set is also relatively low. This low validation loss demonstrates the model's ability to generalize well and make accurate

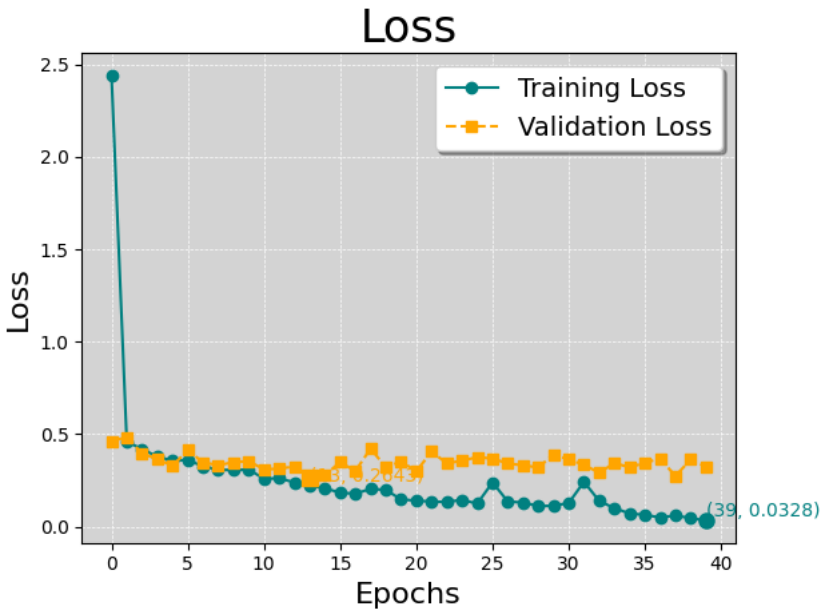

Figure 4. The Validation and Training Loss Functions are graphically represented over a range of epochs..

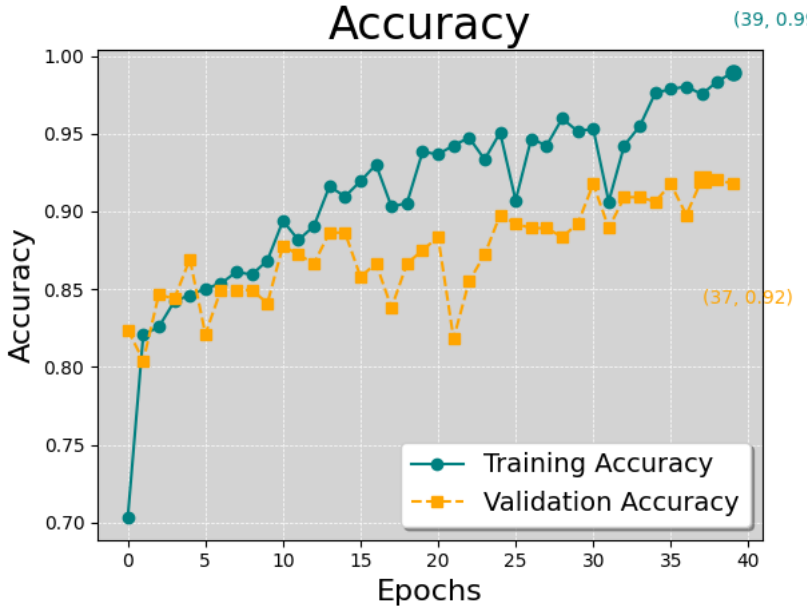

Figure 5. Validation and Training Accuracy Graphical Representation across Various Iterations.

predictions on new, unseen data points.

The fact that these results were achieved at the 40th epoch indicates that the model continued to improve and refine its performance as the training progressed. It demonstrates that the model benefited from additional training iterations, allowing it to reach its maximum performance at the end of the training process.

Overall, the numbers obtained for train accuracy, validation accuracy, train loss, and validation loss indicate that

the model learnt effectively, generalized well, and achieved a high level of accuracy and performance in classifying glaucoma cases.

4.4. Comparison between our study and state-of-art studies

In order to assess the efficacy of our suggested glaucoma prediction model, it is crucial to compare its accuracy with the findings of other studies conducted in the same domain. By examining the outcomes and performance metrics achieved by different models, we can gain valuable insights into the overall efficacy of our approach. This comparative analysis will provide a broader perspective on the state-of-the-art in glaucoma prediction and enable us to assess the reliability and robustness of our model in accurately detecting this sight-threatening condition.

Table 6. Comparative Analysis with Previous Works.

Comparative Analysis with Previous Works			
Cited	*Methodology*	Accuracy (%)	Year
[18]	CNN Semi-Supervised Transfer Learning	92.4	2019
[24]	OCT Probability map using CNN	95.7	2019
[20]	AG-CNN	95.3	2019
[25]	CNN	88.2	2020
Multi-Dataset CNN (Proposed Approach)		98.93	-

5. Limitations

5.1. Dataset Constraints:

ORIGA, ACRIMA, REFUGE Limitations: Each dataset utilized in this study possesses distinct characteristics, including variations in image quality, size, and representativeness. These variations across datasets could potentially impact the model's performance and its ability to generalize across diverse datasets. Inadequate Sample Size: Despite deliberate efforts to merge multiple datasets to augment the sample size, it remains possible that the combined dataset's size might still be insufficient for conducting a more comprehensive analysis. This limitation could potentially affect the model's capability to generalize effectively.

5.2. Assumptions and Biases:

Inherent Biases: The datasets utilized in this study might carry inherent biases stemming from data collection methods, variations in imaging quality, or discrepancies in demographic representation among datasets. Such biases might have inadvertently influenced the model's training and predictions. Assumptions in Model Development: Certain assumptions made during data preprocessing, feature extraction, or model development might have impacted the model's outcomes. These assumptions could have introduced bias or limitations in the model's performance.

5.3. Generalizability:

Model Applicability Beyond Current Context: Limitations might exist regarding the model's applicability to different populations, settings, or scenarios due to dataset-specific characteristics or potential biases. This could potentially restrict the model's generalizability beyond the studied contexts. Limited Computing Resources: Constraints related to computational power or hardware limitations might have impacted the complexity of model training or testing. This limitation could have influenced the depth or scope of the analysis conducted.

5.4. Future Directions:

Methodology Improvement: There is a need to explore alternative methods or consider incorporating additional datasets to mitigate the identified limitations. This step aims to enhance the model's robustness and generalizability. Enhancement Opportunities: Utilizing more advanced computational resources for intricate model training and testing could address the limitations related to computing constraints. Integrating more diverse datasets could enhance the variability and reliability of the model, thereby improving its overall performance.

6. Conclusion

In conclusion, our study aimed to develop and evaluate a multi-dataset Convolutional Neural Network (CNN) for glaucoma prediction. By leveraging diverse datasets and employing a comprehensive methodology, we successfully built a robust model.

Our results demonstrated the effectiveness of the multi-dataset CNN in predicting glaucoma. With a training accuracy of 98.93% and a validation accuracy of 92.05%, the model exhibited strong performance. Furthermore, the model's performance was evaluated based on loss metrics. The training loss reached 0.0328%, and the validation loss was 0.2043%, indicating effective optimization and error minimization.

Comparative analysis with other studies highlighted the advantages of our multi-dataset CNN, showcasing competitive performance.

The visualizations of the accuracy and loss graphs provided valuable insights into the model's training progress, demonstrating its ability to capture relevant features and minimize errors.

In summary, our study contributes to the advancement of glaucoma prediction using a multi-dataset CNN. The achieved results and comparative analysis validate the effectiveness of our approach. Future research can focus on enhancing the model's performance and exploring interpretability techniques. Accurate glaucoma prediction models have the potential to improve patient outcomes.

REFERENCES

1. Q. Abbas, *Glaucoma-deep: detection of glaucoma eye disease on retinal fundus images using deep learning*, Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 6, 2017.
2. C. Cook and P. Foster, *Epidemiology of glaucoma: what's new?*, Can. J. Ophthalmol., vol. 47, no. 3, pp. 223–226, 2012.
3. Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, *Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis*, Ophthalmology, vol. 121, no. 11, pp. 2081–2090, 2014.
4. A. Chakib, N. Ouarrach, M. Haloui, M. Elbelhadji, and A. Amraoui, *Viscocanalostomy: Preliminary clinical results [Viscocanalostomie: résultats cliniques à court terme]*, 2010.
5. K. J. Canuto et al., *A scoping review of Aboriginal and Torres Strait Islander health promotion programs focused on modifying chronic disease risk factors*, Heal. Promot. J. Aust., vol. 32, no. 1, pp. 46–74, 2021.
6. D. B. Rein et al., *The economic burden of major adult visual disorders in the United States*, Arch. Ophthalmol., vol. 124, no. 12, pp. 1754–1760, 2006.
7. Y. Hagiwara et al., *Computer-aided diagnosis of glaucoma using fundus images: A review*, Comput. Methods Programs Biomed., vol. 165, pp. 1–12, 2018.
8. B. A. Francis et al., *Novel glaucoma procedures: a report by the American Academy of Ophthalmology*, Ophthalmology, vol. 118, no. 7, pp. 1466–1480, 2011.
9. Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature, vol. 521, no. 7553, pp. 436–444, 2015.
10. M. D. Abràmoff et al., *Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning*, Invest. Ophthalmol. Vis. Sci., vol. 57, no. 13, pp. 5200–5206, 2016.
11. V. Gulshan et al., *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs*, jama, vol. 316, no. 22, pp. 2402–2410, 2016.
12. N. Gupta and Y. H. Yücel, *Glaucoma as a neurodegenerative disease*, Curr. Opin. Ophthalmol., vol. 18, no. 2, pp. 110–114, 2007.
13. M. Hashemi and H. Karimi, *Weighted machine learning*, Stat. Optim. Inf. Comput., vol. 6, no. 4, pp. 497–525, 2018.
14. M. Fayaz, A. Abadi, and S. Khodakarim, *The Functional Regression With Reconstructed Functions From Hybrid Principal Components Analysis: With EEG-fMRI Application*, Stat. Optim. Inf. Comput., vol. 10, no. 3, pp. 890–903, 2022.
15. I. Grabec, E. Švegl, and M. Sok, *A method for automatic medical diagnosis*, Stat. Optim. Inf. Comput., vol. 7, no. 1, pp. 26–39, 2019.
16. R. Daneshvar et al., *Prediction of glaucoma progression with structural parameters: comparison of optical coherence tomography and clinical disc parameters*, Am. J. Ophthalmol., vol. 208, pp. 19–29, 2019.

17. J. M. Ahn, S. Kim, K.-S. Ahn, S.-H. Cho, K. B. Lee, and U. S. Kim, *A deep learning model for the detection of both advanced and early glaucoma using fundus photography*, PLoS One, vol. 13, no. 11, p. e0207982, 2018.
18. G. An et al., *Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images*, J. Healthc. Eng., vol. 2019, 2019.
19. N. Sengar, M. K. Dutta, R. Burget, and M. Ranjoha, *Automated detection of suspected glaucoma in digital fundus images*, in 2017 40th International Conference on Telecommunications and Signal Processing (TSP), 2017, pp. 749–752.
20. M. Aamir et al., *An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification*, Diagnostics, vol. 10, no. 8, p. 602, 2020.
21. V. S. Mary, E. B. Rajsingh, and G. R. Naik, *Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey*, IEEE Access, 2016.
22. X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, and J. Liu, *Glaucoma detection based on deep convolutional neural network*, in 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC), 2015, pp. 715–718.
23. A. Sevastopolsky, *Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network*, Pattern Recognit. Image Anal., vol. 27, pp. 618–624, 2017.
24. A. Serener and S. Serte, *Transfer learning for early and advanced glaucoma detection with convolutional neural networks*, in 2019 Medical technologies congress (TIPTEKNO), 2019, pp. 1–4.
25. A. Saxena, A. Vyas, L. Parashar, and U. Singh, *A glaucoma detection using convolutional neural network*, in 2020 international conference on electronics and sustainable communication systems (ICESC), 2020, pp. 815–820.