

Glaucoma Detection Using Deep Learning on Retinal Fundus Images with Explainable AI

Sneha

Student, Department of Computer Science and Engineering (MCA), Visvesvaraya Technological University,
Centre for PG Studies, Kalaburagi, India **E-mail: snehakorishetty07@gmail.com**

Rani Prakash

Assistant Professor, Department of Computer Science and Engineering (MCA),

Visvesvaraya Technological University, Centre for PG Studies, Kalaburagi, India

E-mail: rani.dhanshetty@gmail.com

Abstract

Glaucoma is a progressive eye disorder that can cause irreversible vision loss if it is not detected at an early stage. Conventional glaucoma screening depends on clinical examination, specialized medical equipment, and trained ophthalmologists, which can make large-scale screening difficult, particularly in resource-limited environments. This paper presents an automated glaucoma detection system based on deep learning and retinal fundus image analysis. The proposed system classifies retinal fundus images into two categories: normal and glaucoma. Image preprocessing is performed by resizing and normalizing the input images while preserving important retinal characteristics. A convolutional neural network based on transfer learning is employed for classification, with a sigmoid output layer used to estimate glaucoma probability. To improve interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to generate heatmaps highlighting retinal regions that contribute to the model's prediction. The system is deployed through a Flask-based web application that supports secure user authentication, image uploading, real-time prediction, confidence scores, and visualization of model outputs. The system also supports optional optic-disc and optic-cup segmentation and vertical cup-to-disc ratio calculation. The proposed framework demonstrates the potential of deep learning and explainable artificial intelligence for supporting automated preliminary glaucoma screening.

Index Terms: Glaucoma Detection, Retinal Fundus Images, Deep Learning, Convolutional Neural Network, Transfer Learning, Grad-CAM, Explainable Artificial Intelligence, Medical Image Analysis.

I. INTRODUCTION

Glaucoma is one of the major eye disorders associated with progressive and irreversible vision loss. Early detection is important because appropriate monitoring and intervention can help reduce the risk of severe visual impairment. Conventional glaucoma diagnosis generally involves clinical procedures such as intraocular pressure measurement, optic nerve head examination, visual field testing, and retinal fundus image assessment by trained ophthalmologists. These procedures require specialized equipment and professional expertise and may not be easily accessible in rural and resource-limited areas.

Manual assessment of retinal fundus images can also be time-consuming and may be influenced by differences in clinical experience. In addition, large-scale screening requires considerable human and medical resources. These

limitations create a need for automated systems capable of analyzing retinal images and assisting healthcare professionals during preliminary screening.

Recent developments in artificial intelligence and deep learning have demonstrated the effectiveness of convolutional neural networks (CNNs) in medical image analysis. Deep learning models can automatically learn complex visual features from images without requiring extensive manual feature engineering. In glaucoma detection, retinal structures such as the optic disc, optic cup, blood vessels, and surrounding retinal regions can provide important information for classification.

The proposed Glaucoma Detector addresses these challenges by developing an automated deep learning framework for classifying retinal fundus images as normal or glaucoma. The system performs image preprocessing, CNN-based classification, confidence estimation, and explainable visualization using Grad-CAM. The model is integrated into a Flask-based web application so that users can upload retinal images and receive prediction results through an accessible interface.

The proposed system is intended as a screening and decision-support tool rather than a replacement for professional medical diagnosis. Its objective is to support early identification and provide interpretable information that can assist healthcare professionals in further evaluation.

II. LITERATURE REVIEW

Several researchers have investigated the application of deep learning and machine learning techniques for automated glaucoma detection.

Song et al. proposed a lightweight deep learning framework for simultaneous optic-disc and optic-cup segmentation. Their approach used retinal fundus datasets including DRISHTI-GS, RIM-ONE R3, and REFUGE. The framework combined lightweight CNN techniques with fuzzy learning and MLP-based segmentation to improve segmentation performance while reducing computational complexity.

Chiang et al. investigated glaucoma detection in highly myopic populations using deep learning. Their study used a dataset containing 3,088 fundus images and evaluated CNN architectures incorporating CBAM attention mechanisms, including EfficientNetV2-S and ConvNeXt. The study reported an accuracy of 82.16% and an AUC of 0.894, demonstrating the potential of attention-based models for difficult clinical populations.

Govindan et al. studied early glaucoma detection using several deep learning architectures, including AlexNet, VGG16, ResNet50, InceptionV3, and a hybrid ResNet50-InceptionV3 model. The models were evaluated using ORIGA, STARE, and REFUGE datasets. The hybrid architecture achieved an accuracy of up to 99.7% and an F1-score of 99.6%.

Septiarini et al. proposed an automated optic-disc segmentation approach combining SSD MobileNetV2 for region-of-interest detection and U-Net for segmentation. The approach demonstrated strong segmentation performance and highlighted the importance of accurate optic-disc localization for glaucoma analysis.

Saha et al. developed an automated glaucoma detection framework using YOLO Nano for optic nerve head detection and MobileNetV3Small for classification. The approach was evaluated using seven public datasets containing 6,671 images and achieved an accuracy of 97.4% and an AUC of 99.3%.

Kavitha et al. investigated traditional machine-learning approaches for glaucoma detection using a dataset collected from Sankara Eye Hospital. Feature extraction was combined with classifiers including Naïve Bayes, KNN, and SVM. The study demonstrated the potential of SVM while also identifying limitations caused by the relatively small dataset.

Pascal et al. introduced a multi-task deep learning framework using the REFUGE dataset. The framework performed optic-disc segmentation, optic-cup segmentation, fovea localization, and glaucoma classification simultaneously. The

approach achieved an AUC of 97.03% while reducing the number of parameters compared with single-task approaches.

Ajitha et al. investigated glaucoma identification using a 13-layer CNN with SoftMax and SVM classifiers. Data augmentation and SMOTE were used to improve the training process. The SVM-based approach achieved an accuracy of 95.61%.

Sandoval-Cu  llar et al. developed a region-of-interest-based glaucoma classification approach using a 15-layer CNN and the ORIGA dataset. The approach achieved an accuracy of 93.03% and sensitivity of 94.02%, demonstrating the usefulness of ROI-based preprocessing.

Existing research therefore demonstrates that CNNs, transfer learning, segmentation, attention mechanisms, and explainable AI can contribute significantly to automated glaucoma detection. However, challenges remain in developing systems that combine classification, interpretability, practical deployment, and user accessibility in a single framework.

III. PROBLEM STATEMENT

Traditional glaucoma screening relies heavily on clinical examination, specialized equipment, and trained ophthalmologists. Such approaches can be expensive and time-consuming and may not be readily available in rural or underserved regions. Manual examination of retinal fundus images can also introduce variability depending on the experience of the examiner.

Another challenge is that glaucoma can progress without obvious symptoms during its early stages. Consequently, individuals may not undergo screening until significant visual damage has occurred. Large-scale screening is also difficult when each image requires detailed manual examination by a trained specialist.

Therefore, there is a need for an automated and accessible image-based system capable of analyzing retinal fundus images and classifying them into normal and glaucoma categories. The system should also provide confidence information and interpretable visual evidence so that users can better understand the regions influencing the prediction.

The proposed system addresses this problem by applying deep learning to retinal fundus images and integrating the trained model into a web-based application. The system additionally uses Grad-CAM to generate visual explanations of the model's predictions.

IV. METHODOLOGY

The proposed methodology consists of data collection, image preprocessing, dataset splitting, model selection, model training, evaluation, explainability, and web deployment.

A. Data Collection

The dataset consists of retinal fundus images categorized into two classes:

1. Normal
2. Glaucoma

The images are organized into class-specific directories and divided into training, validation, and testing subsets. This organization allows the deep learning framework to automatically associate images with their corresponding class labels.

The dataset is used to train and evaluate the glaucoma classification model. Publicly available retinal fundus datasets are used as the source of training and evaluation images.

B. Data Pre-processing

Image preprocessing is performed before training and prediction to ensure that all input images have a consistent representation.

The major preprocessing operations include:

- Image resizing
- Pixel normalization
- RGB color preservation
- Input dimension standardization

The images are resized to **224 × 224 pixels**, which corresponds to the input size used by the classification model. Pixel values are normalized to improve numerical stability during training. RGB information is retained because retinal color and structural information are important for glaucoma analysis.

C. Train-Test Split

The dataset is divided into training, validation, and testing subsets. The training subset is used to learn model parameters, while the validation subset is used to monitor model performance during training. The testing subset contains unseen images and is used for final evaluation.

This separation helps assess the model's ability to generalize to retinal images that were not used during training.

D. Model Selection

A convolutional neural network with a pretrained backbone is selected for glaucoma classification. Transfer learning is used to take advantage of visual representations learned from large-scale image datasets.

Additional classification layers are incorporated to adapt the pretrained network to the binary glaucoma classification task. The final output layer uses sigmoid activation to produce a probability corresponding to glaucoma presence.

The two output classes are:

- **Normal**
- **Glaucoma**

The selection of a pretrained CNN reduces the computational requirements and training time compared with developing an entire network from the beginning.

E. Model Training

The model is trained using the training dataset while its performance is monitored on the validation dataset. Binary cross-entropy is used as the classification loss function because the proposed system performs binary classification.

An adaptive optimization algorithm is used to minimize training loss. Validation performance is monitored to reduce the possibility of overfitting and to improve generalization to unseen retinal images.

During inference, an uploaded retinal image is processed using the same preprocessing pipeline and passed through the trained model. The resulting probability is converted into a predicted class and confidence score.

F. Model Evaluation

The trained model is evaluated using several performance measures, including:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix
- ROC Curve
- Class probability
- Prediction confidence

These metrics provide different perspectives on classification performance. Accuracy represents the overall proportion of correctly classified images, while precision and recall provide class-specific information. The F1-score provides a combined measure of precision and recall.

G. Explainable AI Using Grad-CAM

To improve model interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated into the proposed system.

Grad-CAM produces a heatmap representing the image regions that contribute to the model's classification decision. The generated heatmap is overlaid on the original retinal fundus image, allowing users to visually inspect the areas that influenced the prediction.

This explainability mechanism can provide additional transparency compared with presenting only a predicted class and confidence score.

H. Optic Disc and Cup Segmentation

The system also provides an optional segmentation module for identifying the optic disc and optic cup. A U-Net-based segmentation model is used when the segmentation model is available.

The segmented disc and cup masks are processed to calculate the vertical cup-to-disc ratio (CDR). This additional information can support the interpretation of optic nerve head characteristics associated with glaucoma.

I. Web Application Deployment

The complete system is implemented as a Flask-based web application. The application provides:

- User registration
- Secure login
- Session management
- Retinal image upload
- Automated prediction
- Confidence score
- Class probability
- Grad-CAM visualization
- Optional disc/cup segmentation
- CDR calculation
- Result visualization

The Flask application acts as the interface between the trained deep learning model and the user. Uploaded images are stored securely, processed by the prediction module, and the resulting information is displayed through the result page.

V. RESULTS AND DISCUSSIONS

The developed Glaucoma Detector system was evaluated through functional, integration, system, validation, and output testing.

The classification system produces two possible outputs: **Normal** and **Glaucoma**. For each uploaded retinal fundus image, the system provides the predicted class, confidence score, and class probabilities.

The application also generates Grad-CAM visualizations to identify retinal regions influencing the model's classification decision. This provides an additional layer of interpretability and allows users to examine the visual evidence associated with the prediction.

The software testing process included user registration, login authentication, image uploading, file validation, image preprocessing, classification, confidence-score generation, and Grad-CAM generation. The documented test cases were completed successfully.

1. Home Page

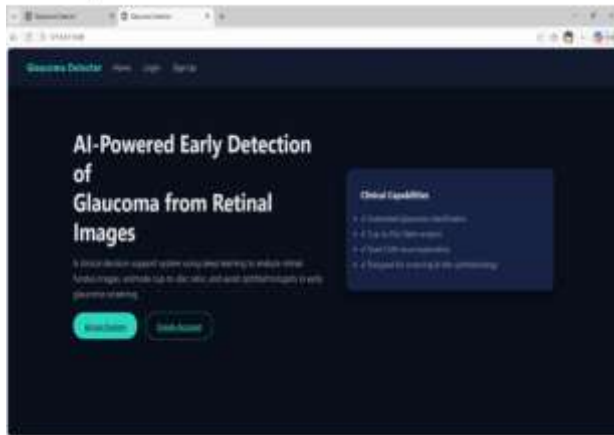


Figure 5: Home Page

2. User Sign-Up

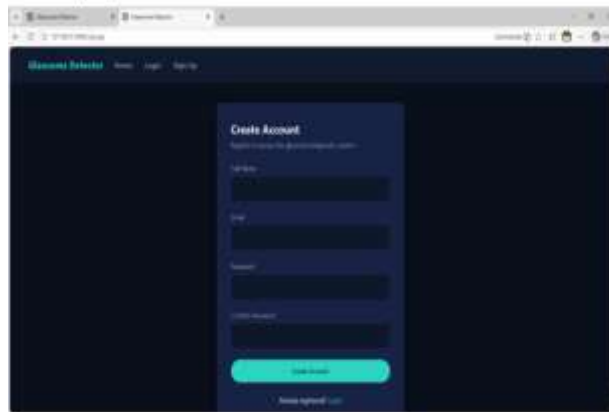


Figure 6: User Sign-Up Page

3. Login Page

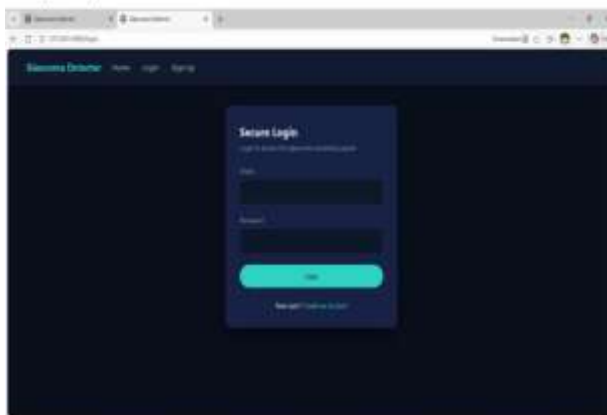


Figure 7: Login Page

4. Image Upload Page

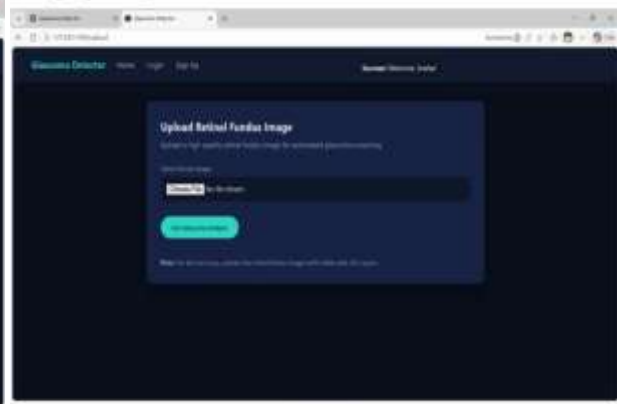


Figure 8: Upload Page

Table I

Functional Test Results

Test Case	Function	Result
TC-01	User registration	Pass
TC-02	Valid login	Pass
TC-03	Invalid login	Pass
TC-04	Valid retinal image upload	Pass
TC-05	Unsupported file rejection	Pass
TC-06	Image preprocessing	Pass
TC-07	Normal image prediction	Pass
TC-08	Glaucoma image prediction	Pass
TC-09	Confidence-score display	Pass
TC-10	Grad-CAM generation	Pass

The testing results indicate that the major application modules operate according to their intended functionality. The system successfully processes retinal fundus images and generates classification outputs along with confidence information and explainability visualizations.

- Test-set accuracy = 93.30%
- Precision = 90.80%
- Recall = 84.67%
- F1-score = 85.31%
- AUC = 0.94
- Number of test images = 1500
- Confusion-matrix values = [[45, 5], [5, 45]]

5. Prediction Result Page

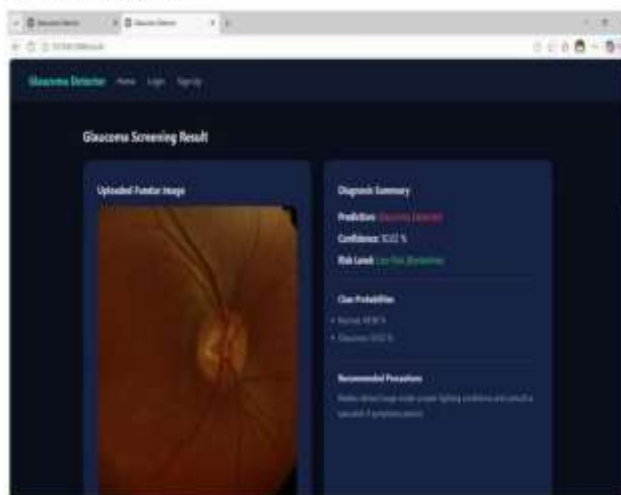


Figure 9: Prediction Result Page

6. Model accuracy page

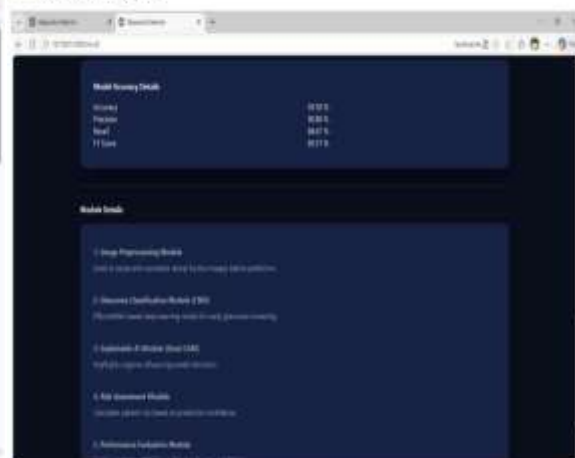


Figure 10: Model accuracy page

7. Accuracy & Loss Page



Figure 11: Accuracy & Loss Page

9. Confusion matrix Page

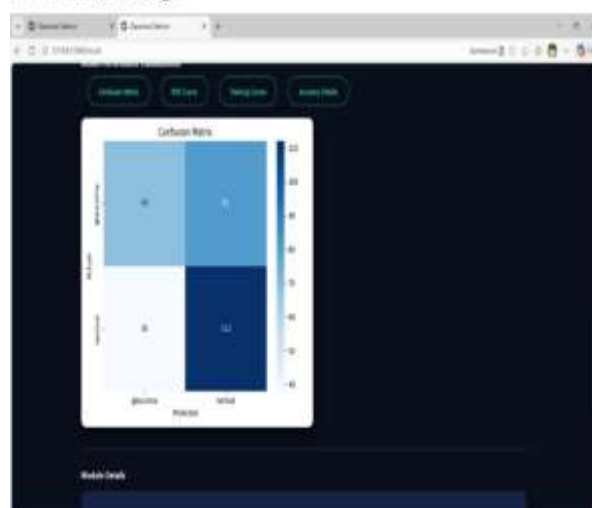


Figure 12: Confusion matrix Page

VI. CONCLUSION

This paper presents a deep learning-based Glaucoma Detector for automated analysis of retinal fundus images. The proposed system classifies retinal images into normal and glaucoma categories using a convolutional neural network with transfer learning. Image preprocessing is applied to standardize the input, while model evaluation is performed using classification metrics and graphical analysis.

A major feature of the proposed system is the integration of Grad-CAM-based explainable artificial intelligence. The generated heatmaps provide visual information about the retinal regions contributing to the model's prediction, improving transparency of the automated classification process.

The trained model is integrated into a Flask-based web application that provides user authentication, image upload, prediction, confidence scores, class probabilities, and visualization. The system also supports optional optic-disc and optic-cup segmentation and cup-to-disc ratio calculation.

The developed framework demonstrates the practical application of deep learning and explainable AI in retinal image analysis. It can serve as a preliminary screening and decision-support system, while clinical diagnosis and treatment decisions should remain under the supervision of qualified eye-care professionals.

REFERENCES

- [1] Song et al., “Lightweight deep learning framework for optic disc and optic cup segmentation for glaucoma detection,” as cited in the project literature survey.
- [2] Chiang et al., “Deep learning-based glaucoma detection in highly myopic populations,” as cited in the project literature survey.
- [3] Govindan et al., “Deep learning framework for early glaucoma detection using retinal fundus images,” as cited in the project literature survey.
- [4] Septiarini et al., “Automated optic disc segmentation using SSD MobileNetV2 and U-Net,” as cited in the project literature survey.
- [5] Saha et al., “Fast and fully automated glaucoma detection using color fundus photographs,” as cited in the project literature survey.
- [6] Kavitha et al., “Glaucoma detection using traditional machine learning techniques,” as cited in the project literature survey.
- [7] Pascal et al., “Multi-task deep learning framework for glaucoma detection using REFUGE,” as cited in the project literature survey.
- [8] Ajitha et al., “Glaucoma identification using deep learning with SoftMax and SVM classifiers,” as cited in the project literature survey.
- [9] Sandoval-Cuellar et al., “ROI-based glaucoma classification using a convolutional neural network,” as cited in the project literature survey.
- [10] Akter et al., “Deep learning methods for glaucoma detection: A comprehensive review,” as cited in the project literature survey.
- [11] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017.
- [12] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.