Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

AI-based Retinal Image Classification for Early Detection of Eye Diseases

Ananthakumari A.[1], Muthu Menaka N.[2], Kaniska M.[3]

[1]Assistant Professor, [2,3]Student, Department of Computer Science and Engineering, Dr. G U Pope College of Engineering, Thoothukudi, India.

E-mail: [1]kumari.cse2007@gmail.com, [2]muthumenaka705@gmail.com, [3]kaniska749@gmail.com

Abstract

Early detection of eye diseases are essential to prevent severe vision impairment and improve patient care. Eye diseases such as cataract, glaucoma, and diabetic retinopathy are among the leading causes of vision loss if not diagnosed at an early stage. The goal of this project is to use artificial intelligence to develop an automated classification system for retinal images, which uses retinal fundus images to classify the four most common retinal diseases (i.e., cataract, diabetic retinopathy, glaucoma, and normal). We designed an image classification system which uses the Vision Transformer (ViT) deep learning model to classify images into four categories: cataract, diabetic retinopathy, glaucoma, and normal. Before training occurred, all images were pre-processed by using various methods of resizing, normalizing, and applying data augmentations to enhance the model's performance and generalization. The performance metrics used to evaluate the overall performance of the trained model were accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. Overall, the proposed model achieved an average overall accuracy of 93.48% on the test dataset.

Keywords: Eye Disease Detection, Retinal Image Classification, Vision Transformer (ViT), Deep Learning, Medical Image Analysis, Streamlit.

1. Introduction

Preventing vision loss and maintaining healthy eyes begins with early detection of eye diseases, including cataracts, glaucoma and diabetic retinopathy which are the most three

DOI: https://doi.org/10.36548/jscp.2026.2.004
Extended version of article received from **International Conference on Next Generation Engineering Innovations, AI and Sustainable Automation - 2026**
Received: 23.03.2026, received in revised form: 27.04.2026, accepted: 14.05.2026, published: 26.05.2026

common causes for developing blindness. Early-stage detection and treatment of those will lower chances for developing permanent vision issues, as well. While traditional techniques require an ophthalmologist, who has received specialized training to manually inspect retinal photographs. Unfortunately, it often takes a long time for the ophthalmologists to complete their exams, making it difficult to provide care to people in need in many parts of the world.

Artificial intelligence (AI) has advanced rapidly in recent years, so automated analysis of medical images has become an option for addressing the challenges faced by eye care providers. In particular, deep learning models such as convolutional neural networks and transformer models have demonstrated superior performance on complex visual data and are capable of learning useful information from retinal images automatically. Therefore, AI-based systems can help healthcare providers reduce the workload associated with diagnosing eye diseases and increase the speed at which those diagnoses are made. The interpretability of medical AI continues to be an important factor in the development of medical AI.

Many systems on the market only detect a single type of eye disease and have complicated installations, which make them less useful to consumers; also, these systems typically have non-user-friendly interfaces, thus making them less usable by non-experts. Therefore, there is a need for a total system that detects multiple types of eye diseases with ease and accessibility to all users. This paper describes an artificial intelligence (AI) based classification system that uses images of the retina taken using retinal fundus cameras to classify common eye diseases. The system uses a Vision Transformer (ViT) model to classify the images as cataracts, diabetic retinopathy, glaucoma, and normal. Different preprocessing methods are applied to improve the quality of the data and the performance of the model. In addition, a web-based user interface allows users to easily upload retinal images and receive predictions on their level of probability (e.g., whether or not they have retinal pathology) and the results. The proposed system offers a fast, accurate, and user-friendly way to detect the presence of eye diseases at an early stage.

2. Literature Review

As the number of people with visual impairments continues to rise, the demand for effective scalable diagnostic systems is increasing as well. According to the World Health Organization, millions worldwide are affected by various types of eye disorders including diabetic retinopathy, glaucoma and cataracts; these disorders are often preventable or treatable if detected at an early stage of development. The use of traditional diagnostic methodologies relies heavily on manual contact-based assessments conducted by trained ophthalmologists. These assessments are time-intensive to conduct, often inaccessible in regions of the globe where trained ophthalmologists do not practice.

Recently, there has been much progress made in the use of deep learning to analyze medical imaging data for diagnostic purposes. The two papers published by LeCun et al. and Schmidhuber provided the basis for much of the ongoing research on neural networks and hierarchical feature learning, as well their application in AI-based systems [2][3]. A further important milestone occurred with the development of convolutional neural networks (CNNs) as highlighted by AlexNet, which dramatically enhanced image classification accuracy compared to previous methods [4]. There have also been substantial advances made with the development of large image classification datasets (e.g., ImageNet) which have helped advance current models and speed up research within the area of computer vision [5].

Deep learning methods have yield strong results via Medical Imaging Technologies. Litjens et al laid the groundwork for deep learning over conventional image processing showing that deep learning techniques are superior to previous techniques for extracting complex features within medical images [6]. Gulshan et al created a system to identify diabetic retinopathy and achieved the same level of accuracy for detection as a highly trained Opthalmologist [7]. Pratt et al., similarly, have created systems that use CNN models for classifying patient images of retinal disease and have provided successful evidence for the detection of diseases in these patients [8].

Several researchers have focused their research efforts on Diabetic Glaucoma. For instance, Isha et al used deep learning techniques for identifying Glaucoma and outlined the potential for Automated Screening Systems [9]; The presence of datasets such as IDRiD has created capabilities for building and validating models that would otherwise not have existed [10]; the use of vessel segmentation techniques within retinal images may also assist in increasing the ability to extract features from images in these patients, creating greater classification ability [11].

Transformer Models created by Vaswani et al have also dramatically improved image classification in a variety of applications, including medical imaging. The use of attention mechanisms in transformer Models extend the ability of the model to establish relationships across the entire image and have provided significantly improved results for image classification, including medical images [14].

Machine Learning and scientific computing can be performed using Scikit-learn and SciPy which offer the basic foundation of tools to efficiently process data and analyse it [15],[16]. Deep Learning frameworks (i.e., PyTorch and TensorFlow) enable scalable development and training of models [17],[9]. Grad-CAM is used to enhance the interpretability of models by visualising which areas of an image are most important in making a prediction [18]. Datasets that can be accessed via Kaggle are also available for training and evaluating models for classifying retinal diseases [20].

Building upon these contributions, the proposed system develops an AI-based retinal image classification model using a Vision Transformer architecture. The system has an web interface so that user can upload retinal eye(s) images and then get back an analysis from a doctor on whether user have eye diseases such as a cataract, diabetic retinopathy, or glaucoma so that user can find out if any of these conditions are present in their early stages and get help from the health system with treatment.

3. Problem Statement

Detecting eye diseases like cataracts, glaucoma, and diabetic retinopathy is a key step in preventing blindness and providing prompt treatment. Currently, the most tests done by ophthalmologists involve visually inspecting retinal fundus photographs. This process is time-consuming and requires specialized skills, making it hard to accommodate many patients over time. Furthermore, most of the automated testing available is either designed to detect only one disease process or requires a complicated setup that is not user-friendly for either the general public or healthcare providers. Thus, there needs to be a system that is both smart and easy to use for automatically analysing retinal photographs, accurately detecting multiple diseases, and providing an easy-to-use web browser interface for both at-home users and healthcare professionals to assist with early diagnosis and improved eye health.

4. Proposed Methodology

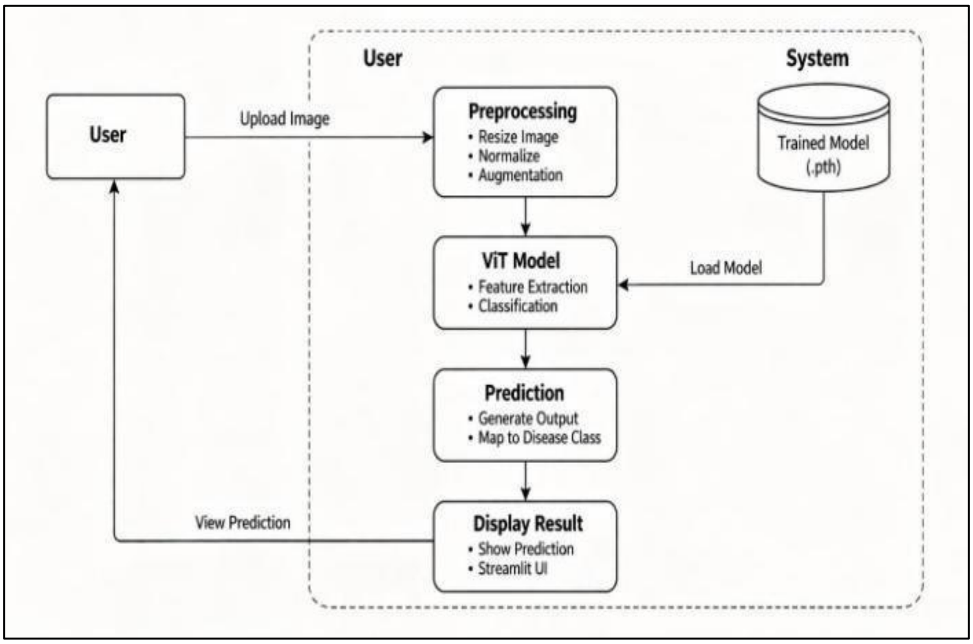

Figure 1. System Architecture

Figure 1 is the system architecture of the Retinal Image Classification. It shows the AI-based retinal imaging system works and the user interacts with the different parts of the system. The first step in this process is for a user to upload a retinal image using the system's user interface. The user will send the image to the preprocessing stage. During preprocessing, the image will undergo three main steps that will standardize it for the deep learning model: 1.) resizing, 2.) normalizing, and 3.) augmenting the measures in the image. This makes sure both the model can learn from it and the model has a better chance at being accurate and reliable.

Once the image is preprocessed, it is given to the Vision Transformer (ViT) model. The system will first retrieve a pre-trained model file (with the file extension of. pth) that contains learned weights from previous training sessions. The ViT model will perform two major operations on the processed retinal image. It extracts features from the image using attention mechanisms with respect to spatial context and will classify the retinal image based upon its analysis of global relationships in the retinal image. This classification allows the system to recognize complex relationships associated with multiple eye diseases.

After executing the two major functions of feature extraction and classification, the ViT system generates a list of output probabilities for each disease class associated with the processed retinal image. The probabilities will be associated with categories of disease such as cataracts, glaucoma, diabetic retinopathy, and normal. The category with the highest probability will be used as the output prediction. The user will be able to visualize the predictions, including related data, via a Streamlit user interface once the image classification system has completed processing the image. The entire workflow for diagnosing any of the four diseases is fast, automated, and user-friendly.

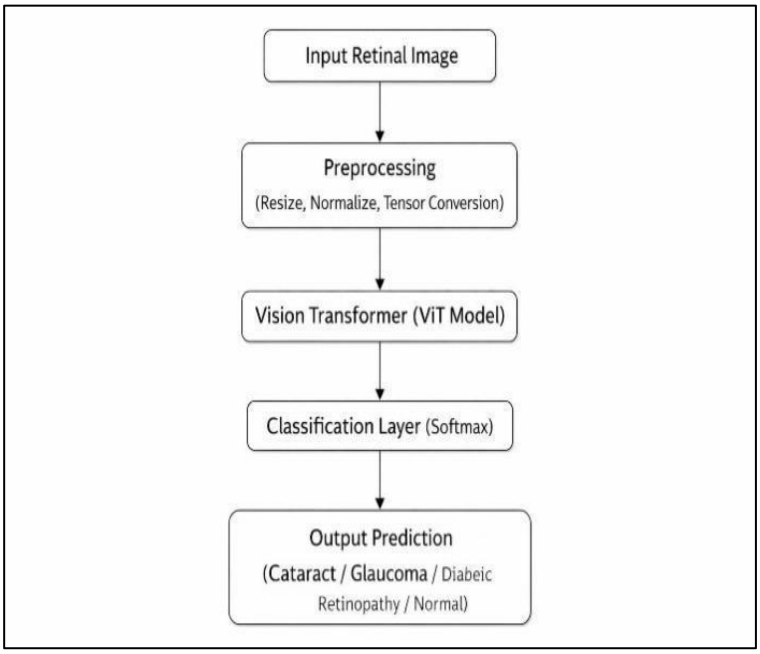

Figure 2. Data Flow Diagram

Figure 2 shows flow diagram of the Retinal Image Classification. An automated retinal image classification pipeline utilizing a deep learning approach is depicted in the diagram below. The pipeline initiates with an input retinal image obtained from a standard fundus camera. The image proceeds to undergo preprocessing, including resizing the image to a specified resolution (e.g. 224 x 224 pixels), normalizing the pixel values to a common scale, and converting the image into a tensor format compatible with the deep learning model. By standardizing images in this fashion, the resulting dataset exhibits greater integrity and facilitates improved performance of the classification model.

The Vision Transformer (ViT) model will classify the pre-processed image from the preprocessing step. The ViT does not rely on local features of the image as in CNNs but instead uses an attention mechanism to model the relationships between all of the pixels of the input image; therefore, the ViT has a better ability to extract and represent features of the complex retinal images than deep learning. Such features can aid in identifying subtle differences among retinal images resulting from ophthalmic diseases (e.g. cataracts, glaucoma, and diabetic retinopathy). Following feature extraction, the features are forwarded to a classification layer (often utilizing a softmax function), which maps the outputs of the ViT model to a probability score for each of the underlying classes in the classification system. The class associated with the highest probability then becomes the model's prediction.

The overall output of the deep-learning-based automated image classification pipeline is an identified condition (cataracts, glaucoma, diabetic retinopathy, or no condition). Therefore, every image can be examined quickly and accurately via an automated pipeline, thereby eliminating the need for examination by human observer(s) and enabling healthcare personnel to identify ophthalmic disorders at the earliest possible date.

5. Results and Discussion

5.1 Simulation Setup

The proposed system is developed using Python for building the deep learning model and performing retinal image classification. The model is implemented using the PyTorch framework along with the timm library, which provides the Vision Transformer architecture used for detecting eye diseases. Image preprocessing techniques such as resizing, normalization, and augmentation are applied using Torch vision to prepare retinal images for model training and testing. Streamlit can create a user interface that enables users to upload their ocular image(s) and see the prediction results. Also used with Streamlit are the libraries NumPy, Matplotlib, Seaborn, and Scikit-learn for data analysis and performance assessment. The combination of deep learning and web technology results in developing an efficient, user-friendly platform for automated detection and analysis of ocular disease.

The dataset for retinal images used here is from Kaggle [20]. The dataset was divided into two subsets for training (80% of the data) and validating (20% of the data), thus enabling the training of the model and the performance evaluation of the model. At the pre-processing stage, multiple pre-processing procedures were carried out to increase the level of generalizability and robustness before passing the images to the model. Foremost, the pixel intensity values were normalized by applying the mean and standard deviation values; this ensures the inputs are consistent and will provide accurate representations to enable efficient learning. The normalization process leads to stability of the training process, as it allows the model to converge faster than it would without normalization. Secondly, data augmentation procedures were used to provide increased variety in the dataset and lower the probability of overfitting. Examples of data augmentation techniques used include horizontally flipping the image randomly, rotating the image slightly (+/- 10°), and changing the brightness of the image; each simulates the variations seen in real life retinal images, which help generalize the prediction probability of the model to new data. The dataset contains four classes representing common eye conditions: cataract, diabetic retinopathy, glaucoma, and normal. Model training was performed using a GPU environment wherever available to accelerate computations.

5.2 Model Training

The Vision Transformer (ViT–small patch16 224) architecture was employed for classification. Table 1 shows the simulation parameter and their values.

Table 1. Simulation Parameter Table

Parameter	Value
Loss Function	CrossEntropyLoss (Label Smoothing = 0.1)
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epochs	20

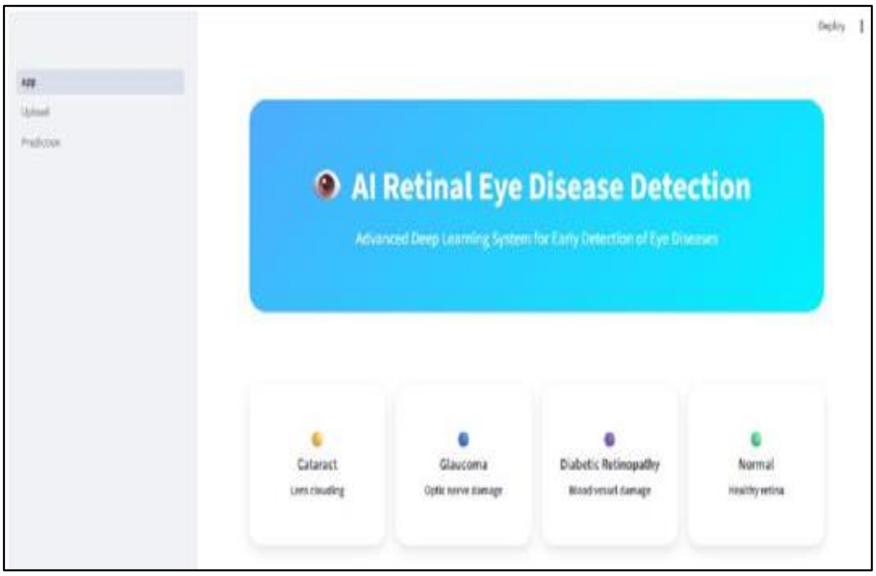

Figure 3. Homepage

The input page allows users to upload retinal images for analysis shown in Figure 3. The uploaded image is displayed for confirmation before processing is shown in Figure 4.

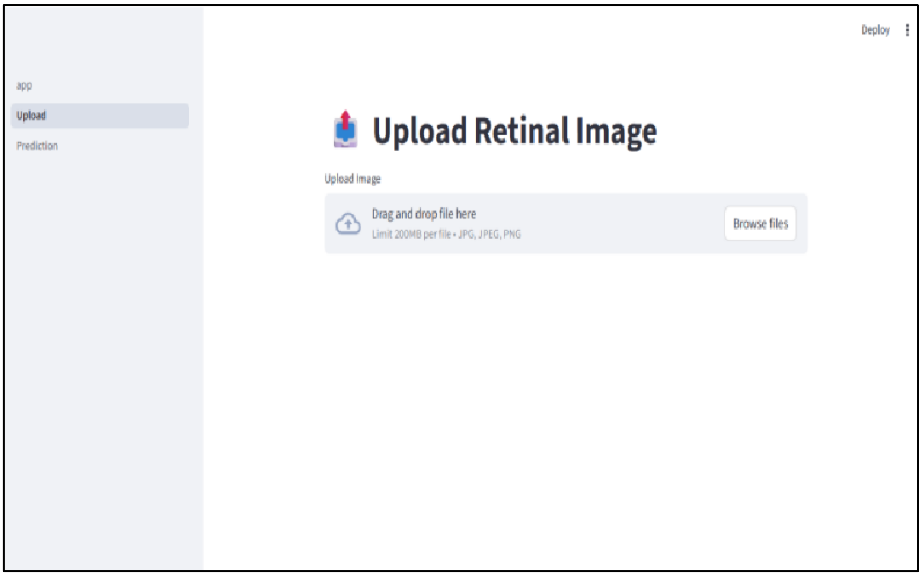

Figure 4. Upload Page

Figure 5 shows the prediction result page displays the detected eye disease along with the confidence score generated by the trained model.

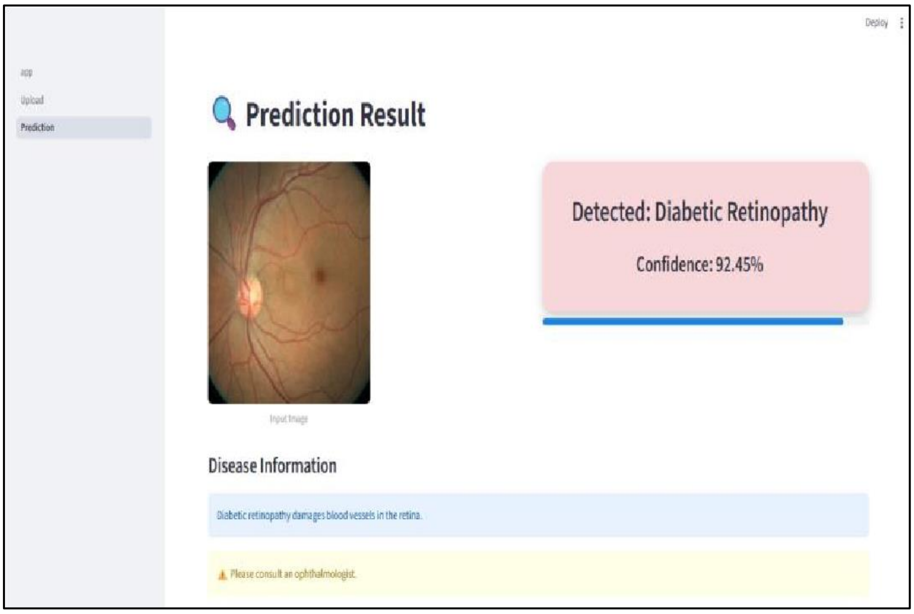

Figure 5. Prediction Page

A bar chart Figure 6, shows that each category of the predictions for eye diseases contains probabilities with each having a very high prediction probability for one category. This means that there is one category for which the final prediction can be made with high confidence in comparison to all other categories. A probability chart is generated to visualize the prediction scores for each disease category. The homepage of the AI-based Retinal Eye Disease users understands the model output clearly. Detection system presents an overview of the application along with information about different eye diseases.

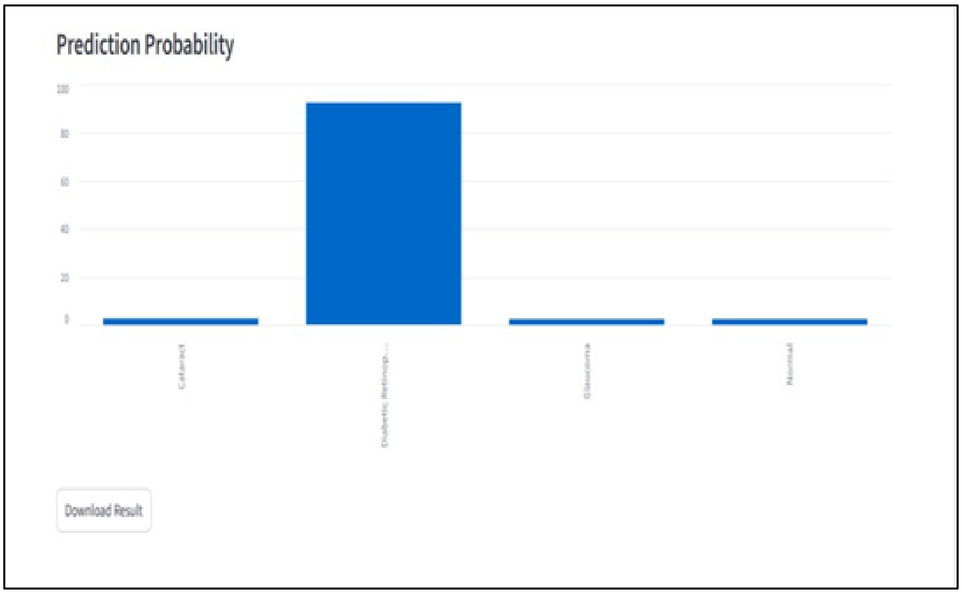

Figure 6. Probability Chart

The model setup allowed the training of the model on distinguishing characteristics across four separate categories of illness while achieving a maximum balance between computational efficiency and accuracy, which is reflected in Table 2.

Table 2. Training and Validation Accuracy at Different Epochs

Epoch	Train Accuracy (%)	Validation Accuracy (%)
1	80.29	85.48
5	95.59	88.74
9	98.11	94.22
13	98.37	93.48
18	98.81	93.78
20	98.74	94.22

At Epoch 13, the best performance from the saved model was recorded, which resulted in a Validation Accuracy of 93.48%.

5.3 Model Testing

The trained model evaluation resulted in a score of 93.48% on the validation dataset.

Classification Report:

Table 3. Classification Report of the Proposed Model (Precision, Recall, and F1-Score for Each Class)

Class	Precision	Recall	F1score
Cataract	0.95	0.93	0.94
Diabetic Retinopathy	0.99	1	1
Glaucoma	0.89	0.91	0.9
Normal	0.91	0.89	0.9

Table 3 shows the model's classification performance for precision, recall and F1-score for each of the disease categories.

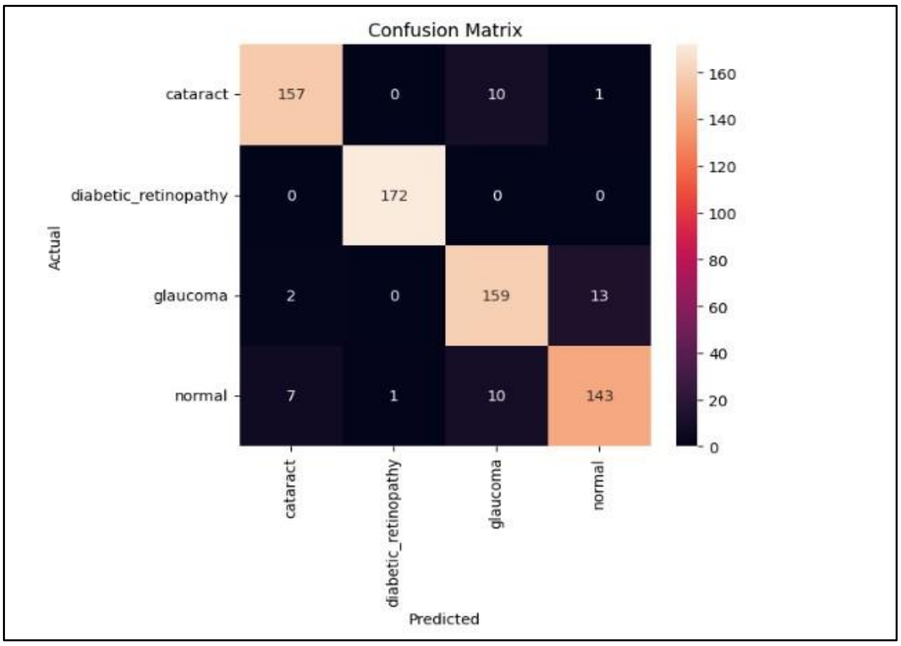

Figure 7. Confusion Matrix

The Confusion Matrix depicted in Figure 7 shows strong Classification performance (Accuracy, Precision & Recall) across all Class categories. The majority of the misclassifications observed are between Diabetic Retinopathy and Glaucoma (due to similar retinal characteristics).

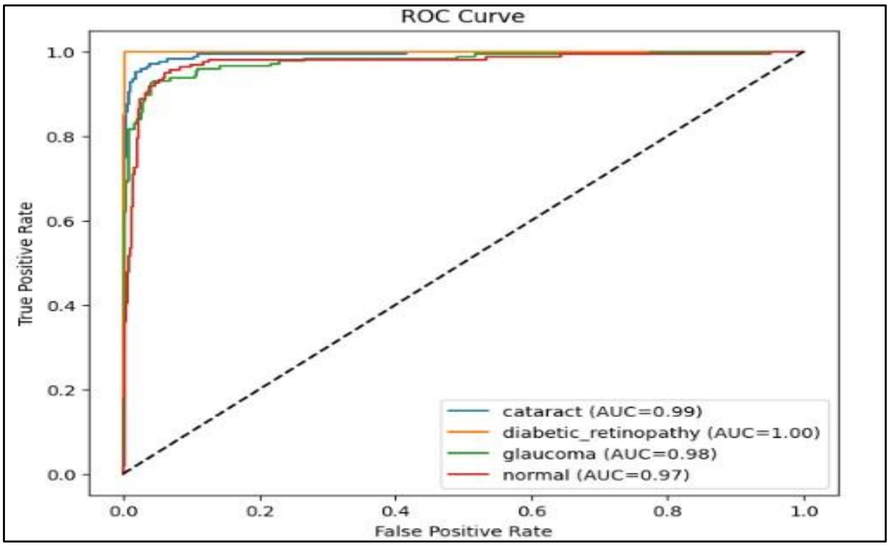

Figure 8. ROC Curve

Figure 8 shows the ROC curves demonstrates the separation between different types of disease categories for all five types of diseases, as indicated by the relatively high AUC values for each of the five diseases. A test accuracy of 93.48% demonstrates that this model has the ability to analyze and classify retinal fundus images correctly (at least) when analyzing previously unseen examples of these types of images. The model training included the implementation of a process to provide a wider range of training images through random flipping and rotating to improve overall performance, as the lighting and orientation of the images. Nevertheless, there were several misclassifications between the two classes (diabetic retinopathy as well as glaucoma) due to their similar appearance/differences as seen via fundus images. The application was designed with usability in mind through a Streamlit web application interface that allows users to upload images and receive predictions and associated confidence scores with the results. On the practicality side, a GPU was required during training for optimal model development, this resource may not always be available in low resource settings. Furthermore, the performance of the model may vary from the specified accuracy and if it is evaluated with new datasets whose image quality, acquisition methods and lighting levels differ from those of the development dataset.

The proposed system demonstrates effective performance in retinal image classification by accurately identifying multiple eye diseases using a Vision Transformer-based approach. This model opposed to conventional single disease identification systems, provides an automated solution for multi-class classification. We made it easy for users to generate predictions of retinal photographs through integration of a web interface based on Streamlit.

While this model is highly accurate, the prediction accuracy will change with the introduction of different datasets that have different imaging characteristics. Improvements in the future will include the addition of numerous and more diverse data sets; optimizing for low resource environments and deploying it through scalable cloud-based solutions to make use of this model in real world environments.

6. Conclusion and Future Scope

This study proposes an artificial intelligence (AI)-based approach for the early identification of common eye disorders based on retinal fundus images. The new AI-based technique employs a Vision Transformer (ViT) model to effectively categorize fundus images into four groups: cataract, diabetic retinopathy, glaucoma, and normal. The utilization of various preprocessing techniques (i.e., resizing, normalization, and data augmentation) significantly enhances the model's performance. Additionally, with an impressive classification rate of 93.48%, the AI-based solution demonstrates its ability to reliably and automatically detect prevalent eye diseases. Along with providing excellent predictive performance, the creation of a simple-to-use web application via Streamilt enhances the usability of the AI-assisted imaging system. Therefore, the new AI solution provides a scalable, cost-effective, and accessible automated imaging solution for early detection of coursing eye disorders that can help prevent blindness due to eye diseases. Although the current results obtained are encouraging, the model's performance may differ based on the source of the input data used (e.g., different dataset(s) or imaging modalities). Future research efforts will investigate how to improve the robustness of the model through the development and/or utilization of larger and more diverse image datasets. Furthermore, the deployment of the proposed AI solution can be completed once a sufficient number of retinal images obtained from actual clinical practices are available.

References

[1] World Health Organization: WHO. (2026, February 10). Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." nature 2015, vol. 521, no. 7553: 436-444.

[3] Schmidhuber, Jürgen. "Deep Learning in Neural Networks: An overview." Neural networks 2015, vol. 61: 85-117.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." Communications of the ACM 2017, vol. 60, no. 6: 84-90.

[5] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet Large Scale Visual Recognition Challenge." International journal of computer vision 2015, vol. 115, no. 3, 211-252.

[6] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez. "A Survey on Deep Learning in Medical Image Analysis." Medical image analysis 2017, vol. 42: 60-88.

[7] Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan et al. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." Jama 2016, vol. 316, no. 22, 2402-2410.

[8] Pratt, Harry, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, and Yalin Zheng. "Convolutional Neural Networks for Diabetic Retinopathy." Procedia computer science 2016, vol. 90, 200-205.

[9] Isha, I., C, J. G. T., Anusha, A., S, A., & Harsha, H. "Glaucoma Detection Using Deep Learning and Image Processing." Interantional Journal of Scientific Research in Engineering and Management 2025, vol. 09, no. 12, 1–9.

[10] Porwal, Prasanna, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research." Data 2018, vol. 3, no. 3: 25.

[11] Budai, Attila, Rüdiger Bock, Andreas Maier, Joachim Hornegger, and Georg Michelson. "Robust Vessel Segmentation in Fundus Images." International journal of biomedical imaging 2013, vol. 2013, no. 1: 154860.

[12] Jason Brownlee, "A Gentle Introduction to Object Recognition with Deep Learning," Machine Learning Mastery 2021. [Online]. Available: https://machinelearningmastery.com/ [Accessed: Mar. 20, 2026].

[13] Albawi, Saad, Oguz Bayat, Saad Al-Azawi, and Osman N. Ucan. "Social Touch Gesture Recognition Using Convolutional Neural Network." Computational Intelligence and Neuroscience 2018, vol. 2018, no. 1: 6973103.

[14] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is All You Need." 31st Conference on Neural Information Processing Systems (NIPS 2017), 6000 – 6010.

[15] Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel et al. "Scikit-Learn: Machine Learning in Python." the Journal of machine Learning research 2011, vol. 12: 2825-2830.

[16] Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." Nature methods 2020, vol. 17, no. 3: 261-272.

[17] Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." Advances in Neural Information Processing Systems: 8024-8035.

[18] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization." In Proceedings of the IEEE international conference on computer vision 2017, 618-626.

[19] Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin et al. "TensorFlow: A System for Large-Scale Machine Learning." 12th USENIX symposium on operating systems design and implementation (OSDI 16) 2016, 265-283.

[20] Retinal Disease Classification (Dataset)- https://www.kaggle.com/datasets/andrewmvd/retinal-disease-classification.