

ClearSight: Deep Learning-Based Classification of Retinal Diseases

Cyrus Gomes, Bhakti Palghadmal, Gayatri Nadar and Prof. Sachin Narkhede

Department of Computer Engineering, Fr. Conceicao Rodrigues College of Engineering Bandra, Mumbai, Maharashtra

Abstract—Retinal diseases such as Diabetic Retinopathy (DR), Glaucoma, Cataract and Age-Related Macular Degeneration (ARMD) have been factors behind partial vision loss or even in serious cases complete blindness in our country. Often they are not caught at their infancy stages due to lack of testing, or simple oversight by the patients and doctors themselves. Early diagnoses and treatment can rectify and heal the eye before causing too much of a significant damage. However manual diagnoses can be tough as India lacks ophthalmologist per population head. We offer to solve this by creating a tool for diagnosing and reducing the workload on already encumbered health professionals. ClearSight: a system built using deep learning, using multiple classes for classification of retinal diseases using the fundus images of the eye. Using preprocessing and training three models: ResNet, EfficientNet, InceptionV3 on robust, labeled dataset by healthcare professionals. Combined with ensemble to give accurate results taking the mean probability for the classes. We focus on three classes in this paper i.e Normal, Diabetic Retinopathy, Cataracts, Glaucoma, Macular Degeneration The system also incorporates understandable AI methods - particularly Grad-CAM - to offer clear visual explanations that emphasize the retinal areas affecting model predictions. A web interface built on Streamlit enables users to load fundus images and obtain instant disease identification accompanied by visual heatmaps and assurance scores. Tests show that the combined model reaches higher precision and clarity than single-model standard comparisons. Hence, ClearSight provides a dependable, straightforward and easily reachable AI-supported tool for screening retinal diseases - especially useful in areas with scarce ophthalmic knowledge.

Index Terms—Exoplanets, Habitability, NASA Dataset, Habitable Zone, Atmospheric Composition, Gravitational Effects, Planetary Dynamics, Space Exploration, Three.js, Extraterrestrial Life

I. INTRODUCTION

Traditional techniques for diagnosing retinal conditions either depend on rudimentary machine learning models or depend on the evaluation and inspection of their respective ophthalmologists. The machine learning models use specific designed properties of an image, such as texture analysis and vascular segmentation. The process can often be subjective, slow, and limited to diagnosing a single condition, such as diabetic retinopathy. These approaches also suffer from issues of class imbalance, a lack of generalization ability, and decipherability. On the other hand, the proposed ClearSight system recognizes unique features from retinal fundus images, depending on deep learning ensemble of three CNN architectures: ResNet18, EfficientNet-B0, and InceptionV3. The model is trained to make classification of multiple diseases. The model is also pre-processed with Discrete Wavelet Transform (DWT) to facilitate the capture of small texture features. To back up its prediction in a clinical environment, ClearSight uses explainable AI (XAI) methods that includes Grad-CAM to provide visual explanations for its predictions. When deployed through a Streamlit web platform, ClearSight also provides straightforward, and interpretable disease detection that exceeds traditional methods. On a global scale, retinal disorders such as age-related macular degeneration, cataracts, glaucoma, and diabetic retinopathy put a major burden on sight. The insufficiency of trained ophthalmologists, along with the disadvantages of manual detection methods, continues to make it difficult to arrive at early diagnosis. Current AI solutions have many limitations and generally, narrow on one specific disease or utilize narrow and inconsistent datasets, which can undermine practical reliability. In addition, the vast majority lack model transparency, making it difficult to understand or

trust their predictions in a clinical environment. ClearSight was founded to develop a holistic, varied structure for analysing retinal diseases, with the aim of bolstering model reliability and understanding, in combination with improving diagnostic accuracy through ensemble deep learning. In this way, it aims to both bridge the gap between automated image detection and reliable, clinical decision-making, and to advance professionally-supported AI solutions in eye-health.

A. Applications of ClearSight

1. Clinical Screening and Diagnosis: For the targeted eye diseases, ClearSight offers ophthalmologists instant diagnoses. In clinical environments, it acts as a decision support system.
2. Teleophthalmology and Remote Healthcare: In isolated or resource-poor areas, without access to ocular specialists, ClearSight provides AI-assisted ocular assessment. Providers can upload a fundus photo and within minutes receive diagnoses enabling early detection.
3. Mass Screening and Preventive Care: A technology capable of providing widespread diagnosis of retinal diseases in hospitals, diabetes clinics, and community health camps can facilitate identification of at-risk patients even before vision loss occurs.
4. Medical Education and Training: ClearSight, an explainable artificial intelligence feature such as Grad-CAM, can help educate medical students and residents in ophthalmology. Students can visualize retinal patterns that are characteristic of a specific condition, as well as see the reasoning for the model reasoning after the label is generated.
5. Research and Dataset Benchmarking: Researchers can use the ClearSight framework to evaluate new models, analyze feature extraction methods, or investigate how explainable AI based medical imaging tools can be explained.
6. Integrating into existing Healthcare systems: This application can be integrated into proper medical practice as a diagnostician tool to help the healthcare provider and connect to their EHR i.e Electronic Health Record for further monitoring and with the consent of the patient be further used for model training.

II. LITERATURE REVIEW

Recent advances in deep learning, particularly in the automated detection of cataracts, glaucoma, age-related macular degeneration (AMD), and diabetic retinopathy (DR), have transformed the diagnostic landscape for retinal diseases. While conventional inspection

processes are typically accurate, they are often labor-intensive, subjective, and largely dependent on visual ability. In recent research, convolutional neural networks (CNNs), wavelet-based preprocessing methods, and optimization algorithms for accurate disease classification have been increasingly employed to address these shortcomings.

This section presents three relevant papers Asia et al. (2022), Thanikachalam et al. (2024), and Vardhan et al. (2024)—that collectively illustrate the progression of computer-vision-based eye image analysis using CNNs and provide the methodological background for ClearSight.

A. Foundational Papers for ClearSight

- 1) *Asia et al. (2022) – Detection of Diabetic Retinopathy Using CNN Classification Models*: Asia et al. (2022) focused on using CNN architectures—VGG16, ResNet50, and ResNet101—to detect diabetic retinopathy. They utilized the Xiangya No. 2 Hospital (XHO) dataset of 1,607 fundus photos and additional datasets (HRF, STARE, DIARETDB0, MESSIDOR) for validation. To handle issues like limited dataset size, class imbalance, and image quality variations, preprocessing steps such as green channel extraction, histogram equalization, and noise removal were applied. Data augmentation techniques (rotation, flipping, noise injection, scaling) increased the training set by 147. However, this study was limited to binary classification (symptomatic vs. asymptomatic) and did not address multi-class severity ratings. Despite this, it remains a strong example of CNN generalization for DR detection under varying imaging conditions.

- 2) *Thanikachalam et al. (2024) – Optimized Deep CNN for DR and DME Detection*: Thanikachalam et al. (2024) introduced an improved deep CNN pipeline based on MESSIDOR dataset for the classification of DME and DR. The developed system used Discrete Wavelet Transform (DWT) for noise suppression and restoration which even performed better than the conventional filters on PSNR (29.67 dB) and SSIM (0.965). They employed Random Forest (RF) features selection, Adaptive Gabor Filters (AGF) for lesion extraction and Artificial Neural Networks (ANNs) for segmentation. Additionally, the Chicken Swarm Algorithm (CSA) optimized CNN weights, balancing exploration and exploitation, improving convergence. The CSA-enhanced CNN achieved accuracy of 97.91%. The main contributions include

biologically-inspired optimization and integration of traditional image processing with deep learning. Limitations include evaluation on a single dataset, no real-time testing, and lack of explainability methods.

3) *Vardhan et al. (2024) – Eye Disease Detection Using Transfer Learning:* Vardhan et al. (2024) leveraged transfer learning to generalize retinal diagnosis across DR, glaucoma, cataracts, and healthy eyes using a Kaggle composite dataset of 4,217 images. Preprocessing steps encompassed ImageNet normalization and data augmentation strategies (flipping, rotation, brightness adjustments) to improve model robustness. The study evaluated VGG19, InceptionV3, and ResNet50 architectures. Regularization techniques like batch normalization and dropout were applied. ResNet50 achieved a training accuracy of 99.94. Although transfer learning substantially improved performance, discrepancies in ROC-AUC values were observed.

B. Other Reference Papers

Recent work has progressed retinal disease detection by means of transformers, GANs, and ensemble learning. ResNetGAN systems boosted sensitivity and specificity by combining characteristic extraction and noise reduction. The alignment with the explanation-sensitive AI approach Grad-CAM and other XAI strategies promote clinical interpretability and thereby boost adoption.

Multi-label classification, imbalanced data, and effective models, e.g., EfficientNet, were taken up in additional investigation to help large-scale deployment. Transformer-based models support modality-spanning diagnostics, and hybrid and ensemble learning prevailed apart singularity for accuracy and robustness

To conclude, three trends are noted from these studies sophisticated preprocessing (e.g., DWT, Gabor filtering), ensemble and hybrid architectures, and explainable AI approaches. ClearSight extends these ideas by integrating DWT preconditioner, CNN ensemble fusion and Grad-CAM heatmap to create a robust, interpretable, and clinically relevant tool for general retinal disease screening.

III. METHODOLOGY

This project focus on using deep learning models to develop detection of disease in the eye using the fundus images, proving a simple and precise diagnosis. The system framework is built by creating a friendly

online platform for submitting images, assessment of prediction results, and utilizing Streamlit, which serves as the basis for the application. The relevant frameworks involve three key modules which engage the following activities. First, data preparation, collecting experimental data based on the use of deep learning algorithms to develop an eye disease classifier using TensorFlow or Pytorch framework on the Google Colab environment. The second represents a multi-class disease classifier based on adequate training. The third and last topic discusses the ability for the classifications to be evaluated in real time practice. This system framework serves to assist the health professional in the diagnosis and treatment of eye diseases and to assist the early identification and diagnosis of ocular diseases. Technology Stack: In constructing the system for identifying eye disorders, we used Python, TensorFlow and PyTorch for deep learning. This was ascertained and trained in Google Colab. Streamlit was employed to make disease predictions and visualizations interactive on a web application to provide value to end-users. Other libraries such as NumPy, Pandas and OpenCV were also very beneficial in facilitating image analysis and preprocessing. Expected Outcome: A deep learning platform can utilize images of the retinal fundus to identify and categorize eye diseases. It is thought to detect very many diseases accurately and reliably. This method will be implemented in an interactive web application called Streamlit.

A. System Overview

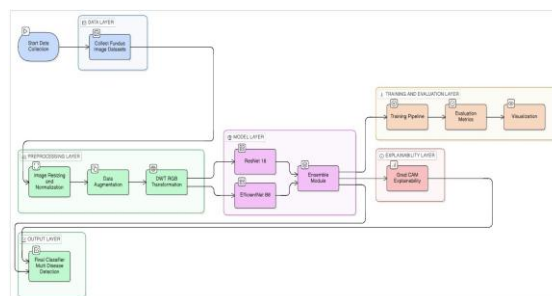


Fig. 1. Workflow Diagram

To improve the diagnosing the diseases and added interpretability, we have created a end-to-end workflow for analyzing the fundus images present. First the image is uploaded and directed to the pre processing pipeline where it is resized and restructured to the normal conventions. Multiple techniques are applied such a DWT(Discrete Wavelet Transform) and color conversion and then resized to a standard, uniform size.

The architecture of the system consists of various layered components that work together to facilitate high-quality disease classification through the use of fundus images. First, in the Data and Preprocessing Layer, training sets of fundus images are downloaded from resources such as Kaggle and organized into specific diseases. The fundus images are preprocessed utilizing resizing, normalization, and augmentation techniques that feed into diversity of data through rotations, flipping, noise, and zooming. The Discrete Wavelet Transform (DWT) plus RGB decomposition provide improved feature representation since they describe detailed spatial-frequency information. Second, in the Model Layer, two contemporary convolutional neural network architectures ResNet-18 and EfficientNet-B0 are used to compute outputs that are combined in the Ensemble Module to improve accuracy and robustness. Third, in the Training and Evaluation Layer, the training of the model is completed using Adam optimizer with CrossEntropy loss and using metrics including accuracy, precision, recall, and F1 score and visualizing with a confusion matrix. Fourth, the Explainability Layer utilizes Grad-CAM to identify the areas of individual images that most heavily influenced the model. Finally, the Output Layer delivers the system's diagnostic results, providing both disease classification and visual explanations that assist ophthalmologists in validating and understanding the model's predictions.

B. Dataset Acquisition and Curation

We used a combined dataset, comprising 2500 images of the fundus of the eye. We have five classes and each class has 500 images. The five classes being diabetic retinopathy, glaucoma, cataracts, age related macular degeneration and last normal. All these images were sourced from multiple data sources, through kaggle and other open public locations. Each image is pretagged to a particular class by an ophthalmologist for easier use. To ensure integrity of data, rigorous testing was done to eliminate unfit images. This was done through technical validation i.e the technical quality of the image, resolution, being up to standard and also clinical review by trained professionals.

C. Data Preprocessing and Augmentation

All images in the dataset underwent a standardized preprocessing pipeline to maintain consistency across the entire collection. The initial step involved

converting all images to PNG format, a lossless compression method that preserves image quality without degradation. Subsequently, all images were resized to a uniform resolution of 512x512 pixels for computational efficiency and ease of processing. A crucial preprocessing technique use was Discrete Wavelet Transform (DWT), which enables the particular model to better capture the finer details of the image, particularly the vascularity of the fundus image i.e the several capillaries, arteries and veins in the retinal fundus. The DWT performs channel decomposition, retaining the crucial colour information, which helps with accuracy. Along with this the standard data augmentation techniques are also used such as resizing, including rotations, zooming in, zooming out, brightness and contrast adjustments. These changes are include to increase variability as realistically the picture the users upload will always not be uniform, this also lessens overfitting in the model training, and again mimics how a user might upload less than quality pictures, which accounts into using different cameras, different lighting due to diverse conditions from when their photo was acquired. Since the dataset was balanced with 500 images per class, there wasn't a need to adjust class weights, which simplified the process of optimizing. All these data preprocessing and augmentation was used to help improve the training data.

D. Model Development

For model development we used three models to train the dataset on. These three being ResNET, EfficientNET and InceptionV3, utilizing their deep learning architectures for

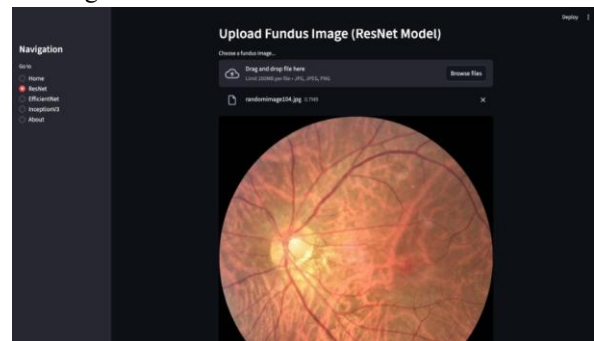


Fig. 2. Implementation through Streamlit robust analysis. ResNET was chosen for its efficiency and its reliability as a model. It creates deep feature hierarchies preventing the vanishing gradient issue faced. It is quite stable and one of the more popular

models. The key idea behind EfficientNet is compound scaling, it scales all three i.e depth, width and resolution in a balanced way. Developed by Google AI in 2019, it gives accurate results with far less computational requirements. As for InceptionV3, instead of utilizing a single filter size, it uses multiple filters in parallel such 1x1, 2x2, 3x3 and so forth. These are concatenated to display features at multiple different scales at the same time. It is very good at feature extraction for complex images such as satellite and medical images.

E. Ensemble Strategy



Fig. 4. Probabilities after ensemble

We used soft voting as our ensemble strategy, In soft voting all three models give out probability on the class, for examples supposedly for a fundus image, ResNet gives out the following probabilities, [0.8, 0.1, 0.05, 0.05, 0]. Therefore in soft voting it takes the probabilities from all models, and averages it. The class with the highest average across all the models therefore becomes the final result. This is better than traditional hard voting where it is majority rules, soft voting provides much higher accuracy. Along with soft voting we also have used prediction uncertainty metric, it shows how confident the model is about its output prediction, showing the reliability of the model. For example, if the probabilities by a model are [0.4, 0.4, 0.1, 0.1, 0], it shows that the model is unsure between the first two classes, while a higher average might give a class, it will still be not a confident prediction, so it is displayed with a much higher prediction uncertainty score.

F. Evaluation

To ensure a robust and clinically relevant assessment, the performance of all models was evaluated using a suite of complementary metrics that go beyond simple accuracy. This multi-faceted approach provides a

holistic view of each model's diagnostic strengths and weaknesses. The evaluation framework relied on three key matrices to paint a complete picture. Classification reports were generated for each model, detailing precision, recall, and F1-score on a per-disease basis. This allowed for a nuanced understanding of performance for each specific condition, such as glaucoma or diabetic retinopathy, ensuring that high overall accuracy was not achieved at the expense of reliably detecting a particular disease. Furthermore, confusion matrices were meticulously analyzed to uncover the exact nature of classification errors. These matrices served as a critical diagnostic tool, visually revealing which conditions were most frequently confused for one another and highlighting potential areas for model improvement. Finally, visual analytics including learning curves and performance heatmaps were employed. These graphs were instrumental in tracking training stability, diagnosing issues like overfitting, and providing an intuitive, high-level summary of the model's capabilities across the different diagnostic categories.

G. Model Explainability

Model explainability was one of our crucial goals during this project. We achieved this by implementing Gradient-weighted Class Activation Mapping (Grad-CAM). We considered GRAD-CAM because it generates heatmaps of the fundus image, highlighting the regions which influence the results of the model. It is very crucial in interpreting the results. The heatmaps consisted of localized anatomical focus highlighting a bursted capillary or vein, damage to the macular region, white spots etc. This explainable AI helped in giving us a short written note on the source of the issue along with the heatmap. GRAD-CAM works by looking at the penultimate layers of the model before its decision by focusing on the regions the model is particularly interested at.

H. User Interface and Deployment

All the three trained models, the final soft voting results, uncertainty prediction and the GRAD-CAM explainability were deployed using Streamlit. The models were uploaded on Hugging Face. The interface is pretty standard with easy to use and understand the user interface, with a simple how to on the landing page. Simply upload the fundus image and it will give out a report about the probabilities predicted by each model.

The final result after the ensemble soft voting and the uncertainty of the prediction. If you want an indepth exaplanability you can click on the GRAD-CAM tab to view the heatmaps about the where the model is focusing on to predict the results, along with an simple medical explanation about the location of the issue.

IV. ANALYSIS OF RESULTS

This section compares the three CNN architectures ResNet18, EfficientNet-B0, and InceptionV3 on a balanced fundus dataset from five categories. Our data demonstrate a complete performance profile, including overall accuracy and classspecific precision, recall, and F1 scores. We conclude the analysis with macro and weighted averages that capture the performance of each model.

A. Overall Accuracy & Averages

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score	Weighted F1-Score
ResNet-18	0.96	0.97	0.96	0.96	0.96
EfficientNet-B0	0.99	0.99	0.99	0.99	0.99
Inception V3	100%	100%	100%	100%	0.98

TABLE I PERFORMANCE COMPARISON OF MODELS

InceptionV3 was the best of all models, achieving a perfect score on the validation set. EfficientNet-B0 placed a close second, earning a 99ResNet-18, which had a respectable 96for the best performance on these data.

TABLE II- CLASS-WISE F1-SCORE COMPARISON ACROSS MODELS

Disease Class	ResNet-18	EfficientNet-B0	InceptionV3
Age-related Macular Degeneration	0.98	0.99	1.00
Cataract	0.97	0.99	1.00
Diabetic Retinopathy	0.98	0.99	1.00
Glaucoma	0.95	0.99	1.00
Normal	0.94	0.99	0.99

As shown in the table above, it shows the accuracy for each class as the three models. ResNet gave a safe baseline accuracy being most accurate for Diabetic Retinopathy and Age-Related Macular Degeneration. EfficientNet was all round accurate with no flaws in a particular class and Inception was found out to be the most accurate but it was slower. This was tested on our testing slit of the dataset. We further increase the accuracy by using ensemble learning - soft voting to choose the most probable class. Also results with high uncertainty were deemed inaccurate and thus considered a failure of the model.

V. CONCLUSION AND FUTURE WORK

Our research shows that a reliable system for diagnosing eye diseases is possible and also practically viable. However success depends on deliberate preparation, pre-processing and an intelligent approach to ensemble learning. The models trained by us held well against test cases, and edge cases. We saw considerable improvement in sensitivity and specificity for diabetic retinopathy and glaucoma. Also with interpretability it gave us a basic idea where the problem is, helping the clinician make final judgement. To make it accessible we have deployed it through streamlit for usage by anyone. Our goal was to create a hub for diagnosing all diseases under one tool, and we have succeeded in differentiating among five classes. Our next step is to present this to ophthalmologists and fix minor kinks in our model.

First we continue to training and testing datasets, from a wide variety of different demographic of people. Covering a wide range of demographic makes the model as accurate as possible. Also the future steps include to several more classes of diseases apart from the standard four developed here. Improving accessibility to this project and opening it for open-source so we can get contributions, and ideas from general populace. Improving the sites UI and improving the interpretability further, helping understand us the background of diagnoses.

REFERENCES

- [1] A.-O. Asia, C.-Z. Zhu, S. A. Althubiti, D. Al-Alimi, Y.-L. Xiao, P.-B. Ouyang, and M. A. A. Al-Qaness, "Detection of diabetic retinopathy in retinal fundus images using CNN classification models,".

- Electronics, vol. 11, no. 17, p. 2740, 2022, doi: 10.3390/electronics11172740
- [2] V. Thanikachalam, K. Kabilan, and S. K. Erramchetty, *Optimized deep CNN for detection and classification of diabetic retinopathy and diabetic macular edema*, BMC Medical Imaging, vol. 24, no. 227, 2024, doi: 10.1186/s12880-024-01406-1.
- [3] K. B. Vardhan, M. Nidhish, S. K. C., N. S. D., S. C. V., and R. M. Bhavadharini, *dEye disease detection using deep learning models with transfer learning techniques*, EAI Endorsed Transactions on Scalable Information Systems, vol. X, pp. 1–8, Jul. 2024, doi: 10.4108/eetsis.5971.
- [4] Bilal, A. Keles, and M. Bendeche, *Advances in disease detection through retinal imaging: A systematic review*, Computers in Biology and Medicine, vol. 194, p. 110412, 2025, doi: 10.1016/j.compbiomed.2025.110412.
- [5] G. Jeyasri and R. Karthiyayini, *Deep learning based retinal disease classification using an autoencoder and generative adversarial network*, Biomedical Signal Processing and Control, vol. 108, p. 107852, 2025, doi: 10.1016/j.bspc.2025.107852.
- [6] M. M. Aghabeigi Alooghareh, M. M. Sheikhey, A. Sahafi, H. Pirnejad, and A. Naemi, *Deep learning for comprehensive analysis of retinal fundus images: Detection of systemic and ocular conditions*, Bioengineering (Basel, Switzerland), vol. 12, no. 8, p. 840, 2025.
- [7] S. Ejaz, R. Baig, Z. Ashraf, M. M. Alnfai, M. M. Alnahari, et al., *A deep learning framework for the early detection of multi-retinal diseases*, PLOS ONE, vol. 19, no. 7, p. e0307317, 2024.
- [8] O. J. Patel, K. Verma, C. Vyas, T. N. Pandey, B. B. Dash, and S. S. Patra, *Detection of eye disease using deep learning algorithms*, in 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp. 1164–1170, doi:10.1109/ICSES63445.2024.10763329.
- [9] T. Hu, B. Yang, J. Guo, W. Zhang, H. Liu, N. Wang, and H. Li, *A fundus image classification framework for learning with noisy labels*, Computerized Medical Imaging and Graphics, vol. 108, p. 102278, 2023, doi: 10.1016/j.compmedimag.2023.102278.
- [10] D. Muhammad and M. Bendeche, *Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis*, Computational and Structural Biotechnology Journal, vol. 24, pp. 542–560, 2024, doi: 10.1016/j.csbj.2024.08.005.
- [11] L. Arora, S. K. Singh, S. Kumar, et al., *Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy*, Scientific Reports, vol. 14, p. 30554, 2024, doi: 10.1038/s41598-024-81132-4.