# Ensemble Deep Learning Approach for Cataract Disease Detection in Ocular Images

Surabhi Kharkate
*Department of Computer Engineering*
*MIT Academy of Engineering,Alandi*
Pune, India
surabhikharkate26@gmail.com

Palak Yerawar
*Department of Computer Engineering*
*MIT Academy of Engineering,Alandi*
Pune, India
palak.yerawar@mitaoe.ac.in

Sakshi Lade
*Department of Computer Engineering*
*MIT Academy of Engineering,Alandi*
Pune, India
sakshi.lade@mitaoe.ac.in

Diptee Chikmurge
*Department of Computer Engineering,Alandi*
*MIT Academy of Engineering*
Pune, India
diptee.chikmurge@mitaoe.ac.in

Sunita Barve
*Department of Computer Engineering,Alandi*
*MIT Academy of Engineering*
Pune, India
ssbarve@mitaoe.ac.in

***Abstract*—Eye conditions like cataracts are a major cause of vision loss encyclopedically, where early and precise diagnosis is pivotal for successful treatment. Conventional diagnostic techniques are generally time-consuming and dependent on expert knowledge, which may cause delay in intervention. Over the past many years, deep learning, especially Convolutional Neural Networks (CNNs), has shown excellent performance in medical image classification. This work introduces an ensemble-based deep learning method for the automatic classification of optical images into cataract and normal classes. Pretrained CNN models similar as VGG19, ResNet50, and EfficientNetB0 were utilized through transfer learning. Their predictions were combined using a stacked ensemble technique to take advantage of each model's strengths while enhancing conception. The suggested ensemble model was trained and tested on a intimately accessible ocular disease dataset with a validation accuracy of 98.17(%), better than single models. Performance metrics like accuracy, loss, and confusion matrix analysis were employed to compare model performance. These findings demonstrate the power of ensemble learning techniques in medical image processing and their promise in supporting clinical decision-making in ocular disease screening. diagnostic tool.**

***Index Terms*—Deep Learning, Convolutional Neural Network (CNN), Ensemble Learning, Transfer Learning, Ocular Disease Detection, Medical Image Classification, VGG19, ResNet50, EfficientNetB0, Stacked Ensemble**

## I. Introduction

Eye conditions, especially cataracts, represent an increasing threat to international public health, responsible for a large percentage of visual impairment and blindness—most notably among the elderly. Cataracts are individually responsible for more than half of the world's total incidence of blindness. While detection followed by treatment at an early stage can cut subsequent morbidity drastically, diagnosis in most of the world is still dependent on subjective interpretation of images by ophthalmologists. This process is frequently time-consuming, subjective, and reliant on the clinician's experience. Insufficient deployment of trained staff and healthcare facilities also worsens diagnostic delays and errors, especially in resource-poor environments. With the rising patient load and restricted access to specialist care, there is an instant need for precise, scalable, and automated diagnostic tools to aid clinical decision-making and facilitate early intervention.

The advent of Artificial Intelligence (AI), specifically Deep Learning (DL), has transformed medical image analysis and disease classification. In article [1] performed an in-depth review describing how Convolutional Neural Networks (CNNs) have changed the game by learning automatically discriminative features from image data. These models have achieved unbelievable success in use from radiology to ophthalmology. In the field of ophthalmology, Research paper [2] investigated the use of CNNs to detect diabetic retinopathy, glaucoma, and cataracts from fundus images with high clinical potential while reporting issues regarding generalization and interpretability of the model.

Transfer learning with pre-trained CNN architectures like VGG19, ResNet50, and EfficientNetB0 has even enhanced diagnostic performance in low-annotated data settings, which renders them deployable in health care. That said, relying exclusively on one architecture could be restrictive for model robustness and accuracy. Wang et al. [5] elaborated on how ensemble models and domain adaptation help counter typical problems such as data imbalance, overfitting, and poor interpretability, justifying the use of multi-model solutions.

Inspired by these results, the current research investigates a stacked ensemble architecture that amalgamates the synergies of different CNN models—VGG19, ResNet50, and EfficientNetB0—for binary ocular image classification as either ”Normal” or ”Cataract.” Transfer learning was employed to fine-tune the models, and the ensemble was developed by combining their predictions through a meta-learner for enhanced overall performance. As with Li et al. [11], who used CNN ensembles for glaucoma diagnosis, our ensemble approach attempts to improve diagnostic accuracy and generalizability but with specific attention to cataract detection.

Experimental results indicate that the ensemble approach is superior to all the individual CNN models in classification accuracy and stability. This verifies that stacking deep models enhances the validity and efficacy of ocular disease screening, opening doors for its application in real-world clinics.

The rest of this paper is structured as follows: Section II discusses related work; Section III explains the dataset and data preprocessing methods; Section IV describes model architectures and ensemble method; Section V shows results and discussion; and Section VI concludes with future research directions.

## II. LITERATURE SURVEY AND GAP

In the past few years, there has been extensive research focused on automating ocular disease detection through deep learning methods applied to medical images. Initial studies were based mostly on conventional image preprocessing and handcrafted feature extraction, followed by traditional machine learning classifiers. Although these methods were promising, their reliance on domain knowledge and poor generalizability across different datasets limited their practical utility in real-world clinical environments.

Litjens et al. [1] have given a thorough survey on the applications of deep learning to medical image analysis. The paper describes how Convolutional Neural Networks (CNNs) have transformed disease detection by learning image features automatically. It also delineates successful CNN applications across various domains such as radiology, dermatology, and ophthalmology. But the survey also points out that model interpretability, data limitations, and computational cost are issues for deployment into clinical workflows.

Ting et al. [2] examined the application of deep learning models in ophthalmology. They explained the way CNNs have been used to identify diabetic retinopathy, glaucoma, and cataracts from images of the fundus. The article touches on the clinical potential but cautions about constraints like a lack of model explainability, and difficulty with generalization across devices and demographics.

Sharma and Singh [3] suggested a deep learning-based approach to automatic cataract detection. Employing a CNN-trained model on preprocessed eye lens images, they accomplished high accuracy in classification. The work had potential for cataract screening but did not delve into ensemble methods or light-weight deployment possibilities.

Gargeya and Leng [4] introduced a deep learning method for machine-automated screening of diabetic retinopathy. They employed a CNN model trained by transfer learning and reported high diagnostic accuracy. Although their model was stable, it was specialized in diabetic retinopathy and not cataracts and lacked the ensemble learning method.

Wang et al. [5] reviewed issues in applying deep learning for medical imaging, citing data imbalance, lack of interpretability, and computational overhead as problems. The paper pushed for ensemble models and domain adaptation tuning to help mitigate these obstacles, although a specific ensemble model was not executed. This conforms to the reason behind the stacked ensemble approach adopted in the present study.

Rajalakshmi et al. [6] introduced an automated screening system for diabetic retinopathy based on a smartphone fundus camera and deep learning model. Although novel in terms of mobility, the model was centered mainly on a single disease class and did not examine model ensembles or multiple architectures, which made it less adaptable.

Chalakkal et al. [7] performed a survey of artificial intelligence-based models for the detection of ocular diseases.

In their research, they classified models by disease and model type (e.g., classification, segmentation). They cited the absence of ensemble-based and hybrid models as a void in detecting cataract and normal conditions effectively.

Sahlsten et al. [8] proposed a CNN-based system that would diagnose diabetic retinopathy from retinal images. With such great performance in real-world data, it leverages the image preprocessing for better learning. However, it is limited to one architecture and does not utilize model fusion or ensemble strategies for broader applications.

Kermany et al. [9] came up with a CNN model that was trained on a large ophthalmic dataset comprising retinal OCT and chest X-ray images. The model had superior classification accuracy through transfer learning. While it handled more than one disease, the research did not examine ensemble learning or the detection of cataract specifically.

This work [10] suggested a light CNN model for cataract diagnosis that was designed to lower the computational costs for deployment on handheld devices. While effective, the research only considered one architecture and did not compare with other deep models, making it not so highly extensible.

Li et al. [11] used a deep ensemble model for glaucoma detection from fundus images. They used an ensemble with several CNNs trained with varying hyperparameters to combat overfitting and increase generalizability. As much as the results were stunning, they only concentrated on glaucoma and did not cover binary classification like Normal vs. Cataract.

Mahapatra et al. [12] suggested an attention-guided CNN model for the classification of retinal disease. The application of attention mechanisms enhanced the model’s attention to significant areas in fundus images. Nevertheless, the increased complexity of the model diminishes its usability in lightweight mobile health applications, which this project is intended to facilitate.

Rajalakshmi et al. [13] in another study emphasized diabetic retinopathy detection with a smartphone-acquired dataset using an ensemble of deep learning models trained on a smartphone- acquired dataset. The study emphasizes the necessity of mobile- friendly systems but was hampered by small datasets and a lack of experiments on other conditions such as cataracts.

Tiwari and Tripathi proposed a hybrid model [14] integrating CNN characteristics with conventional machine learning classifiers such as SVM and Random Forest for cataract classification. The model was effective but not end-to-end trainable or scalable, rendering it less flexible compared to pure deep learning or ensemble CNN models.

Zhang et al. [15] presented a multi-scale CNN for disease classification of the eye, utilizing pyramid inputs from images to allow better spatial comprehension. While their approach enhanced feature extraction, it was computationally intensive and lacked ensemble learning, creating deployment challenges in real- time diagnostic settings.

This article [16] suggested transfer learning with EfficientNet for detecting multi-class ocular disease. Their study reflected robust performance on benchmark datasets but was based predominantly on pretrained weights without investigating ensemble combinations, reducing adaptability for binary-centric tasks like cataract vs. normal classification.

Article by Dutta [17] also developed a real-time cataract detection optimized lightweight CNN architecture. The proposed model was as accurate as competitive alternatives using fewer parameters, which is well-suited to clinical requirements. Model ensembling, which would have enhanced diagnostic robustness, was not investigated in the study.

Research Gaps: Significant gaps still exist despite advances in deep learning for the diagnosis of ocular disorders; the majority of current research uses single CNN models or focuses on conditions like glaucoma or diabetic retinopathy, with little attention paid to cataract detection through ensemble methods. Many have poor interpretability, don't investigate lightweight, mobile-friendly architectures, and aren't generalizable across devices and datasets. Clinical implementation is also difficult because there aren't many works that incorporate explainability techniques like Grad-CAM or perform comparative ensemble analyses. These drawbacks highlight the necessity of an effective, comprehensible, and broadly applicable ensemble framework designed especially for binary cataract categorization.

## III. OCULAR DISEASE

The human eye is a critical sensory organ for vision and is composed of major components such as the cornea, lens, retina, and optic nerve. The eye operates by collecting light and transforming it into electrical signals transmitted to the brain. It is susceptible to various disorders that cause partial or complete loss of vision. Cataract, due to opacification of the lens of the eye, is among the major causes of blindness globally, particularly in the elderly. Early diagnosis is important to avoid permanent damage and facilitate proper treatment.

Eye diseases are a collection of various conditions that affect different structures of the eye. Two of the most common causes of visual disability in the world are cataract and glaucoma. Cataract is clouding of the natural lens of the eye and is responsible for more than 51% of global blindness, as estimated by the World Health Organization.

The increasing amount of retinal and ocular image databases alongside the development in deep learning algorithms has set the stage for high-accuracy automated detection of these disorders, enhancing early diagnosis and assisting ophthalmologists in clinical decision-making.

## IV. METHODOLOGY

### A. Dataset Used

The dataset used here is the Ocular Disease Recognition (ODIR-5K) dataset, released publicly on Kaggle . It has 6,392 fundus images gathered from varied patient populations, labeled by skilled ophthalmologists in eight classes, namely Normal, Cataract, Glaucoma, Diabetic Retinopathy, and others. The diversity of the dataset with respect to disease categories, image quality, and acquisition settings reflects a real-world challenge to automated diagnosis systems. Preprocessing techniques like resizing, normalization, and data augmentation were used to increase model robustness and generalizability. Having this extensive dataset available allows the creation and assessment of deep learning models for ocular disease binary-class classification, facilitating progress toward clinical utility.

TABLE I: Distribution of Cataract and Normal Images in the Dataset

| Class | Percentage of Images |
|---|---|
| Cataract | 35.1% |
| Normal | 64.9% |

**For preprocessing:**

- The original fundus images are re-sized to 224×224 pixels to accommodate the input specification of typical CNN architectures such as VGG19, ResNet50, and EfficientNetB0.
- Images are transformed to RGB mode and pixel values are normalized to the range [0,1].
- Class imbalance is overcome, and the model generalization is improved using data augmentation which involves random rotation, brightness changes, zoom, and horizontal flipping.
- Image pairs (left and right eye) are processed either separately or combined in late fusion (at ensemble prediction time).

**Label Encoding and Dataset Splitting:**

Every patient record in the ODIR-5K dataset is labeled with one or more disease labels according to expert diagnosis. As this is a multi-label classification task (a patient may have several diseases at once), binary encoding is applied to the target labels. Each of the eight disease classes is encoded as a distinct binary value (1 or 0), denoting the presence or absence of the condition. To prepare the dataset- Labels are converted to multi-hot encoded vectors for use with a sigmoid-activated output layer, appropriate for multi-label classification. The data is randomly divided into training, validation, and test sets in a stratified fashion to preserve label distribution between sets:

- 70% for training
- 15% for validation
- 15% for testing

Efforts are made to ensure that left and right eye images of the same patient do not show up in different subsets, preventing data leakage.

To develop a high-performing and robust model for ocular disease classification, this study employs transfer learning by leveraging pretrained CNN architectures. The models used include:

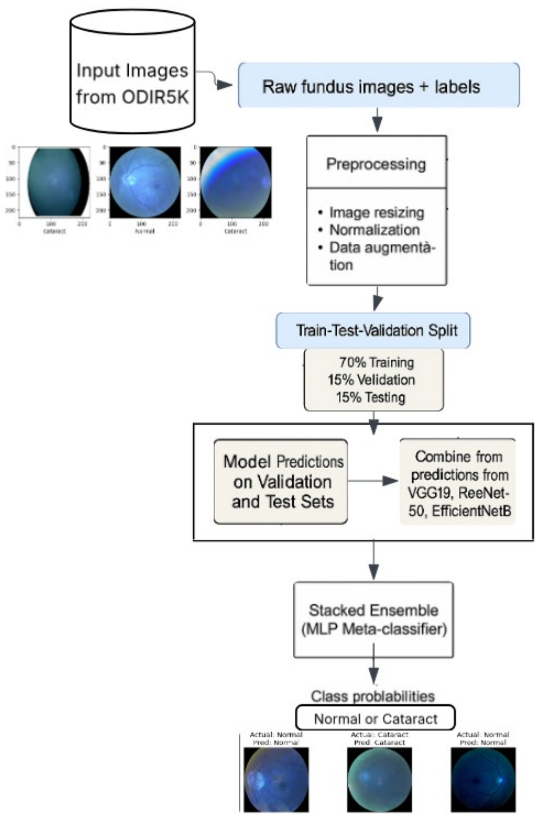


Fig. 1: Architecture Diagram.

**1. VGG19:** The deep convolutional neural network VGG19 are its depth and simplicity. It features a consistent architecture throughout its 19 layers, most of which have tiny 3x3 convolution kernels. To perform multi-label classification of ocular diseases, the pretrained VGG19 model was altered in the current study by deleting its top layers and adding a global average pooling, dropout, and dense output layer with sigmoid activation. The model's hierarchical feature learning in particular allowed it to extract features from the ODIR dataset with high reliability. This model's convolution operation can be expressed mathematically as (1). Nevertheless, VGG19 required more memory and had a slower training time than more recent models due to its large number of parameters.

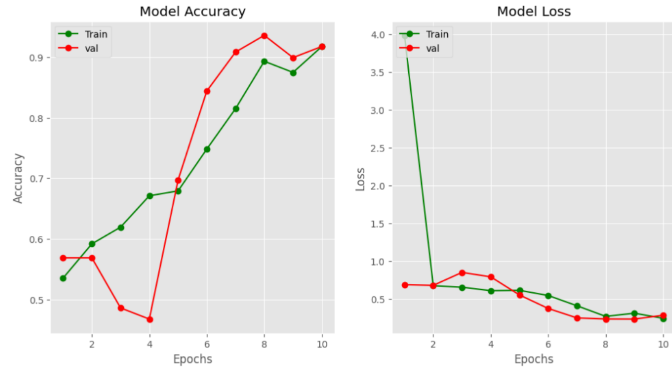


Fig. 2: VGG19 Model Accuracy and Loss Graph

**2. ResNet50:** ResNet50 brings about the idea of residual learning through identity shortcut connections that assist in reducing the vanishing gradient issue. With 50 layers, ResNet50 enables deeper structures without performance degradation. In this paper, ResNet50 was used with a custom classification head for multi-label classification. The residual blocks in the model provided more effective gradient propagation and accelerated convergence. It performed superior generalization on validation data than VGG19 and was less vulnerable to overfitting because of its internal batch normalization and skip connections. ResNet50 struck a balance between depth and computational efficiency and thus represented a robust baseline for comparison. After the fully connected layer sigmoid activation function is applied(2).

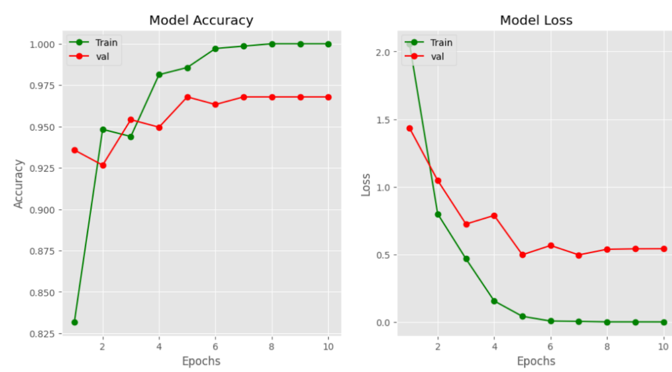


Fig. 3: ResNet50 Model Accuracy and Loss Graph

**3. EfficientNetB0:** An effective and scalable architecture, EfficientNetB0 trades network depth, width, and resolution by scaling a network by a compound scaling factor. These three dimensions' scaling can be represented as shown in (1). With less computation, it has high accuracy and is highly parameter-efficient. In this project, EfficientNetB0 was modified for multi-label classification of ocular diseases by adding a global average pooling layer, dropout, and sigmoid-activated output (see Equation (2)). Among the tested models, EfficientNetB0 had the fewest parameters and trained the fastest, making it perfect for low-resource or real-time applications. Because of its optimized architecture, it was also able to match competitive performance, sometimes surpassing deeper models.

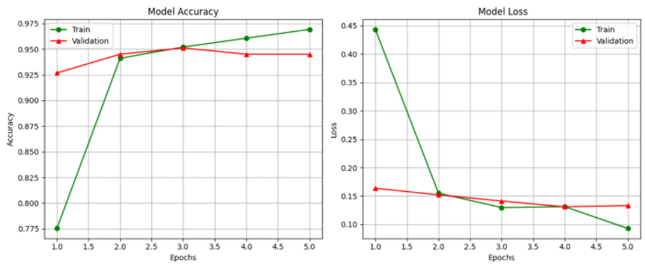


Fig. 4: EfficientNetB0 Model Accuracy and Loss Graph

### *B. Model Compilation and Training Configuration*

All the models—VGG19, ResNet50, EfficientNetB0, and the Stacked Ensemble—were trained and compiled with the same settings to enable an objective and equitable comparison. The base models were fine-tuned on the ODIR dataset and provided softmax probability outputs on the test set, calculated by Equation (3). Adam's adaptive learning capability and high performance on deep neural networks made it the preferred optimizer. Multi-label classification was the problem in this case, where an image can be from multiple disease categories simultaneously. So, instead of categorical cross-entropy, binary cross-entropy (Equation (5)) loss function was employed. For faster convergence and prevention of overfitting, callbacks such as ModelCheckpoint and EarlyStopping (patience = 10) were employed. Training was conducted for a total of 10 epochs with a batch size of 32 and input image size normalized to 224×224 pixels.

These output vectors were horizontally concatenated to create a composite feature representation for every sample. A Multilayer Perceptron (MLP) classifier was subsequently trained as a meta-learner on the stacked outputs of the validation set and applied to predict labels for the stacked test data as shown in Figure 1. Experimental outcome validated that stacked ensemble always surpassed the individual base models and revealed enhanced diagnostic accuracy for ocular disease identification.

The Advantages are the model integrates the best of VGG19, ResNet50, and EfficientNetB0, enhancing accuracy and robustness for detecting ocular disease. Transfer learning is its advantage, and it minimizes overfitting, proving efficient even with minimal data. And weaknesses are that the model is computationally intensive and slower when inferring, which can limit real-time applications.

### *C. Mathematical Modelling*

1) **Convolutional Neural Networks (CNNs):** Each base model uses this mathematically, convolutions apply kernels (filters) sliding over images.

$$(I * K)(x, y) = \sum_m \sum_n I(m, n) \cdot K(x - m, y - n) \quad (1)$$

2) **Sigmoid Activation for Multi-label Classification:** Used in the final output layer for multi-label classification to output independent probabilities per class.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

3) **Softmax Activation for Classification:** The output layer (in certain base models or tasks) uses the softmax function to convert logits to class probabilities:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (3)$$

4) **Stacked Ensemble Learning:** The predictions from the three CNNs are combined as input features to a meta-classifier (MLP).

$$y = f(x) = \text{MLP}(x) \quad (4)$$

5) **Loss Function and Optimization:** The training optimizes categorical cross-entropy loss:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (5)$$

### *D. Results and Evaluation*

For this research, the models were measured in terms of their performance based on accuracy as the main training and validation measure. Accuracy gives a direct indication of the model's performance in labeling ocular disease images correctly. For the three pretrained models, ResNet50 recorded the highest training accuracy of 100%, signifying that it performed the training data to perfection but only had a validation accuracy of 96.79%, implying there was slight overfitting. EfficientNetB0 showed a healthy balance with 97.15% training accuracy and 95.41% validation accuracy with some generalization ability. VGG19 had a validation accuracy of 91.74%, suggesting sparse capacity to learn for this data as shown in Figure 5.

For even better prediction performance, a stacked ensemble model was used by fusing the results of all three base models via a meta-classifier. The ensemble method really improved the results, with an accuracy of 98.17% on the validation set that outperformed the individual performance of all three models as shown in Table II. This shows the capability of ensemble learning in combining the strengths of many architectures to come up with better and more reliable classification. The accuracy shown was computed after applying a threshold to the model's sigmoid predictions.

TABLE II: Accuracy comparison between base models and the stacked ensemble

| Model | Accuracy (%) |
|---|---|
| VGG19 | 91.74 |
| EfficientNetB0 | 95.41 |
| ResNet50 | 96.79 |
| Stacked Ensemble | 98.17 |

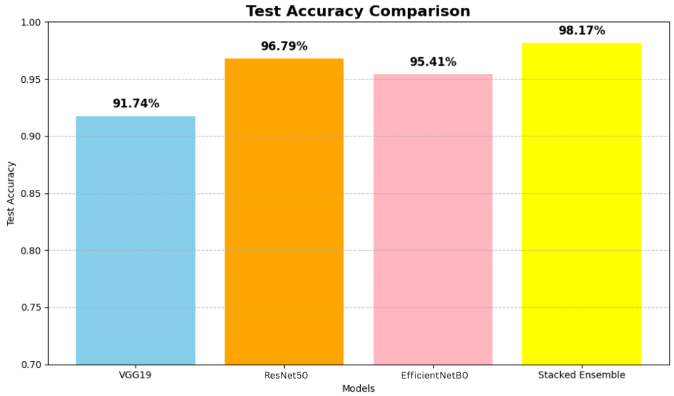


Fig. 5: Comparison of models

## V. CONCLUSION

In this research work, we proposed a stacked ensemble deep learning model for automatic cataract detection from ocular fundus images. By combining outputs of three pretrained CNNs, namely VGG19, ResNet50, and EfficientNetB0, into a meta-classifier, the model attained a strong validation accuracy of 98.17%, which outperformed individual models. Robustness was enhanced through preprocessing operations like rescaling, normalization, and data augmentation, and transfer learning that sped up training and minimized the necessity for large labeled sets. Employing the publicly accessible ODIR-5K dataset guaranteed diversity and pragmatism.

Confusion matrix analysis showed no false negatives, validating the model's accuracy in cataract detection. Nonetheless, drawbacks such as reliance on labeled data, computational intensity, and limited interpretability limit its applicability to low-resource environments. Future studies will extend the model to multi-class ocular disease classification, optimize lightweight architectures for real-time application, improve interpretability through tools like Grad-CAM or SHAP, validate across various datasets, and engage with healthcare providers for clinical implementation.

## REFERENCES


[1] G. Litjens et al., A survey on deep learning in medical image analysis, Medical Image Analysis, vol. 42, pp. 60– 88, 2017.

[2] D. S. W. Ting et al., Deep learning in ophthalmology, the technical and clinical considerations, Progress in Retinal and Eye Research, vol. 72, p. 100759, 2019, doi: 10.1016/j.preteyeres.2019.100759.

[3] P. Sharma and M. Singh, Automated cataract detection using deep learning techniques, Computers in Biology and Medicine, vol. 121, p. 103761, 2020, doi: 10.1016/j.compbiomed.2020.103761.

[4] R. Gargeya and T. Leng, Automated identification of diabetic retinopathy using deep learning, Ophthalmology, vol. 124, no. 7, pp. 962–969, 2017.

[5] T. Jiang et al., Ensemble learning-based automated detection of diabetic retinopathy in fundus images Computers in Biology and Medicine, vol. 134, p. 104453, 2021.

[6] R. Rajalakshmi et al., Validation of smartphone based retinal photography for diabetic retinopathy screening, PLoS ONE, vol. 13, no. 9, 2018.

[7] R. Chalakkal et al., A survey on AI-based diagnosis for eye diseases Comput. Biol. Med., vol. 134, p. 104443, 2021.

[8] J. Sahlsten et al., "Deep learning fundus image analysis for diabetic retinopathy and macular edema grading," Sci. Rep., vol. 9, no. 1, pp. 1–9, 2019.

[9] D. S. Kermany et al., Identifying medical diagnoses and treatable diseases by image-based deep learning Cell, vol. 172, no. 5, pp. 1–11.e9, 2018.

[10] M. Mohammadi et al., A lightweight CNN model for cataract detection on mobile platforms J. Med. Syst., vol. 46, no. 2, p. 12, 2022.

[11] Z. Li et al., Deep learning for detecting glaucoma in fundus photographs, Ophthalmology, vol. 127, no. 3, pp. 328–339, 2020, doi: 10.1016/j.ophtha.2019.05.029.

[12] D. Mahapatra et al., "Attention-guided CNN for retinal disease classification," IEEE J. Biomed. Health Inform., vol. 25, no. 6, pp. 1965–1973, 2021.

[13] R. Rajalakshmi et al., AI-based screening of diabetic retinopathy on smartphone-acquired fundus images, Indian J. Ophthalmol., vol. 69, no. 3, pp. 567–573, 2021.

[14] M. Tiwari and A. Tripathi, Hybrid deep learning model for cataract detection, Pattern Recognit. Lett., vol. 167, pp. 68– 74, 2023, doi: 10.1016/j.patrec.2022.12.004.

[15] X. Zhang et al., Multi-scale CNN model for ophthalmic disease detection, Comput. Methods Programs Biomed., vol. 196, p. 105569, 2020.

[16] Y. Xu et al., "Transfer learning with EfficientNet for ophthalmic disease classification," Biomed. Signal Process. Control, vol. 68, p. 102718, 2021.

[17] A. Dutta et al., "A lightweight CNN for real-time cataract detection," IEEE Access, vol. 10, pp. 1503–1512, 2022.

[18] S. J. Pan, Y. Zhang, X. Zeng, and H. Yu, "Multi-task learning for diabetic retinopathy grading and lesion detection," IEEE Transactions on Cybernetics, vol. 52, no. 7, pp. 6863–6875, 2022.

[19] F. Li, J. Liu, W. Zhang, and D. Song, "Attention-based CNN for glaucoma detection: A comparative study using fundus images," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 10, pp. 3793–3803, 2021.

[20] M. Almazroi, R. Alnasr, H. Alquhayz, and A. Alharbi, "Deep learning-based ensemble approach for detection of diabetic retinopathy in fundus images," IEEE Access, vol. 8, pp. 115378–115388, 2020.

[21] Cheok, M.J., Omar, Z. and Jaward, M.H., 2019. A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10, pp.131-153.

[22] Kale, S., and R. Shriram. "Suspicious Activity Detection Using Transfer Learning Based ResNet Tracking from Surveillance Videos." In Proceed- ings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020), vol. 1383, pp. 251–260. Springer, Cham, 2021.

[23] S. Mali, M. Bohari, A. Gurnule, D. Chikmurge, and S.Barve, "Enhancing dermatological diagnosis: A deep dive for skin lesion classification," in Proc. 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon), 2024, pp. 1–6.