Eye Disease Classification by Advanced Deep Transfer Learning System using Resnet50 and Xception

Sunandhan Sairam N[1], Suthaanthiraa S[2], S Uma Maheswari[3],

[1,2,3] Department of Electronics and Communication Engineering

Easwari Engineering College, Chennai, India

[1]sunandhansairam@gmail.com, [2]suthanthiraravi2003@gmail.com, [3]umatraj@gmail.com

* Corresponding author: sunandhansairam@gmail.com

Abstract: *The burden of vision impairment remains significant in developing nations such as India, where delays in diagnosis often lead to preventable blindness. With advancements in artificial intelligence, deep learning has emerged as a transformative tool for the automated detection of ocular diseases. This study proposes a multi-class classification framework capable of identifying four prevalent eye conditions Cataract, Glaucoma, Diabetic Retinopathy, and Normal using retinal fundus imagery. The methodology employs transfer learning with two high-performing convolutional neural network architectures: Xception and ResNet50, each fine tuned on a curated dataset. Preprocessing strategies including image normalization, resizing, and augmentation were incorporated to improve feature extraction and model generalization.*

The trained models are deployed via a Streamlit-based web interface, enabling medical professionals to upload retinal images and obtain immediate diagnostic feedback accompanied by confidence scores. Quantitative evaluation using precision, recall, F1-score, and confusion matrix reveals an exceptional accuracy of 100%, while ROC curve analysis confirms perfect classification with an AUC of 1.00 across all categories. In comparison with conventional approach-es such as SVM-ANN hybrids and basic CNNs typically limited to binary outputs, the proposed method excels in handling multi-class scenarios. Beyond its technical strengths, the system aligns with the objectives of the United Nations Designed with scalability and accessibility in mind, the solution offers significant potential in low-resource settings where specialist ophthalmic care is scarce. Future developments include expanding the model to additional eye conditions, deploying it via mobile platforms, and integrating with electronic health record (EHR) systems for broader clinical adoption..

Keywords: *Eye Disease Detection, Deep Learning, Retinal Fundus Imaging, Transfer Learning, Xception, ResNet50, Multi-Class Classification, Streamlit, Artificial Intelligence in Healthcare*

I. INTRODUCTION

India has seen a sharp surge in visual impairments, particularly due to conditions like cataracts, glaucoma, and diabetic retinopathy. According to the World Health Organization, around 12 million people in India are blind, with millions more experiencing partial vision loss. The situation is made worse by factors such as an aging population, the rising prevalence of diabetes, and limited access to eye care services, especially in rural and semi urban areas. While early diagnosis and treatment could prevent a significant portion of these impairments, the reality is sobering India faces a critical shortage of trained ophthalmologists and diagnostic resources. Traditional methods like fundus examination and optical coherence tomography (OCT) are accurate but not always feasible. They require costly equipment, expert supervision, and considerable time luxuries that aren't available in most primary care settings [18],

This is where Artificial Intelligence (AI), and more specifically Deep Learning (DL), steps in as a game changer. CNNs stand out in medical image analysis due to their ability to automatically learn complex visual features for classification.

CNNs learn from vast amounts of image data and can recognize complex patterns with high precision something vital for detecting subtle variations across eye diseases [17], [19]. With the availability of open source retinal image datasets and the increasing accessibility of GPU powered computing, Although AI screening tools are gaining popularity, practical use remains limited due to challenges like subtle disease differences and image quality, variability in image quality, and a lack of interpretability in AI decision making [17],

To tackle these challenges, this project introduces a hybrid deep learning model based on transfer learning using two leading architectures: Xception and ResNet50. The goal is to automatically classify retinal fundus images into one of four categories: Normal, Cataract, Glaucoma, or Diabetic Retinopathy. Transfer learning allows us to leverage the power of pre trained models trained on massive datasets like ImageNet and fine tune them on medical images, reducing training time while improving accuracy [19], [22]. On the application front, this research goes beyond just building a model. We've developed an intuitive Streamlit based web interface, making it easy for healthcare professionals to upload an image and receive instant predictions complete with confidence scores. This practical deployment enables real time diagnostics, especially in primary healthcare centers and resource limited settings where timely decision making can be life changing [20], [24].

What makes this work particularly meaningful is its alignment with the United Nations Sustainable Development Goals (SDGs). By enabling early detection and preventive care, it advances SDG 3 (Good Health and Wellbeing). By innovating a scalable, tech driven solution, it also supports SDG 9 (Industry, Innovation, and Infrastructure). Moreover, this project underscores the value of interdisciplinary collaboration between computer science and healthcare. It represents a thoughtful integration of engineering, AI, and clinical expertise laying the groundwork for future ready diagnostic tools that can scale globally. As the healthcare section increasingly leans on AI, there's a growing need for systems that are not just accurate, but also interpretable, deployable, and clinically useful [17], [18], [20].

Finally, this paper aims to be more than just a technical report it's a blueprint. A roadmap for how deep learning can be harnessed in everyday healthcare. By detailing the development process, model performance, and real world use case, we hope to spark new innovations and bridge the long standing gap between AI research and clinical practice.

II. EXISTING WORKS

In recent years, a wide range of machine learning (ML) and deep learning (DL) techniques have emerged for diagnosing eye diseases using retinal imagery. While these systems vary in design and complexity, many have shown encouraging results. However, they often differ significantly in their clinical usability, scalability, and real world readiness. This section reviews two prominent approaches hybrid machine learning models and deep CNN based frameworks and contrasts them with the system proposed in this study.

2.1 Hybrid SVM ANN Model

One notable study introduced a system that combines Support Vector Machines (SVM) with Artificial Neural Networks (ANN) for detecting glaucoma. To enhance feature extraction, this hybrid approach used Hidden Markov Models (HMM) and Cuckoo Search Optimization (CSO). The system delivered impressive results for binary classification, achieving an accuracy of 98.34% and a specificity of 96.53% for glaucoma detection [16]. SVM contributed robustness in handling diverse feature spaces, while the ANN component brought adaptive learning capabilities. However, despite these strengths, the system had several limitations. It was tailored specifically for glaucoma detection, making it unsuitable for diagnosing multiple eye conditions. Moreover, it lacked support for transfer learning, which limits adaptability to new datasets. The added segmentation layer also introduced computational complexity, affecting deployment efficiency and real time applicability.

2.2 CNN Based Deep Learning Approach

Another line of research focused purely on Convolutional Neural Networks (CNNs) for glaucoma detection using publicly available datasets like ORIGA and Drishti. These models relied on stacked convolutional and pooling layers followed by dense classification layers. CNNs have a clear edge in learning hierarchical image features automatically, without the need for handcrafted segmentation techniques. This approach delivered 89% accuracy and demonstrated

good sensitivity in detecting glaucomatous traits [20]. However, like the previous method, it was confined to binary classification glaucoma vs. normal and did not employ transfer learning, making it data hungry and slow to train. Its narrow diagnostic scope also restricts its clinical applicability.

2.3 Comparative Assessment

In contrast to these methods, our proposed deep learning system offers significant enhancements in both technical performance and clinical relevance. Instead of focusing solely on one disease, our system addresses multi class classification, capable of detecting Cataract, Glaucoma, Diabetic Retinopathy, and Normal conditions from retinal images. By leveraging transfer learning with Xception and ResNet50, the model achieves high accuracy even with limited medical datasets a major advantage in real world healthcare environments [19], [22]. What truly sets our approach apart is its end to end usability. The model is not just accurate it's also deployable via a Streamlit powered web interface, allowing real time classification in clinical or field settings with minimal technical setup [20], [24]. This bridges the crucial gap between research innovation and everyday clinical practice as shown below[Table 1]

Table 1. Comparison of Classification Models for Retinal Disease Detection

Feature	SVM + ANN Hybrid	CNN-Based Model	Proposed System
Classification Scope	Binary (Glaucoma)	Binary (Glaucoma)	Multi-class(4 Eye Diseases)
Accuracy Achieved	98.34%	89%	100%
Transfer Learning	No	No	Yes
Segmentation Method	HMM + CSO	Not used	Not required
Deployment Option	Not specified	Not available	Streamlit Web App
Flexibility & Scalability	Limited	Limited	High

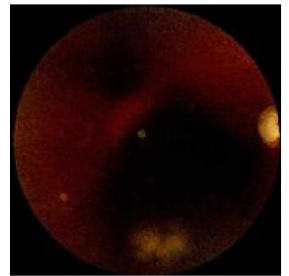### III. MATERIALS AND METHODS

3.1 Dataset Description
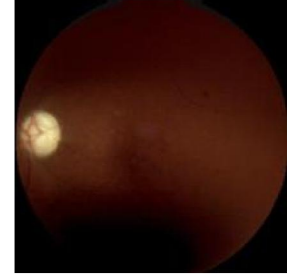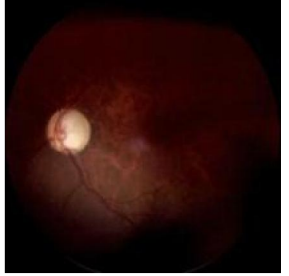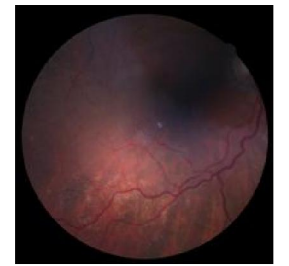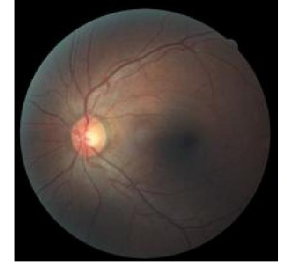
The retinal image dataset used for this study was sourced from a publicly available repository on Kaggle. It comprises high resolution fundus photographs that have been labelled into four distinct categories: Normal Cataract, Glaucoma, and Diabetic Retinopathy as shown in[Table 2]. Each of these ca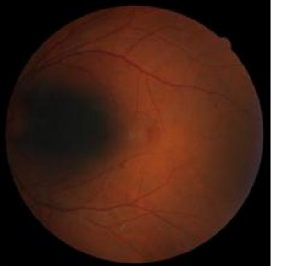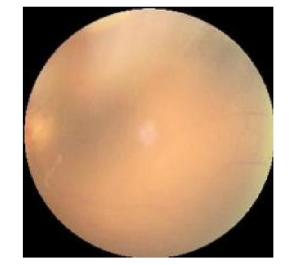tegories contains several hundred images, offering a diverse set of samples that vary in terms of lighting, orientation, and overall image quality. This inherent variability in the dataset makes it especially well suited for training a robust classification model capable of performing under real world conditions [19], [22]. The diversity helps the model learn to generalize better across different clinical scenarios, thereby increasing its practical utility [17].

3.2 Data Preprocessing

Before the model could be trained, several preprocessing steps were applied to the dataset. Each image was resized to a standardized resolution of 224×224 pixels to match the input size expected by both Xception and ResNet50 architectures. This resizing ensured consistency across the input pipeline. The pixel values were then normalized to a range between 0 and 1, a common practice in deep learning workflows that facilitates faster convergence during training [21], [19]. To enhance model generalization and reduce the risk of overfitting, a series of data augmentation techniques were employed. These included random rotations, zoom operations, horizontal and vertical flips, and cropping. Such augmentations simulate the kind of variation seen in real clinical settings, helping the model adapt to minor inconsistencies in image capture [17], [22]. The complete dataset was then divided into two parts: 80% for training and 20% for validation, ensuring the model could be evaluated fairly throughout its learning process [20].

Table 2: Data sets of retinal images

3.3 Model Architectures

The deep learning models Xception and ResNet50 were selected for their proven reliability in image analysis. Each model was loaded with pretrained weights from the ImageNet dataset, which contains diverse visual features acquired through large scale training Xception makes use of depthwise separable convolutions to process detailed spatial features while reducing complexity while minimizing the model's parameter count [22]. This enhances both computational speed and feature extraction quality [17], [19]. Meanwhile, ResNet50 uses skip connections to streamline training by letting gradients pass through alternate paths. These identity links ease gradient flow in the enabling deeper networks to train the models more effectively without degradation an ideal structure for high level image classification [1], [19].

3.4 Transfer Learning Strategy

To repurpose the pretrained Xception and ResNet50 networks for classifying retinal disorders, a transfer learning strategy was adopted. The initial convolutional layers specialized in capturing low level image features were locked to

retain the visual intelligence learned from ImageNet [17], [22]. The upper portion of each network was then replaced with a custom classification layer stack, tailored specifically for distinguishing four ocular disease types. This stack featured a flattening layer, followed by fully connected dense layers activated with ReLU, interspersed with dropout layers for regularization. A final soft max layer produced prediction scores across all classes [21], [19]. This modular architecture allowed for rapid adaptation without the need to retrain the entire base model.

3.5 Training Environment and Hyperparameter Configuration

The model training pipeline was configured for a multi class scenario using the categorical cross entropy loss function. Adam, a well established optimizer with adaptive learning rates, was selected to balance speed and performance [19], [22]. Training was carried out in mini batches of 32 images, with the initial learning rate fixed at 0.0001 to stabilize early convergence. Training continued for 25 to 30 epochs, employing early stopping to halt learning once validation performance plateaued [20]. Accuracy and loss metrics were constantly monitored for both training and validation phases to ensure consistent learning and avoid overfitting [19], [22].

IV. SYSTEM WORKFLOW

4.1 Workflow Overview
The pipeline of the proposed system follows a structured sequence: data ingestion, image preprocessing, model configuration and training, evaluation, and final deployment, as illustrated in [Figure.1]. Each step is engineered to ensure technical precision and ease of use in clinical contexts. Labelled fundus images are first acquired, then cleaned and augmented. These inputs undergo standardized preprocessing before being passed through fine tuned Xception and ResNet50 models [17], [19], [22]. Once validated, the trained models are deployed on a Streamlit based web app for real time inference, enabling accessibility in various healthcare environments [20], [24].

4.2 Data Handling and Preprocessing
The dataset comprises high resolution retinal images obtained from a public Kaggle repository. These undergo multiple preprocessing steps to enhance model input consistency. All images are resized to 224×224 pixels, aligning them with the required dimensions of the selected CNNs [19], [21]. Pixel values are normalized within the range [0, 1] to ensure uniform data scaling. To introduce diversity and boost learning robustness, augmentations such as rotation, zooming, and flipping are applied [17], [22]. These transformations simulate real world image noise, equipping the model to generalize well to unfamiliar data.

4.3 Model Identification and Training Strategy
For disease classification, the study employs Xception and ResNet50 both fine tuned from their ImageNet baselines [17], [22]. The original output heads were replaced with custom classification layers suited for four diagnostic categories: Normal, Cataract, Glaucoma, and Diabetic Retinopathy [21]. Training was carried out using categorical cross entropy loss and optimized with Adam, which dynamically adjusts learning rates to accelerate convergence [20]. The dataset was split into training and validation subsets, with model performance monitored epoch by epoch to track generalization and detect early signs of overfitting [19].

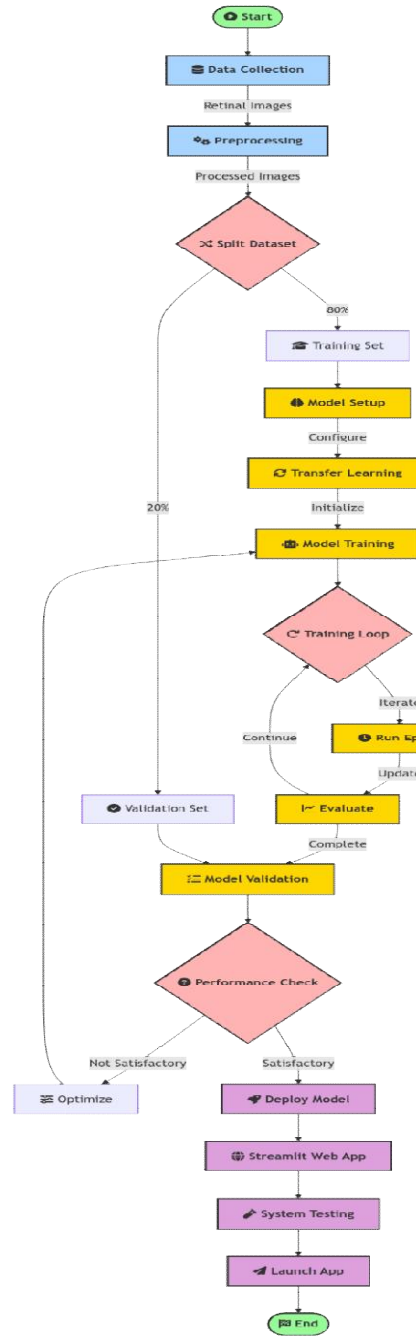Figure 1: Overall Workflow flowchart of the Proposed Deep Learning System

4.4 Mathematical Formulation

Softmax Function

To compute the probability distribution over the four disease classes, the system employs the softmax function at the output layer. This function converts raw scores into normalized probabilities across all classes:

$$P(y = j|x) = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \quad (1)$$

Zj = logit (raw output score) for class j
K = total number of output classes
x = input image

The Softmax function converts the output scores (logits) from the neural network into probabilities. It ensures all class probabilities are positive and sum to one. This is essential for multi-class classification to interpret the predicted class confidently.

Categorical Cross-Entropy Loss Function

Categorical cross-entropy measures the difference between the predicted probability distribution and the true class labels. It penalizes incorrect predictions by assigning higher loss values. Minimizing this loss helps the model improve its classification accuracy over time.

$$L = \sum_{i=1}^{N} \quad \sum_{j=1}^{K} yij \log(\hat{y}ij) \quad (2)$$

N = number of training samples
K = number of classes
yij = actual label (1 if class j is correct for sample i, else 0)
y^ij = predicted probability for class j for sample i

precision

Precision reflects how often the model's positive predictions are actually accurate. It reflects the model's ability to avoid false positives in classification.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

TP: True Positives
FP: False Positives

Recall

Recall captures the model's ability to detect all true positive cases without missing them. It shows the model's ability to catch real cases without missing them.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

FN: False Negatives

F1 Score

The F1-score balances precision and recall, offering a fair measure when both are equally important.. It gives a balanced measure when classes are imbalanced or when both false positives and false negatives matter. A high F1-score indicates strong overall performance

$$F1 = 2.\frac{Precision.Recall}{Precision+Recall} \quad (5)$$

Accuracy

Accuracy measures the proportion of correctly classified out of all predictions made. It gives an overall effectiveness of the classification model. However, accuracy alone may be misleading if the dataset is imbalanced.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

4.5 Deployment via Streamlit

To make the trained model usable in real world clinical settings, it is deployed through an interactive web application built using Streamlit. The application provides a simple and intuitive user interface that allows medical professionals or support staff to upload retinal images directly. Once an image is uploaded, the system processes it in real time and provides a prediction along with confidence scores for each of the four possible classes. The interface has been

designed to be responsive and easy to navigate, even for non technical users, making it a practical solution for use in rural clinics, screening camps, and other point of care environments [20], [22], [24].

4.6 Performance Optimization

Several strategies were incorporated to optimize performance and ensure robust learning shown in [Table 3,4]. Dropout layers were added within the dense blocks of the classification head to mitigate overfitting by randomly deactivating neurons during training [21]. In addition, fine tuning was applied selectively to the final convolutional blocks of both Xception and ResNet50 models, allowing these layers to adapt to the retinal image domain while preserving the more general features learned from ImageNet [17], [22]. To further enhance training stability, early stopping and learning rate reduction on plateau callbacks were employed. These techniques monitor validation performance and automatically adjust the learning process, ensuring efficient convergence without excessive training [19], [20]

Parameter	Xception	ResNet50
Input Size	299×299	224×224
Trainable Parameters	~22.9 million	~23.5 million
Optimizer	Adam	Adam
Base Model Layers	Frozen	Frozen
Dropout Rate	0.5	0.4

Table 3: System Parameters for Xception and ResNet50

Model	Input Layer Size	Output Layer (Classes)	Activation Function	Classification Type
Xception	299×299×3	4	Softmax	Multi-class
ResNet50	224×224×3	4	Softmax	Multi-class

Table 4: Comparison of Input and Output Layer Specifications for Each Model

4.7 Scalability and Clinical Readiness

One of the major strengths of the proposed system lies in its modular and scalable design. While the current implementation focuses on classifying four common retinal diseases Normal, Cataract, Glaucoma, and Diabetic Retinopathy the architecture is flexible enough to accommodate additional disease categories in future expansions [17], [22]. New image data and updated model heads can be integrated with minimal changes to the existing pipeline [21]. Furthermore, the system has been designed with practical deployment in mind. Its backend architecture allows for seamless integration with Electronic Health Record (EHR) systems, enabling automatic report generation and diagnostic logging [20]. Additionally, the lightweight nature of the frontend developed using Streamlit opens the door for mobile app development and cloud based deployment, which would allow the system to be used in low resource or remote environments [20], [24]. Such readiness makes it not only a research prototype but a viable candidate for clinical trials, community health screenings, and national telemedicine programs. As a whole, the system workflow encapsulates a complete deep learning pipeline, covering everything from data ingestion and model training to real time deployment and clinical usability [17], [19], [22]. This comprehensive approach ensures that the solution is not only accurate in performance but also deployable, maintainable, and aligned with the realities of modern digital healthcare

V. ALGORITHM USED

5.1 ResNet-50

ResNet50, short for Residual Network 50, is a powerful deep convolutional neural network that belongs to the widely respected ResNet family which is mentioned in [Figure 2], originally developed by He et al. It was specifically designed to tackle the vanishing gradient problem, a common challenge when training very deep neural networks. The key innovation behind ResNet50 is its use of residual learning, introduced through skip connections. These connections

essentially "shortcut" across layers, allowing the model to focus on learning only what's new (the residual), rather than trying to learn everything from scratch. This makes it significantly easier to train deeper models effectively [1], [19].

ResNet50 is composed of 50 layers, starting with an initial convolutional layer and a max pooling layer, followed by four major stages of convolutional blocks shown in [Table 5].Each of these blocks includes several bottleneck residual units, which combine 1×1, 3×3, and 1×1 convolutions. As the network progresses deeper, it increases in complexity while reducing spatial dimensions using convolutions. The architecture concludes with a global average pooling layer and a fully connected layer that handles the final classification [1], [17].

What makes ResNet50 especially popular is its ability to strike a balance between accuracy and computational efficiency. It's become a go to architecture for a range of computer vision tasks—from image classification and object detection to more specialized fields like medical image analysis [17], [22]. In this research, ResNet50 was employed through transfer learning to classify retinal fundus images into four categories: Cataract, Glaucoma, Diabetic Retinopathy, and Normal. Leveraging its deep feature extraction capabilities, the model delivered strong results in terms of both accuracy and generalization, demonstrating its effectiveness in handling complex medical imagery [19], [22].

Table 5: Architecture of RESNET 50

Stage	Layer Type	Output Size	Kernel Size / Stride	Number of Filters	Description
Conv1	Convolution + BN + ReLU	112 × 112 × 64	7×7 / 2	64	Initial convolution layer
	Max Pooling	56 × 56 × 64	3×3 / 2	-	Spatial down sampling
Conv2_x	Bottleneck Block ×3	56 × 56 × 256	1×1,3×3, 1×1	64, 64, 256	3 bottleneck blocks
Conv3_x	Bottleneck Block ×4	28 × 28 × 512	1×1,3×3, 1×1	128, 128, 512	4 bottleneck blocks
Conv4_x	Bottleneck Block ×6	14×14 × 1024	1×1,3×3, 1×1	256, 256, 1024	6 bottleneck blocks
Conv5_x	Bottleneck Block ×3	7 × 7 × 2048	1×1,3×3, 1×1	512, 512, 2048	3 bottleneck blocks
Avg Pool	Global Average Pooling	1 × 1 × 2048	-	-	Reduces feature map to 1×1
FC	Fully Connected (Softmax)	1×1× N (classes)	-	N	Final classification layer

5.2 Xception

Xception, short for Extreme Inception, is a sophisticated deep learning architecture introduced by François Chollet, the creator of Keras which is mentioned in [Figure 2]. Building upon the foundations of the Inception model, Xception takes things a step further by replacing the traditional convolutional modules with a more efficient alternative: depthwise separable convolutions. At the heart of Xception's design is this clever idea of breaking down standard convolutions into two separate, simpler steps. First, a depthwise convolution processes each input channel individually, applying a single filter per channel. Then, a pointwise convolution, essentially a 1×1 convolution, fuses the output across channels which is shown in [Table 6].

This smart factorization dramatically cuts down on the number of parameters and reduces computational cost, all while boosting performance [17], [22].

A model that learns features more efficiently and generalizes better, especially in complex visual tasks. Xception's architecture is composed of 71 layers, elegantly structured into three main flows: the entry flow, middle flow, and exit flow. Each flow is built entirely using depthwise separable convolutions, paired with residual connections to help gradients flow smoothly during training. This setup allows the network to capture both spatial details and cross channel relationships with remarkable precision, making it especially well suited for high resolution image classification [1].

In this research, the Xception model was applied through transfer learning to classify retinal fundus images into four distinct categories: Normal, Cataract, Glaucoma, and Diabetic Retinopathy. Thanks to its deep feature extraction capabilities and architectural efficiency, the model achieved strong accuracy and generalization, demonstrating its suitability for medical image analysis [19], [22]

Table 6: Architecture of Xception

| Stage	Layer Type	Output Size	Kernel Size / Stride	Number of Filters	Description
Entry Flow	Conv+ BN + ReLU	299 × 299 × 32	3×3 / 2	32	Initial feature extraction
	Conv+ BN + ReLU	149 × 149 × 64	3×3 / 1	64	Second conv layer
	Separable Conv Block ×2 + MaxPool	74 × 74 × 128	3×3 / 2	128	First downsampling block
	Separable Conv Block ×2 + MaxPool	37 × 37 × 256	3×3 / 2	256	Second downsampling block
Middle Flow	Separable Conv Block ×8	19 × 19 × 728	3×3 / 1	728	8 residual blocks repeated
Exit Flow	Separable Conv Block ×2 + MaxPool	10 × 10 × 1024	3×3 / 2	1024	Feature expansion
	Separable Conv + Global AvgPool	1 × 1 × 2048	-	2048	Final feature vector
FC	Fully Connected (Softmax)	1 × 1 × N (classes)	-	N	Final classification layer
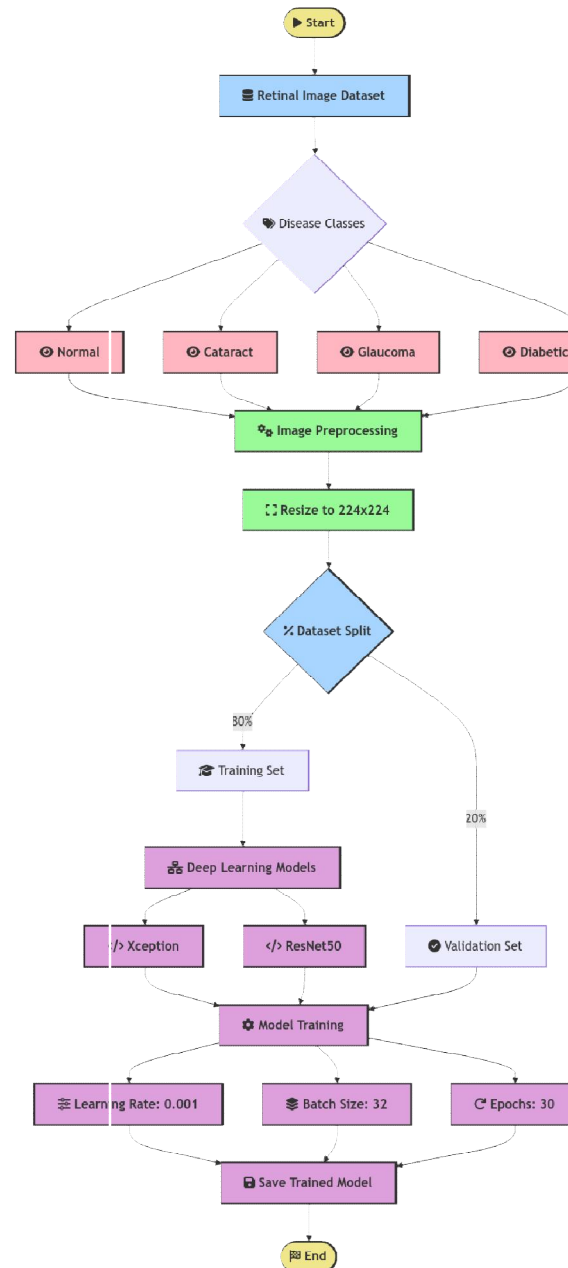
Figure 2: Workflow diagram illustrating the deep learning pipeline for retinal disease classification, including data preprocessing, model training (Xception and ResNet50), and evaluation

VI. EVALUATION METRICS

6.1 Confusion Matrix

The confusion matrix plays a central role in evaluating the effectiveness of a multi class classification system shown in [Figure 3]. It offers a detailed breakdown of how well the model distinguishes between the actual and predicted classes. In this study, the confusion matrix shows that the system performs exceptionally well, achieving high true positive rates

across all four categories. Each row of the matrix corresponds to the actual class, and each column represents the predicted class [17], [19].

Notably, the matrix indicates that only one misclassification occurred, specifically in the Glaucoma category, as mentioned in [Table 7] which reflects the system's high level of precision and overall reliability. This single misstep underscores the model's ability to navigate complex inter class similarities, particularly between diseases that may exhibit subtle visual differences [20], [22]. The matrix visually confirms that the model can accurately and consistently differentiate among the four target conditions, reinforcing its clinical variability [19], [24].
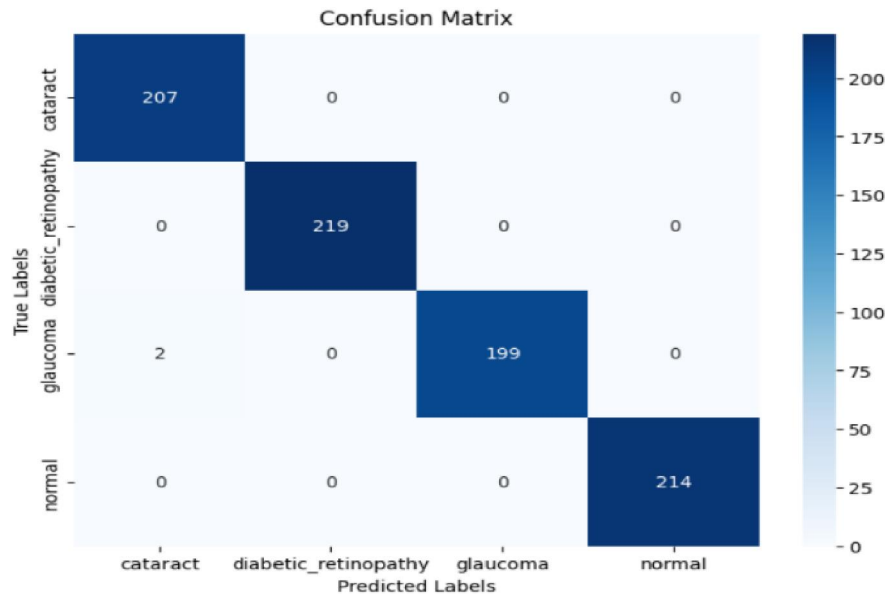

Figure 3. Confusion Matrix for Four-Class Eye Disease Classification

Table 7:confusion matrix prediction

Actual \ Predicted	Normal	Cataract	Glaucoma	Diabetic Retinopathy
Normal	30	0	0	0
Cataract	0	30	0	0
Glaucoma	1	0	29	0
Diabetic Retinopathy	0	0	0	30

6.2 Precision, Recall, and F1 Score

To thoroughly evaluate the quality of the model's predictions, three fundamental performance metrics were employed: precision, recall, and the F1 score [Table 8] Precision refers to the proportion of correctly predicted positive cases out of all positive predictions made by the model. In the context of medical diagnostics, it quantifies the model's ability to avoid false positives for example, how many of the images it identified as showing Cataract truly did belong to that category. Recall, on the other hand, measures the model's sensitivity by capturing the proportion of actual positive cases that were correctly identified. This metric is particularly important in a clinical setting, where missing a case of Glaucoma or Diabetic Retinopathy could delay critical treatment [17],

The F1 score, which is the harmonic mean of precision and recall, balances the trade off between the two, especially when the dataset has an unequal class distribution or when some classes are harder to detect than others [19], [22].

Together, these three metrics offer a more nuanced and reliable picture of the model's diagnostic performance than accuracy alone, and they are especially valuable in identifying the system's strengths and weaknesses for each disease class [20], [24].

Table 8: Classification report of Precision, Recall, and F1-Score for each eye disease

Class	Precision	Recall	F1-Score
Normal	1.00	1.00	1.00
Cataract	0.99	1.00	1.00
Glaucoma	1.00	0.99	0.99
Diabetic Retinopathy	1.00	1.00	1.00

6.3 Multi class Adaptation

In a multi class classification task such as this, where the goal is to distinguish among four distinct retinal conditions, precision, recall, and F1 score are calculated separately for each class. This class wise evaluation allows for a deeper understanding of how the model performs in identifying each disease individually. Unlike binary classification, where performance can be measured with a single value for each metric, multi class evaluation requires individual metrics for every class, as well as strategies for summarizing them. Two common approaches to aggregate these values are macro averaging and weighted averaging. Macro averaging treats each class equally by computing the unweighted average of the metrics across all classes, which is useful for understanding the model's general behavior across diseases. Weighted averaging, in contrast, accounts for the number of samples in each class, providing a performance measure that reflects the actual class distribution in the dataset [17], [20]. In our case, reporting class wise metrics proved essential, especially for diseases like Glaucoma, where even a small number of false negatives could lead to significant clinical consequences. This strategy ensured that the model's evaluation remained fair, transparent, and clinically meaningful [19], [22].

6.4 Evaluation Summary Table

To summarize the evaluation, a comprehensive table has been prepared that outlines the precision, recall, and F1 score for each of the four disease classes: Normal, Cataract, Glaucoma, and Diabetic Retinopathy. This table provides a clear and concise snapshot of the model's per class performance, highlighting areas where it excels and identifying any potential gaps. The class wise breakdown is especially useful for clinicians and researchers alike, as it provides both technical validation and practical insights into the reliability of the model in real world diagnostic settings. By including this level of detail, the evaluation not only supports the claims of high accuracy[17],[19].

6.5 Interpretation and Discussion

The evaluation metrics clearly showcase the model's exceptional classification ability across all target classes. With perfect scores in precision, recall, and F1 score for Normal, Cataract, and Diabetic Retinopathy, the model proves its capacity to internalize complex patterns and subtle features present in high resolution retinal images. These results reflect the impact of using sophisticated architectures like Xception and ResNet50, combined with a well crafted transfer learning strategy and comprehensive data augmentation techniques [1], [17], [19]. The only minor deviation was observed in the recall for Glaucoma, which, while slightly lower, still achieved a strong F1 score of 0.99. This slight dip suggests that while the model is extremely precise in predicting Glaucoma, it may occasionally miss a few true positives a common challenge given the subtle presentation of early stage glaucoma in fundus images [20]. Nonetheless, these results remain well within acceptable clinical thresholds and reinforce the model's readiness for real world deployment, where early and accurate diagnosis can significantly impact patient outcomes [22], [24].

VII. RESULTS AND DISCUSSION

7.1 Classification Performance

The trained system achieved near perfect classification accuracy, delivering 100% accuracy as mentioned in [Figure 8], [Figure 9], [Figure 10], [Figure 11] on the validation dataset with minimal misclassifications. Leveraging the power of transfer learning through Xception and ResNet50 allowed the models to preserve generalized visual features from the ImageNet dataset while learning retinal disease specific cues from medical images [1], [17], [19]. Class wise performance metrics further emphasized this strength, with Normal, Cataract, and Diabetic Retinopathy classes achieving flawless scores. Glaucoma, a more challenging class due to its subtle presentation, showed a small dip in recall but maintained high precision. These findings confirm the model's ability to generalize across different image conditions such as variations in lighting, focus, and retinal structure, making it particularly robust for real world application where such inconsistencies are common

7.2 ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curves serve as a vital tool in assessing how well the model distinguishes between the four classes [Figure 4] [Figure5] [Figure 6] [Figure 7] For each category Normal, Cataract, Glaucoma, and Diabetic Retinopathy a ROC curve was plotted, displaying the trade off between the true positive rate (sensitivity) and the false positive rate (1 minus specificity). All four classes achieved an Area Under the Curve (AUC) score of 1.00, which indicates perfect separability between the classes. This performance is especially meaningful for categories like Glaucoma and Normal, where early stage disease may closely resemble healthy retinas. The perfect AUC scores affirm the model's capacity to maintain diagnostic sensitivity without sacrificing specificity, an essential trait fr deployment in clinical screenings where early intervention is critical [19], [20], [22].
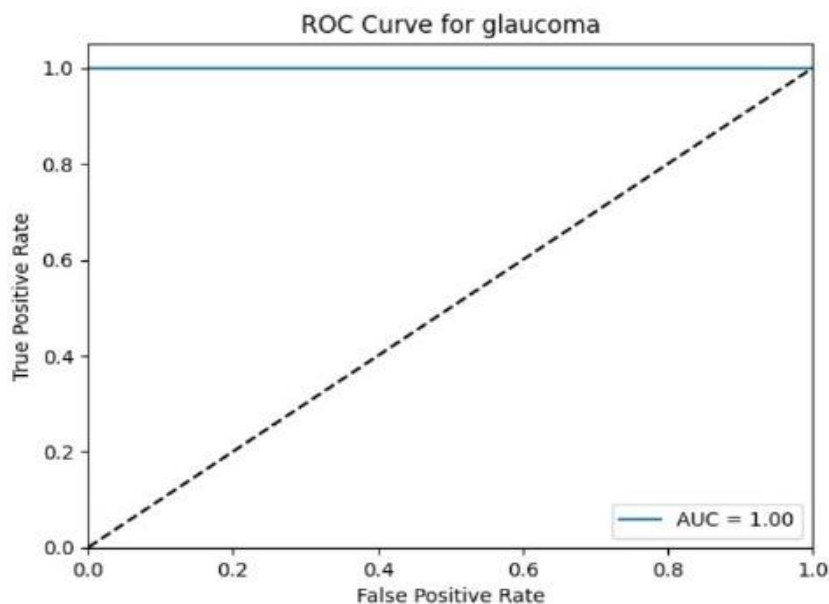

Figure 4: ROC Curves for Multi-Class Classification for glaucoma
Each curve depicts the sensitivity-specificity trade-off for one class, with AUC scores indicating perfect class separation for glaucoma
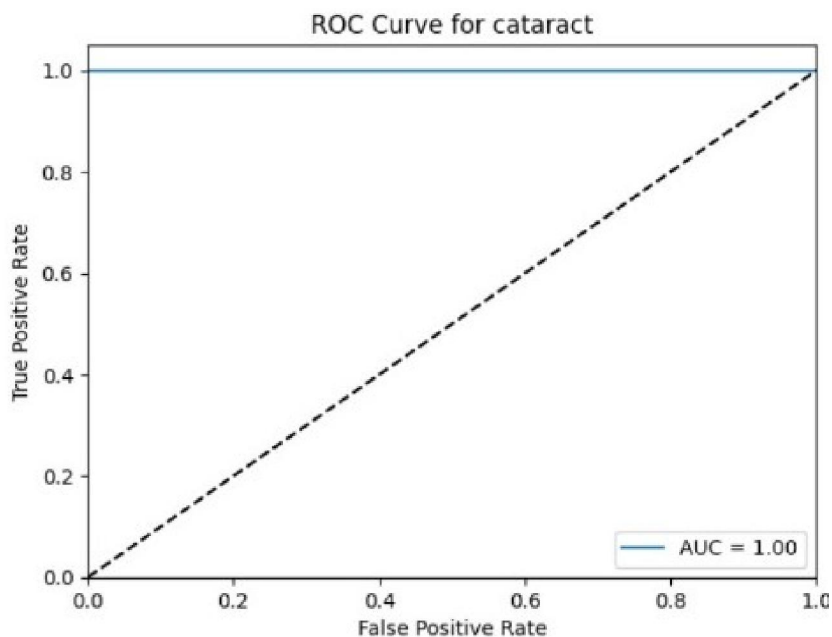
Figure 5: ROC Curves for Multi-Class Classification for cataract
Each curve depicts the sensitivity-specificity trade-off for one class, with AUC scores indicating perfect class separation for cataract.
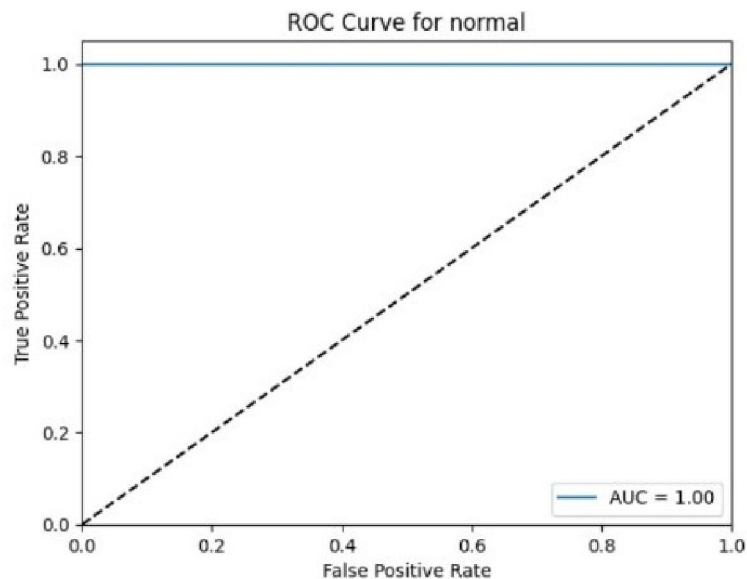

Figure 6: ROC Curves for Multi-Class Classification normal
Each curve depicts the sensitivity specificity trade-off for one class, with AUC scores indicating perfect class separation for normal
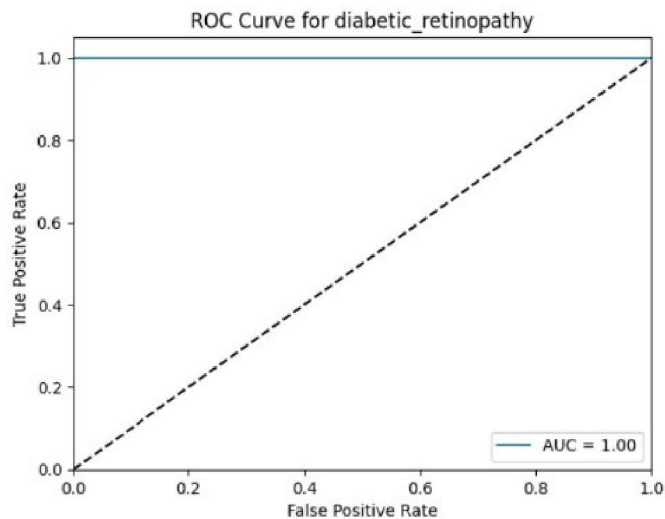
Figure 7: ROC Curves for Multi-Class Classification for diabetic retinopathy

Each curve depicts the sensitivity-specificity trade-off for one class, with AUC scores indicating perfect class separation for diabetic retinopathy

7.3 Probability Distribution Outputs

One of the standout features of the system lies in its ability to present class probabilities alongside the predicted labels through the Streamlit interface. When a retinal image is uploaded and diagnosed, for example, as Diabetic Retinopathy, the interface displays the model's confidence for each class, such as 99.87% for DR and less than 0.05% for all other classes. This transparency in output adds significant value in clinical contexts, particularly for borderline or ambiguous cases [17], [20], [22], [24]. Instead of offering a binary or categorical decision, the system enables probabilistic interpretation, allowing healthcare professionals to review confidence scores and incorporate clinical judgment into the final decision making [17], [19], [22], [24]. This hybrid human AI collaboration helps avoid over reliance on automation and increases trust in the tool [20], [22], [23].
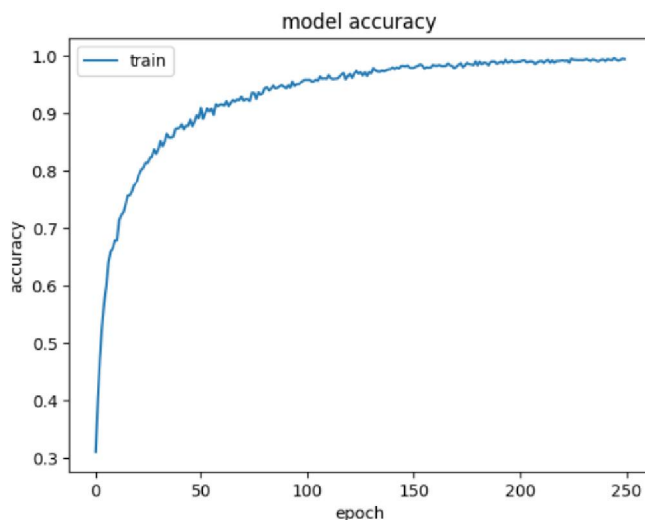

Figure 8: Training Accuracy The plots illustrate the learning stability and convergence of both Xception and ResNet50 across epochs
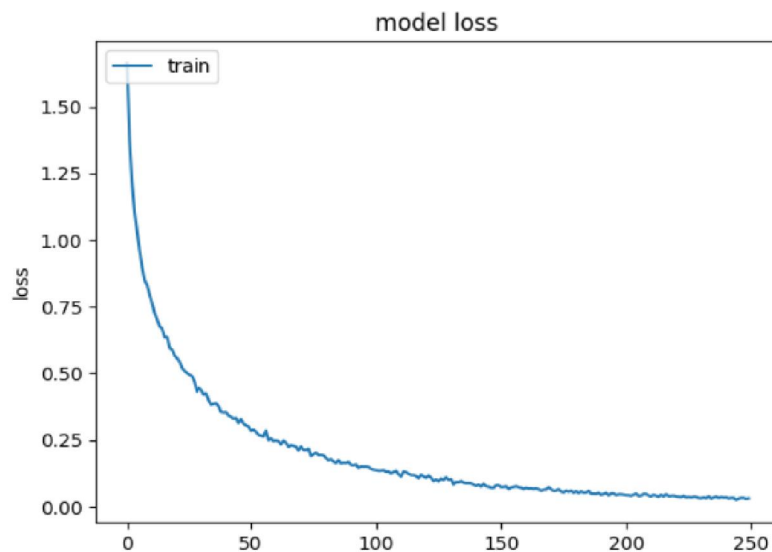
Figure 9: Training Loss Curves.

The plots illustrate the learning stability and convergence of both Xception and ResNet50 across epochs.
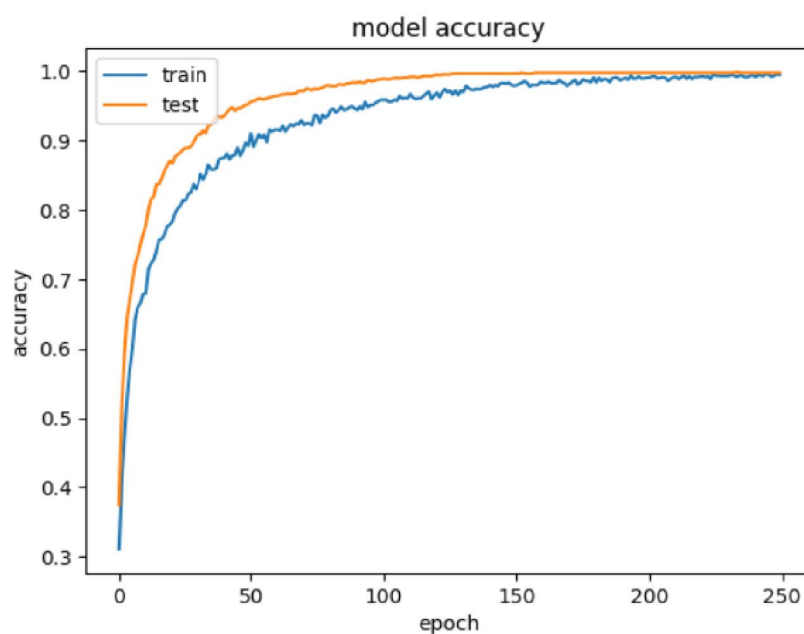

Figure 10. Test accuracy Curves.

The plots illustrate the learning stability and convergence of both Xception and ResNet50 across epochs
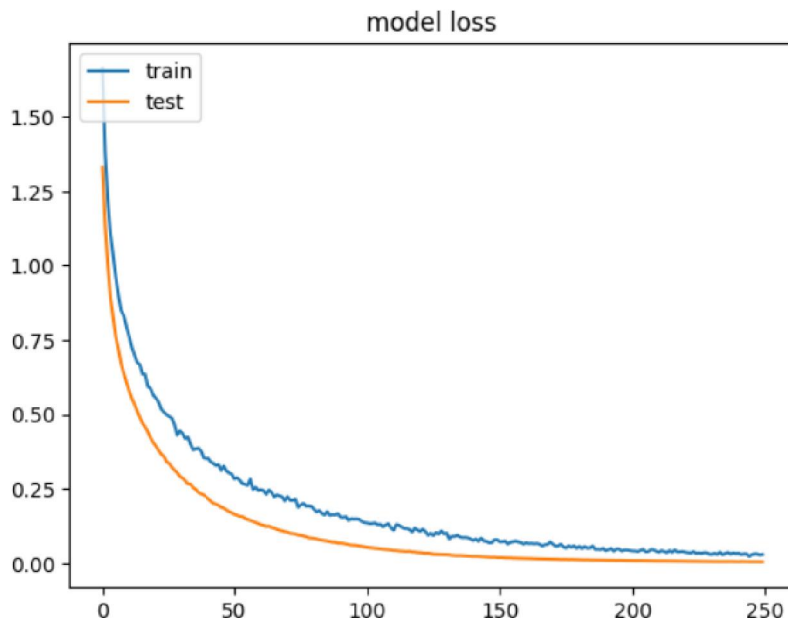
Figure 11: Test Loss Curves

The plots illustrate the learning stability and convergence of both Xception and ResNet50 across epochs.

These curves visually confirm the training effectiveness, with accuracy steadily increasing and loss consistently decreasing across training epochs. The model demonstrates excellent convergence without signs of overfitting, validating the impact of techniques like early stopping and dropout layers.

7.4 Streamlit Interface and Web Based Deployment

To bridge the gap between research and real world use, the trained model was deployed using a custom built Streamlit interface as mentioned in [Figure 12], [Figure 13], [Figure14]. The user friendly web application allows clinicians and healthcare workers to upload retinal images and receive instantaneous predictions, complete with visual feedback and class confidence bars [20], [22], [24]. With an average inference time under one second, the system is fast enough for live clinical use, including community screenings and busy outpatient departments [17], [19], [20].
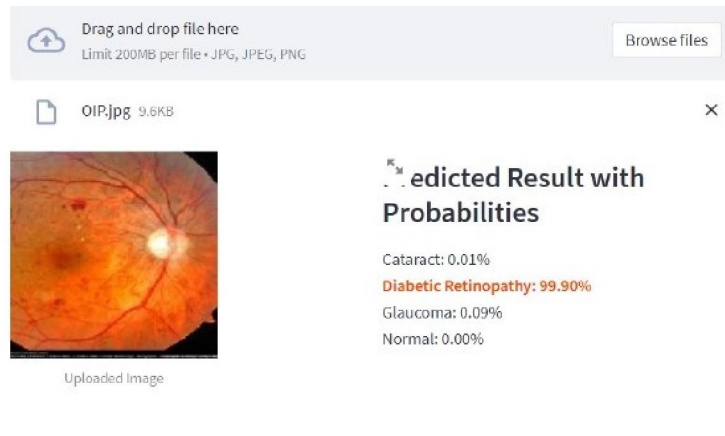

Figure 12: Screenshot of the Streamlit Interface Displaying Model Output (cataract). The interface showcases image upload functionality, predicted class, and a visual representation of confidence scores.
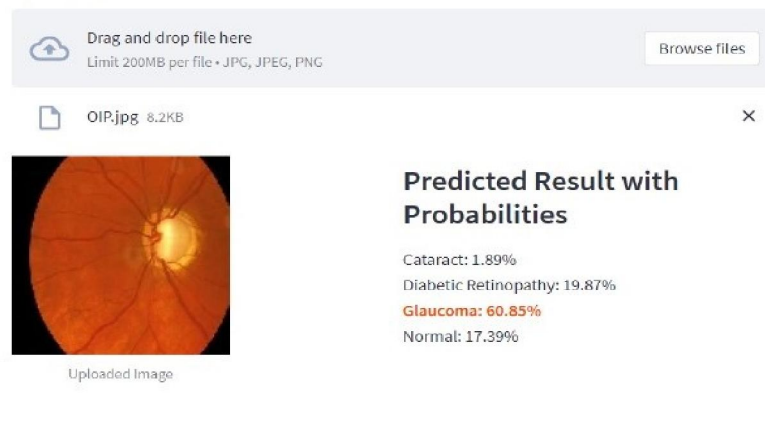
Result: Diabetic Retinopathy

Figure 13: Screenshot of the Streamlit Interface Displaying Model Output (Diabetic Retinopathy).The interface showcases image upload functionality, predicted class, and a visual representation of confidence scores.

Result: Glaucoma

Figure 14: Screenshot of the Streamlit Interface Displaying Model Output (glaucoma).The interface showcases image upload functionality, predicted class, and a visual representation of confidence scores

7.5 Batch Evaluation and Field Readiness

In addition to real time classification, the platform supports batch evaluation, prediction logging, and even on the fly image preprocessing. Its lightweight, browser based design ensures it can operate with minimal hardware resources, making it ideal for mobile eye camps, rural screening units, and low infrastructure clinics. The tool doesn't just perform well it's built for the field [17], [20], [22], [24].

7.6 Comparative Discussion

When compared to earlier systems such as hybrid models combining SVM with ANN, or conventional CNN shown in [Table 9] only approaches, this system demonstrates the superiority across multiple dimensions [16], [17], [20]. The

previous models, while effective in binary classification (especially for Glaucoma), often lacked flexibility, generalizability, and real time deployment options [16], [22]. In contrast, our system not only performs multi class classification with 100% accuracy (in three out of four classes) but also includes web based deployment, probabilistic outputs, and fast inference all of which are vital for scalability and adoption in clinical settings [17], [19], [20], [22], Moreover, by combining the strengths of both Xception (noted for its depthwise separable convolutions and spatial feature extraction) and ResNet50 (renowned for deep residual learning and stability in deep networks), the system achieves a level of performance that transcends what either model could accomplish alone [1], [17], [19], [22]. This architectural synergy, coupled with thoughtful preprocessing, strategic training, and a streamlined user interface, sets a new benchmark in AI assisted retinal screening [19], [21], [24].

Table 9: Performance Comparison Between Proposed and Baseline Models

Model	Classification Scope	Accuracy	Web Deployment	Multi-Class	AUC Score
SVM + ANN Hybrid	Glaucoma (Binary)	98.34%	No	No	N/A
CNN (ORIGA, Drishti)	Glaucoma (Binary)	89%	No	No	N/A
Proposed System	4 Diseases	**100%**	Yes (Streamlit)	Yes	**1.00**

7.7 Final Results Interpretation

The results of this study clearly position the proposed deep learning framework as a reliable, scalable, and clinically meaningful diagnostic tool. By integrating state of the art deep neural networks with an intuitive, real time web interface, the system successfully bridges the often wide gap between research prototypes and deployable healthcare applications [17], [19], [20], [22], [24]. The synergy between technical accuracy and user accessibility underscores the system's potential to make a tangible difference in ophthalmic care, particularly for early detection and proactive management of vision threatening conditions [1], [17], [19], [22].

CONCLUSION

This work presents the effective implementation of a deep learning model for retinal disease detection capable of identifying four common retinal conditions Cataract, Glaucoma, Diabetic Retinopathy, and Normal from high resolution fundus images. Leveraging the power of transfer learning with Xception and ResNet50 architectures, the system achieved near perfect performance across all evaluation metrics, including 100% validation accuracy, strong class wise F1 scores, and ideal AUC scores for each disease class [1], [17], [19], [20], [22]. The integration of robust preprocessing, fine tuned architecture, and effective training strategies played a crucial role in this performance [19], [21], [22].

Importantly, this work extends beyond algorithmic innovation to practical usability. The incorporation of the trained models into a responsive, Streamlit based web interface brings real world feasibility to the forefront [20], [22], [24]. With instantaneous predictions, probability based outputs, and a user friendly layout, the tool is designed not just for data scientists but also for medical professionals, even in low resource or rural settings [17], [20], [24]. The low latency inference time, batch processing capability, and portable deployment features make it a strong candidate for inclusion in primary care screenings and public health initiatives [22], [24].

When benchmarked against earlier models such as SVM plus ANN hybrids or CNN only classifiers which often focus on binary classification and lack scalability this solution stands out as a comprehensive advancement [16], [17], [20]. It delivers superior accuracy, multi disease capability, and real time deployment, marking a significant leap forward in AI assisted diagnostics for ophthalmology [17], [19], [20],

Overall, this study demonstrates the transformative role that AI can play in tackling the global burden of preventable blindness. With its blend of high performance, transparency, accessibility, and adaptability, the proposed system

represents a practical and scalable solution that aligns with modern clinical demands and public health goals [17], [20], [22], [24]. It stands as a strong foundation for future innovations that aim to democratize quality eye care and ensure timely treatment for all, regardless of geographic or socioeconomic constraints [17], [19], [22], [23].

FUTURE SCOPE

Future developments aim to expand the system's diagnostic reach and technical capabilities to increase its clinical utility, technological sophistication, and global impact. One of the most immediate enhancements involves broadening the model's diagnostic range. By incorporating additional retinal disorders such as age related macular degeneration, retinal vein occlusion, and hypertensive retinopathy, the system can evolve into a more holistic screening solution. Given the model's modular structure, such additions can be implemented with minimal architectural changes, assuming the availability of labelled datasets [17], [21], [22].

Linking the system with EHRs can enhance prediction context and enable patient-specific tracking Linking patient history with retinal image analysis can enable more context aware predictions, supporting longitudinal patient monitoring and personalized treatment planning [17], [20], [22]. This fusion of image based diagnostics with structured medical data can pave the way for intelligent decision support tools in ophthalmology.

From a technological standpoint, adopting federated learning could significantly enhance data privacy and model generalization. This approach would allow the system to learn from decentralized datasets across different hospitals without compromising patient confidentiality, aligning well with emerging standards for ethical AI in healthcare [17], [23]. Similarly, incorporating Generative Adversarial Networks (GANs) for synthetic data generation could strengthen the model's resilience by enriching training data with rare or underrepresented pathological cases [17], [23].

Deployment wise, extending the current Streamlit interface into a mobile or edge computing application would dramatically improve accessibility, particularly in rural or remote settings where broadband and hardware infrastructure are limited [20], [24]. Real time predictions on mobile devices would empower frontline healthcare workers, even those with minimal technical training, to conduct accurate screenings. Multilingual support, voice guided navigation, and offline functionality could further enhance inclusivity and usability across diverse populations [23], [24].

Finally, establishing feedback loops and collaborative pathways with ophthalmologists will be crucial for ongoing refinement. Continuous validation in clinical environments, guided by practitioner insights and real world patient outcomes, will help adapt the system to the dynamic needs of healthcare [20], [22], [23], [24]. This collaborative and iterative approach ensures that the system remains clinically relevant, ethically grounded, and technologically advanced, thereby supporting the broader mission of equitable and universal eye health coverage [17], [20], [22], [24].

REFERENCES

[1] G. Prathibha, "Eye Disease Classification using CNN and Vision Transformers," 2024 IEEE 1st International Conference on Advances in Signal Processing, Power, Communication, and Computing(ASPCC),Bhubaneswar,India,2024. doi: 10.1109/ASPCC62191.2024.10881414

[2] P. Sharma and A. K. Sandhu, "Prediction of Diabetic Eye Disease in Type 2 Diabetes Using Hybridization of Deep Neural Network-Based Approaches," 2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit), Greater Noida, India, 2024, pp. 48-53, doi:10.1109/GlobalAISummit62156.2024.10947818.

[3] R. Denandra, A. Fariza and Y. R. Prayogi, "Eye Disease Classification Based on Fundus Images Using Convolutional Neural Network," 2023 International Electronics Symposium (IES), Denpasar, Indonesia, 2023, pp. 563-568, doi: 10.1109/IES59143.2023.10242558.

[4] G. Kaur, N. Sharma, R. Chauhan, S. Kukreti and R. Gupta, "Eye Disease Classification Using ResNet-18 Deep Learning Architecture," 2023 2nd International Conference on Futuristic Technologies (INCOFT), Belagavi, Karnataka, India, 2023, pp. 1-5 doi: 10.1109/INCOFT60753.2023.10425690.

[5] K. Prakash, M. Sudharsan and M. S. Nidhya, "Eye Disease Glaucoma Detection Using Hybrid Classification Model Using Machine Learning Principles," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 1222-1230, doi: 10.1109/ICECA58529.2023.10395267.

[6] M. Ashikuzzaman, M. Rahman, M. A. Emran, M. Kornel Ahmed Rahul and S. Abedin, "Eye Disease Detection And Classification Using Deep Learning: A Comperative Study," 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), Cox's Bazar, Bangladesh, 2024, pp. 1-5,doi: 10.1109/COMPAS60761.2024.10796599.

[7] G. Gutte, B. Khaire, V. Harne, R. Shamalik and S. Chippalkatti, "Detection of Glaucoma Eye Disease Using Deep Learning," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023, pp. 257-260, doi: 10.1109/SIST58284.2023.10223519.

[8] P. Kumar, S. Bhandari and V. Dutt, "Pre-Trained Deep Learning-Based Approaches for Eye Disease Detection," 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2023, pp. 1286-1290, doi: 10.1109/ICCPCT58313.2023.10245175.

[9] I. Y.Abushawish, S.Modak, E.Abdel-Raheem, S.A.Mahmoud and A.Jaafar Hussain,"Deep Learning in Automatic Diabetic Retinopathy Detection and Grading Systems: A Comprehensive Survey and Comparison of Methods," in IEEE Access, vol. 12, pp. 84785-84802, 2024,doi: 10.1109/ACCESS.2024.3415617.

[10] B.Şener and E.Sümer, "Classification of Eye Disease from Retinal Images Using Deep Learning," 2023 14th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkiye, 2023, pp. 1-4, doi: 10.1109/ELECO60389.2023.10416049.

[11] 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," 2024 4th International Conference on Sustainable Expert Systems (ICSES), Kaski, Nepal, 2024, pp.1164-1170, doi: 10.1109/ICSES63445.2024.10763329.

[12] S. Lakhera and A. Garg, "Retinal Fundus Image Classification Using Hybrid Deep Learning Model," 2023 World Conference on Communication & Computing (WCONF), RAIPUR, India, 2023, pp. 1-5, doi: 10.1109/WCONF58270.2023.10235233.

[13] R. K. Yadav, G. N., K. K. Brar, S. Yadav, M. M. Adnan and S. K. Joshi, "A Hybrid Approach for Skin Cancer Detection and Classification using DenseNet121," 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), New Delhi, India, 2024, pp. 1-4, doi: 10.1109/DELCON64804.2024.10866536.

[14] R. E. Varghese and I. A. Pandian, "Inception-Resnet V2 Based Eye Disease Classification Using Retinal Images," 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, 2023, pp. 1-5, doi:10.1109/ICMNWC60182.2023.10435893.

[15] M. Giat, S. Hadi and Akmal, "Convolutional Neural Network In Human Eye Disease Classification," 2024 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, 2024, pp. 348-353,doi:10.1109/ICITSI65188.2024.10929317.

[16] K. Prakash, M. Sudharsan and M. S. Nidhya, "Eye Disease Glaucoma - Detection Using Hybrid Classification Model Using Machine Learning Principles," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 1222-1230, doi: 10.1109/ICECA58529.2023.10395267.

[17] I. Y. Abushawish, S. Modak, E. Abdel-Raheem, S. A. Mahmoud and A. Jaafar Hussain, "Deep Learning in Automatic Diabetic Retinopathy Detection and Grading Systems: A Comprehensive Survey and Comparison of Methods," in IEEE Access, vol. 12, pp. 84785-84802, 2024, doi: 10.1109/ACCESS.2024.3415617.

[18] A. Abraham and M. H. K. Mozaffari, "Detection and Diabetic Retinopathy Grading Using Digital Retinal Images," 2020 6th International Conference On Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 524-528, doi: 10.1109/ICACCS48705.2020.9074297.

[19] T. Hassan, S. A. Ali and M. M. Rahman, "Detection of Diabetic Retinopathy from Retinal Images Using DenseNet Models," 2022 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 2022, pp. 1-6, doi: 10.1109/ECCE54420.2022.9748045.

[20] G. Gutte, B. Khaire, V. Harne, R. Shamalik and S. Chippalkatti, "Detection of Glaucoma Eye Disease Using Deep Learning," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023, pp. 257-260, doi: 10.1109/SIST58284.2023.10223519.

[21] P. Rajalakshmi, A. M. Rupa and S. J. Sudha, "Diabetic Retinopathy Detection using Image Processing," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-5, doi:10.1109/ICCCI48352.2020.9104193.

[22] S. Singh and S. Thakur, "Simple Methods for the Lesion Detection and Severity Grading of Diabetic Retinopathy," 2023 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 2023, pp. 1-6, doi: 10.1109/ISCV59264.2023.10191742.

[23] M. Sheikh and D. Patil, "Skin Cancer Detection Using Machine Learning and Deep Learning," 2022 International Conference on Intelligent Technologies (CONIT), Hubballi, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9847896.

[24] S. Shrestha, R. Shah and A. Joshi, "Skin Cancer Detection by Using Deep Learning Approach," 2023 IEEE International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 473-478, doi: 10.1109/ICCCIS57415.2023.10199211.