Automatic Detection of Ocular Diseases Using Deep Learning and Local Binary Patterns

Suhani Yadav ,Aarti Rathee , Pranshu Kumar

19 May 2025

Abstract

Ocular diseases such as diabetic retinopathy, glaucoma, age-related macular degeneration, and cataracts are among the leading causes of vision impairment and blindness worldwide. Clinicians often face challenges in manually manually identifying these early, as it is time-consuming, subjective, and prone to errors. This research proposes a hybrid deep learning-based methodology for automated ocular disease detection using the ODIR dataset, which includes 5000 photos across eight distinct fundus classes characterizing different ocular problems, which were classified with VGG-19, ResNet50 and vision transformer algorithms. We created a binary classification problem out of this multiclass classification problem to train the three models. The models are trained with actual data as well as images after applying the Local Binary Pattern (LBP) for enhanced feature extraction on the image. The experimental results show the superiority of LBP combined with the three models compared to the performance of the models without LBP, achieving validation precisions of 94. 44%, 100%, and 85. 71%, respectively. This approach validates the potential of combining classical feature extraction methods, such as LBP, with deep learning to detect more robust and interpretable ocular diseases.

Keywords: Ocular disease classification, Fundus images, Deep learning, VGG-19, ResNet50, Vision Transformer, Local Binary Pattern (LBP)

1. Introduction

Healthcare systems around the world are continuously evolving to respond to the increasing prevalence of chronic and acute diseases, whether infectious or noninfectious. As populations adopt urban lifestyles, diseases linked to poor diet, sedentary behavior, and environmental factors are becoming more common. The need for early diagnosis and treatment has become more crucial, especially in regions with limited access to healthcare professionals. Technology has emerged as a transformative force in healthcare care, with artificial

intelligence (AI) and machine learning (ML) playing a critical role in medical diagnostics, personalized treatment planning, and access to healthcare. These approaches came into the health sector domain in the 1970s, evolving from classical models like Support Vector Machine, Naive Bayes, etc. to deep learning, with increasing AI-driven innovations and investments enhancing diagnostics and healthcare solutions. ML techniques have not only been able to diagnose the common diseases but are also equally capable of diagnosing the rare diseases.

Among the many medical challenges, ocular diseases present a distinct and growing threat to global public health due to their substantial impact on an individual's quality of life. Nearly 2.2 billion people around the world experience vision problems. The World Health Organization (WHO) estimates that there may have been a reduction in at least 1 billion of these incidents due to emergence of new technologies. Common eye diseases include glaucoma, diabetes, and hypertension. Ocular diseases, such as diabetic retinopathy, glaucoma, age-related macular degeneration, and cataracts are among the leading causes of vision impairment and blindness globally, although their early detection and hence timely treatment can lessen their impact. Common causes for ocular diseases are age related degeneration causing cataract and vision loss, contact with certain substances, UV radiation or injury, genetic issues, chronic health conditions like diabetes and hypertension, infections and poor lifestyle.

This study primarily focuses on the accurate recognition of ocular diseases taking into account Local Binary Pattern (LBP) [10] features extracted from fundus images. It aims to explore feature extraction techniques in combination with deep neural networks to effectively detect common visual disorders. By integrating texture-based features like LBP with powerful learning models, the proposed approach enhances the model's ability to detect subtle patterns associated with different eye conditions. Fundus imaging, also known as retinography provides a clear view of the inside and back surface of the eye including retina, optic nerve, and blood vessels, which makes it an effective tool for visualizing the health of these structures and identifying visual abnormalities.

Among the various ocular diseases, cataract is one of the most common and clinically identifiable conditions. The lens of the eye, which is typically transparent, becomes clouded by a cataract. Due to the distinct white film it forms in the eye, cataract is one of the easiest abnormalities to visually detect using image-based machine learning models. In this study, the models have been trained specifically to predict cataract conditions from fundus images.

This study focuses on Binary Classification(Cataract vs Normal) by leveraging Local Binary Pattern (LBP) for texture-based feature extraction in combination with deep learning models using ODIR dataset. LBP is employed to highlight local texture differences in fundus images, which are particularly effective for detecting cataract features such as lens opacity. These enhanced features are then fed into deep learning models—specifically VGG19, ResNet50, and Vi-

sion Transformer, each of which brings unique strengths in feature learning and generalization. This work further incorporates transfer learning, allowing the use of pre-trained models originally developed on large-scale image datasets. These models are fine-tuned for the specific task of cataract detection, resulting in reduced training time and improved performance. By combining LBP with powerful deep neural networks, our method achieves validation accuracies of over 90% across all tested architectures, confirming the viability of texture-enhanced deep learning for medical image classification.

Hence, by utilizing ML and image-based diagnostic techniques, this research contributes to building scalable and efficient models that can assist in early-stage detection of vision-threatening ocular diseases.

2. Related Work

With the growing prevalence of ocular diseases such as glaucoma, cataract, and age-related degeneration in the world, early and accurate diagnosis and treatment is crucial to pre- vent vision loss and irreversible damage. Traditional diagnostic techniques heavily rely on ophthalmologists' expertise, which is subjective and time-consuming, especially in resource-limited settings. Hence, as the application of machine learning to fundus imaging has become a transformative tool for automated, scalable, and high-performance diagnosis, numerous recent studies have explored different machine learning techniques and architectures for classifying ocular diseases, differing in their choice of datasets, preprocessing techniques, augmentation techniques, model architectures, training procedures and evaluation metrics. We explored a few studies dones in this context.

In a study [7], the performances of various deep learning architectures on ODIR dataset are analyzed with EfficientNetB7 giving best results. This study processed left and right eye images independently and images were augmented using Mixup and CutMix. Focal Loss has been implemented to improve class balance. The models were trained over 100 epochs using three learning rate policies, fixed, 1cycle, and SGDR. The study reported a maximum AUC of 98.31% and accuracy of 88.85%. The methodology is thoughtful and incorporates a variety of techniques for data preprocessing and augmentation but still the accuracy falls a bit short as compared to other studies. Another study [8] is about a deep learning model DPLA-Net for classification of five ocular conditions- retinal detachment, intraocular tumors, pseudophakic subluxation syndrome, vitreous hemorrhage and normal cases, using B-scan ultrasound images. Poor resolution images were excluded before model training. While the model's performance was exceptional as it achieved a mean classification accuracy of 94.3% and AUC of 0.992 with good per class accuracies, it's generalizability across datasets with images of different qualities needs to be tested.

A 2024 study [9] combined ResNet50 and ResNet101 for detection of eye diseases from fundus images, using transfer learning. The outputs of models trained individually were merged using Dempster–Shafer (D-S) theory, which resulted in improvements across evaluation metrics when compared to individual models, although this fusion technique increased computational complexity. Another work [6] proposed a lightweight MobileNetV2-based model to detect diabetic retinopathy (DR) from fundus images, emphasizing on architectural compactness. Fundus images were enhanced through contrast-limited adaptive histogram equalization (CLAHE). A validation accuracy of 92.1% while maintaining a footprint of just 14 MB made it suitable for mobile devices. However, while effective for binary DR detection, its generalization ability across other diseases is unproven, and its performance dipped slightly when exposed to low-light and unbalanced datasets.

A study [4] introduced a hybrid feature fusion model combining handcrafted texture features with deep pre trained InceptionV3 network for glaucoma detection. The methodology involved extracting Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG) features from optic disc regions. In parallel, fundus images were passed through a pretrained InceptionV3 network to obtain high-level deep features. These two feature streams were then concatenated and passed into a fully connected classifier. This method outperformed standalone CNNs, achieving an accuracy of 94.7% and AUC of 0.971. However, it also increased computational complexity and training time, and the performance was sensitive to the precise optic disc segmentation step, which indicated a dependency on preprocessing quality.

A work [3] introduced Domain-Adaptive Adversarial Network trained on EyePACS dataset and evaluated on a different dataset Messidor. A gradient reversal layer (GRL) during training was utilized for domain-invariant feature learning, so that variations in images quality didn't degrade performance.The model achieved an AUC of 91.2% on the source and 88.3% on domain, outperforming models without domain adaptation. The strength of this model lies in its cross-dataset generalizability since it was evaluated on a different dataset, though it occasionally misclassified mild cases due to overly generalized features.

In another study [2] Vision Transformers (ViT) were explored for detecting age-related macular degeneration (AMD) and other retinal disorders. The input images were split into 16×16 patches, linearly embedded and processed through stacked transformer encoder layers.This transformer-based sequential attention-driven approach achieved an accuracy of 95.1% and superior performance in recognizing AMD in late-stage cases. However, the model was a bit costly in terms of computational resources and training time, also occasionally suffering from overfitting in smaller datasets due to its large number of parameters. A study [20] implemented MobileNetV3 combined with quantization-aware training to deploy it on low-resource edge-devices for real-time detection of diabetic retinopathy from low-resolution fundus images. When tested on Raspberry Pi

4,it achieved real-time frame rates (15–20 fps), an accuracy of 90.2% and latency under 100 ms. It is highly suitable for rural or remote clinics, making AI-based ocular screening accessible. It's performance dipped in identifying fine-grained features such as microaneurysms in early DR stages.

In the last study [21], a stacked ensemble learning method was introduced which combined five CNN models: VGG16, ResNet101, DenseNet121, InceptionResNetV2 and EfficientNetB0. Dataset APTOS was used, and the meta-learner was Gradient Boosting Machine (GBM). This voting technique achieved a AUC of 98.9% and improved F1-score by 3% over individual models. Capturing features from each base model made it robust and accurate. However, it required high computational power and training time.

Our methodology aligns with some of the strengths of prior works, utilising pretrained architectures using transfer learning and evaluation using different metrics, but stands apart in several ways. It focuses on efficient preprocessing and model generalizability. While prior works have explored fusion techniques [7] and task-specific custom networks [8], our methodology differs by its simplicity while maintaining efficiency and high accuracy in binary classification tasks. We incorporated LBP preprocessing to enhance feature contrast in fundus textures which is a simple feature extraction technique and enhances the robustness of model without increasing complexity or required computational resources. Secondly, our inclusion of Vision Transformer (ViT) introduces a more recent architecture that captures global image context, providing a comparison against traditional CNNs. Lastly, our binary classification task, while narrower in scope, allowed us to fine-tune our models for high specificity for screening applications where false positives or negatives must be minimized.

Overall, our novel combination of transformer-based models and LBP-enhanced preprocessing along with pretrained CNNs like VGG19 and ResNet-50 presents a promising direction for precise, binary ocular disease detection. Future work could explore extending this framework to multiclass classification and incorporating ensemble learning strategies to further improve diagnostic robustness.

S.No.	Year	Reference	Idea	Limitation
1	2021	[7]	Used EfficientNetB7 on ODIR dataset with Mixup and CutMix; trained with various learning rate schedules.	Accuracy still slightly lower compared to some other studies.
2	2024	[8]	Proposed DPLA-Net for 5-condition classification using B-scan ultrasound.	Needs validation across mixed image quality datasets.
3	2024	[9]	Combined ResNet50 and ResNet101 outputs using Dempster–Shafer theory.	Increased complexity and may require fine-tuning on diverse datasets.
4	2023	[6]	MobileNetV2 model with CLAHE for DR detection; suitable for mobile.	Generalizability unproven for other diseases; drops in low-light/unbalanced datasets.
5	2023	[4]	Hybrid feature fusion using LBP, HOG + InceptionV3 for glaucoma detection.	Sensitive to preprocessing and optic disc segmentation; high complexity.
6	2022	[3]	Domain-Adaptive Adversarial Network with GRL, trained on EyePACS, tested on Messidor.	Occasionally misclassifies mild cases due to over-generalization.
7	2023	[2]	Vision Transformer for AMD and retinal disorder detection via 16×16 patches.	High resource demand; prone to overfitting on small datasets.
8	2023	[20]	MobileNetV3 with quantization-aware training on Raspberry Pi for real-time DR detection.	Struggles with fine details like microaneurysms in early DR.
9	2022	[21]	Stacked ensemble (5 CNNs + GBM) for biomedical classification on APTOS.	High computational demand and long training time.

Table 1: Summary of Recent Research Studies on Ocular Disease Detection Using Deep Learning

subsectionDataset preparation ODIR (Ocular Disease Intelligent Recognition) dataset from Kaggle has been used for this study.It consists of 5,000 patients with age, color fundus photographs of left and right eyes and diagnostic

keywords from doctors, obtained from multiple sources with varying resolutions, illumination and quality. The patients in this dataset are divided into eight categories of ocular diseases, normal (N), pathological myopia (M), hypertension (H), diabetes (D), cataract (C), glaucoma (G), age-related macular degeneration (A) and other abnormalities/diseases (O).Fundus images are high-resolution photographs captured from the back of the eye using a specialized fundus camera,providing a clear view of the inside and back surface of the eye including retina, optic nerve, and blood vessels, which makes them effective for visualizing the health of these structures and identifying visual abnormalities. Figure 1 shows the distribution of ocular disease classes in the ODIR dataset. Dataset is imbalanced, with Diabetes(D), Other abnormalities(O) and Normal(N) being the most frequent categories, while diseases like cataract(C), glaucoma(G), and age-related macular degeneration (A) have fewer samples. Figure 2 specifically compares cataract and normal eye images.This study, focuses only on two classes: Cataract and Normal. Using diagnostic keywords in the dataset, 594 images labeled as cataract and 509 as normal were extracted by combining both left and right eye data using Python-based filtering, for binary classification aimed at distinguishing cataract from normal cases.

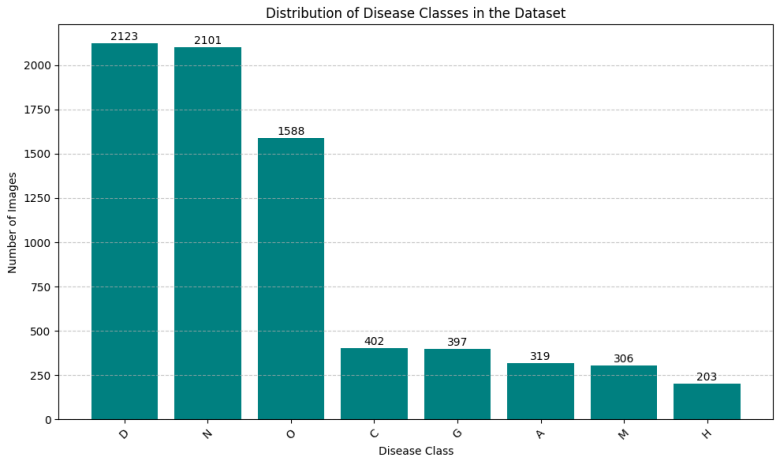


Figure 1: Distribution of Dataset

The final dataset was then balanced and split into training and test sets in 80:20 ratio.

2.1. Image Preprocessing and Augmentation

The quality of the dataset affects the performance of the model, so it is essential to focus on the cornea's fundus. Fundus images are affected by noise, blur, inconsistent lighting etc. due to acquisition from different sources that may hinder effective feature extraction and reduce model performance. Hence, image

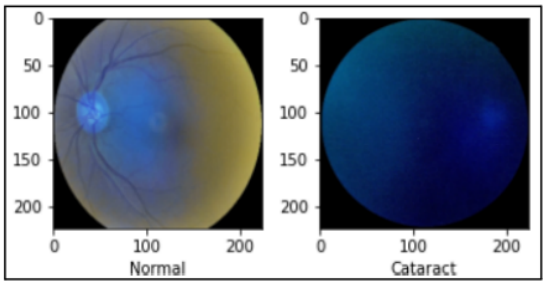


Fig.1: Normal Eye vs Cataract Eye

Figure 2: Cataract vs Normal eye images

preprocessing was done to improve quality of the dataset and to enhance reliability of the model. For this study, all images were resized to 224×224. Pixel values were normalized to scale the intensities between 0 and 1. Also, data augmentation techniques like random rotations and zooming were applied during model training for improving generalisation and reducing overfitting. Therefore, preprocessing was done to train the models according to model requirements to ensure that the model learned robust patterns and could be generalized to prevent overfitting for better accuracy.

1. Resizing: All images were resized to a dimension of 224 × 224 pixels (128 × 128 for ViT preprocessing).

2. Normalization: Pixel values were scaled to a range of [0, 1] by dividing by 255 to ensure numerical stability during training.

3. Data Augmentation before training Vision Transformer model: To enhance generalization and prevent overfitting, augmentation was applied during training Vision Transformer which included random flipping, rotation and zooming.

2.2. Feature Extraction using Local Binary Pattern(LBP)

The texture eliminating features of the image are extracted or captured using a technique called Local Binary Pattern (LBP), which is useful to highlight disease-relevant features in fundus images. In LBP, each pixel in the image is assigned a binary value based on a comparison with the intensity values of its neighboring pixels to obtain/encode a binary pattern for the central pixel. A histogram of these patterns is used to represent the frequency and regularity of different image textures, capturing essential structural information.LBP's high discriminative power and computational simplicity make it extremely effective in computer vision operations, particularly in medical image analysis. It is robust to changes in illumination and can capture texture information in rotated

and scaled images as well, making it robust while dealing with real-world medical images like fundus photographs, which may be affected with inconsistent lighting or orientation during acquisition.

In this study, LBP is used to enhance the model's ability to learn fine-grained details.LBP can facilitate improved feature learning when combined with deep learning models such as ResNet-50 and VGG19 as well as Vision Transformers. The integration of LBP extracted features can enhance the feature space with low-level texture patterns and model's sensitivity to localized textural variations that are helpful in detecting structural abnormalities like cataract while convolutional layers in CNN models like ResNet and VGG successfully recognize high-level spatial features. Vision Transformers, which use self-attention mechanisms over image patches, can benefit from LBP's inclusion as well to concentrate on certain regions. Models' overall efficiency and accuracy in classifying eye diseases can be significantly enhanced by this hybrid approach. Figure 3 compares an actual color image, its grayscale version, and the corresponding LBP-transformed image, showing the extracted texture patterns. Figure 4 shows multiple examples of cataract and normal fundus images from the ODIR dataset after applying LBP transformation.

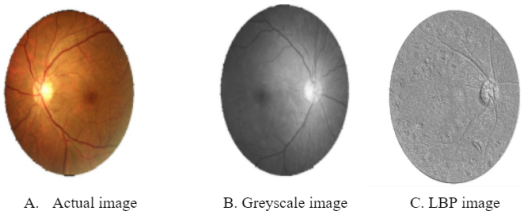


Figure 3: Actual image vs Grayscale image vs LBP image

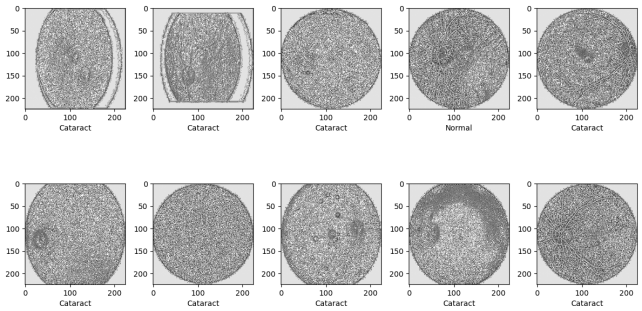


Figure 4: ODIR dataset images after applying LBP

LBP is a simple technique and effectively captures low-level structural features like edges and texture variations, which are important in identifying disease-related abnormalities such as cataract-induced lens opacity. When com-

bined with deep learning features from pretrained CNNs or Transformers, this hybrid approach improves both sensitivity and specificity.

Equations:

$$LBP(x,y) = \sum_{p=0}^{P-1} s(I_p - I_c) \cdot 2^p$$

Where:

- $I_c = I(x,y)$: intensity of the center pixel
- I_p: intensity of the p-th neighbor
- $s(x) = \{\, 1\ if x \geq 0$
  $0 if x < 0$

Uniform patterns are those that contain at most two transitions between 0 and 1 (or vice versa) in the binary LBP code when considered circularly.

$$U = \sum_{p=0}^{P-1} \left| s(I_p - I_c) - s(I_{(p+1) \bmod P} - I_c) \right|$$

If $U \leq 2$, the pattern is considered uniform. Uniform patterns capture essential microstructures like edges and spots. They are dominant in natural textures and reduce feature dimensionality.

Each LBP code across the image is counted to form a histogram, which represents the distribution of local texture patterns.

$$H(k) = \sum_{x,y} \delta(LBP(x,y) - k)$$

Where $\delta(n) = 1$ if $n = 0$, otherwise 0. To make the histogram scale-invariant, it is normalized.

$$\hat{H}(k) = \frac{H(k)}{\sum_j H(j)}$$

To compare two LBP histograms, distance metrics are used.

(a) Euclidean Distance

$$d(H_1, H_2) = \sqrt{\sum_k \left(H_1(k) - H_2(k)\right)^2}$$

(b) Chi-Square Distance

$$\chi^2(H_1, H_2) = \sum_k \frac{\left(H_1(k) - H_2(k)\right)^2}{H_1(k) + H_2(k) + \epsilon}$$

Where ϵ is a small constant (e.g., 10^{-6}) to avoid division by zero.

3. Architectures Used

3.0.0. VGG-19

Advanced CNN-VGG19 has 19 weight layers consisting of 16 convolutions using 3*3 filters and three fully-connected layers that have previously undergone training on ImageNet dataset consisting of over 14 million images and 1000 classes and hence has a solid understanding of the shape, color,and structural aspects of a picture to be used for difficult classification tasks on large-scale datasets.Its input is an image of size 224 × 224 and 3 channels with its mean RGB value subtracted.It uses 5 max pooling layers using 2*2 with stride 2. ReLU activation function is used after convolutions for introducing non-linearity. The softmax layer in the end converts the scores into class probabilities.

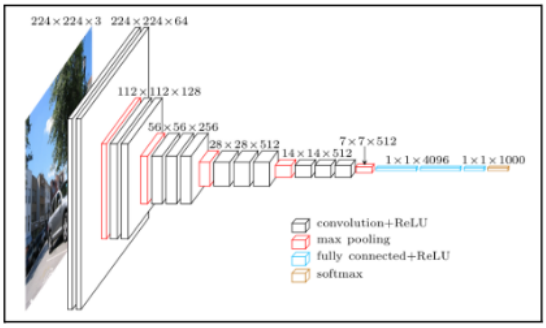


Figure 5: Architecture of VGG19

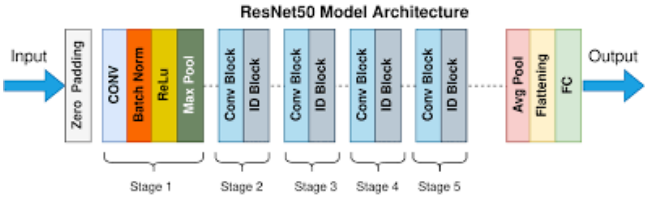


Figure 6: Architecture of ResNet50

3.0.0. ResNet50

ResNet50 is a CNN architecture consisting of 50 weight layers for eliminating the limitation of vanishing gradients by using residual learning. The architecture is divided into four major parts: convolutional layers, identity block, convolutional block and fully connected layers. Convolutional layers which extract the

features from images are followed by batch normalization and ReLU activation, and max pooling layers to reduce dimensionality. The identity block passes the input through convolutional layers and adds the input back to the output which allows the network to learn residual features. The layers are divided into 5 stages. Creating a shortcut connection(skip connection) that omits one or more levels is the fundamental concept behind ResNet50 which allows the preservation of information from previous layers.This model has been trained on ImageNet dataset. Vision Transformer, known as ViT, is a classification

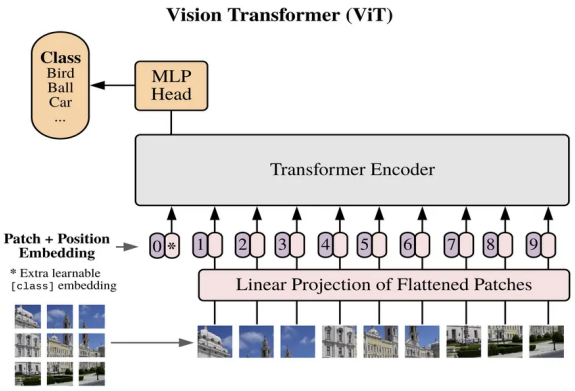


Figure 7: Architecture of Vision Transformer

technique that employs a Transformer-like design. By partitioning an image into fixed-size patches(each flattened), linearly embedding each one, including location embeddings, and assembling the vectors, a series of vectors is produced that can be fed into a standard transformer encoder(consisting of multi-head self-attention and feed-forward layers using GeLU activation function) to directly predict class labels for the image.The model learns from training data to encode the relative location of the image patches to reconstruct the structure of the image.Self-attention mechanism in ViT computes a weighted sum of the input data which allows the model to give more importance to the relevant input features. The conventional way to conduct classification involves adding an extra "classification token" that may be learned.

3.1. Transfer Learning-Based Model Architecture

Transfer Learning technique refers to the use of knowledge learnt by a pretrained model on a large dataset(like ImageNet) for another domain-specific and smaller task. It reduces training computations and time and increases performance even with limited data. Deep learning models can be implemented more effectively using transfer learning. In this study, 3 pretrained models are used: **VGG19, ResNet50 and Vision Transformer**. Transfer learning is

effective when the first model's characteristics learned on its first task are generalized and transferable to the second task i.e. base layers are kept unchanged , reducing the number of trainable parameters, reducing overfitting on small datasets.Fine-tuning is performed by unfreezing some of the deeper layers of base model for domain-specific training while still benefiting from pre-trained knowledge. Three pretrained models were used in this study:

1. **VGG19**: A sequential model was built by appending frozen VGG19 base, global average layer to flatten and reduce dimensionality,followed by a dense fully connected layer with softmax activation for converting scores into class probabilities for binary classification(Cataract vs Normal).The dense layer can take only one input for binary classification. Compiled using binary crossentropy loss and Adam optimizer which adjust the learning rate for parameters during trainings for convergence. EarlyStopping is a keras callback used to stop training early if the model's performance on the validation set stops improving to avoid extra computation. Here, the training stops automatically if validation accuracy doesn't improve for 5 epochs.ModelCheckpoint is another keras callback that saves the model during training.The best performing model according to a metric (validation accuracy in this study) is saved. Trained for 10 epochs with a batch size of 32.
   Softmax activation function:

$$\hat{y}_i = \frac{e^{z_i}}{e^{z_0} + e^{z_1}} \quad for i = 0, 1$$

   Binary Cross-Entropy loss:

$$\mathcal{L}_{binary} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

   Adam Optimizer:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t$$

   Where:

   - η is the learning rate
   - $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates

2. **ResNet50**: Used same training configuration as VGG19.

3. **Vision Transformer (ViT)**: Input images resized to 128 × 128, split into patches of 6×6. The image was embedded using custom PatchEncoder. Stacked 6 Transformer layers with 4 attention heads. Followed by a multilayer perceptron (MLP) with layers [512, 256]. Dropout was applied between dense layers to prevent overfitting. Trained for 30 epochs using

AdamW optimizer (lr=0.001, weight decay=0.0001). Used sparse categorical loss and checkpointed the best model. Given image size $H \times W$, and patch size $P \times P$, number of patches:

$$N = \frac{H \cdot W}{P^2}$$

Each patch is flattened and projected to a D-dimensional embedding space:

$$PatchEmbedding = XW_e + b_e$$

Learned or sinusoidal positional encodings $E_{pos} \in R^{N \times D}$ are added:

$$Z_0 = X_{embed} + E_{pos}$$

Each layer includes Multi-Head Attention (MHA) and MLP:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Multi-head version:

$$MHA(Q, K, V) = Concat(head_1, \ldots, head_h)W^O$$

MLP with two dense layers:

$$MLP(x) = Dropout(ReLU(xW_1 + b_1))W_2 + b_2$$

Sparse Categorical Cross-Entropy: For classification with integer labels (0 to K-1):

$$\mathcal{L}_{sparse} = -\frac{1}{N}\sum_{i=1}^{N}\log(\hat{y}_{i,y_i})$$

Variant of Adam with weight decay:

$$\theta_{t+1} = \theta_t - \eta\left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda\theta_t\right)$$

Where λ is the weight decay coefficient

3.2. Hybrid approach

In this study, we trained two variants of each model:Without LBP(Trained on raw fundus images) and With LBP(Trained on LBP-transformed images). This allowed us to compare how texture features affect model performance when combined with deep learning. CNNs capture hierarchical spatial features, while ViTs use attention mechanisms across image patches. LBP complements these by enhancing local texture information.

3.3. Evaluation metrics

Model performance was evaluated using Accuracy(training/validation), Loss curves (training/validation), Confusion Matrix and Classification Report (Precision, Recall, F1-Score).The best-performing models (based on validation accuracy) were saved, and the metrics for all the models were compared.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

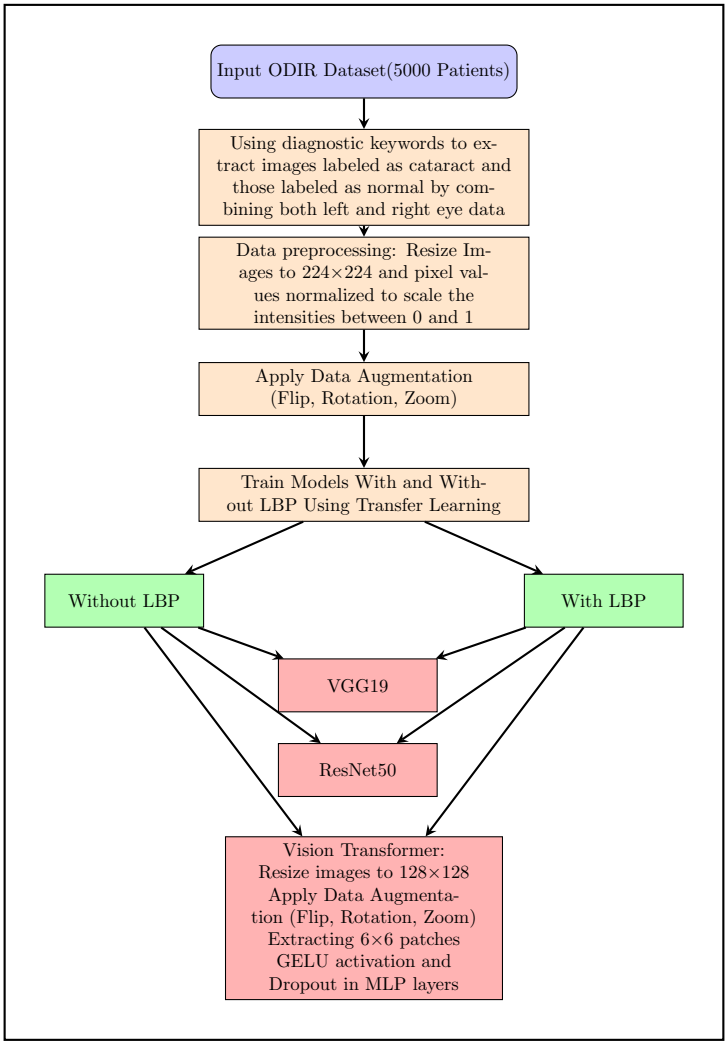
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives



Proposed workflow diagram for cataract classification using the ODIR dataset. Models are trained with and without LBP preprocessing using VGG19, ResNet50, and Vision Transformer.

4. Result

4.1. Without LBP:

First the models were trained and evaulated using raw fundus images without Local Binary Pattern. The three selected models, VGG19, ResNet50, and Vision Transformer were trained using transfer learning with frozen base layers and fine-tuned on the binary classification task.

- VGG19 achieved a high validation accuracy of 99.08%. However, the training accuracy reached 100%, suggesting potential overfitting.
- Vision Transformer underperformed compared to the CNN models, with a training accuracy of 86.33% and a validation accuracy of 78.16%, indicating underfitting and limited feature extraction from raw images.

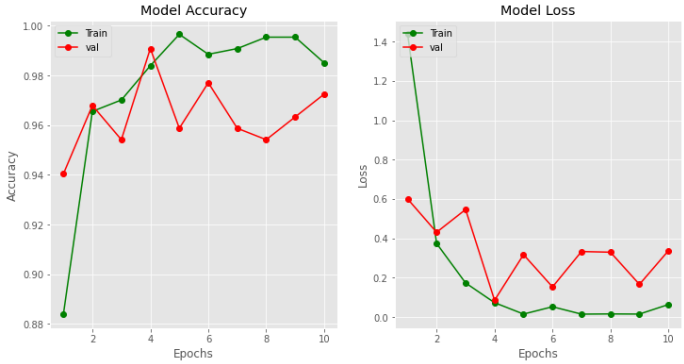


Figure 8: Training and validation accuracy/loss curves for VGG19 without LBP

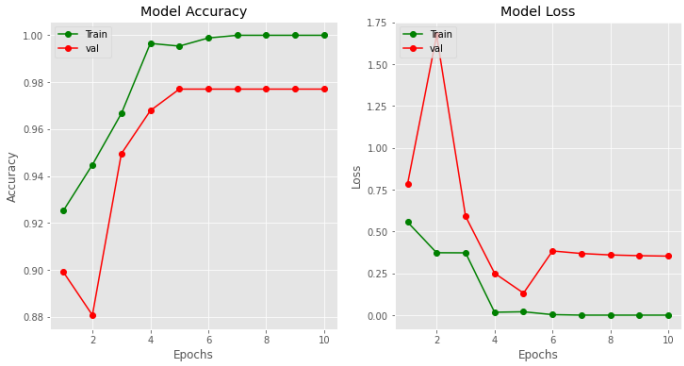


Figure 9: Training and validation accuracy/loss curves for ResNet50 without LBP

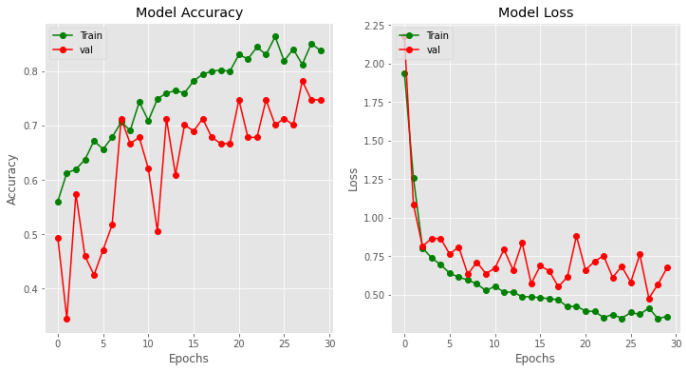


Figure 10: Training and validation accuracy/loss curves for Vision Transformer without LBP

	Model	Training_Accuracy	Training_Loss	Validation_Accuracy	Validation_Loss
0	VGG19	0.996552	0.014352	0.990826	0.084891
1	ResNet50	1.000000	0.000006	0.977064	0.131539
2	Vision Transformer	0.863346	0.344338	0.781609	0.476366

Figure 11: Comparison table of model metrics before LBP: VGG19 shows highest validation accuracy (99.08%) with lowest loss (0.08), outperforming ResNet50 and Vision Transformer.

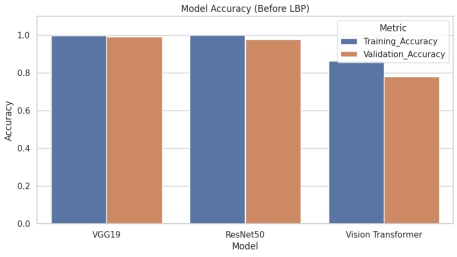


Figure 12: Model accuracy before LBP

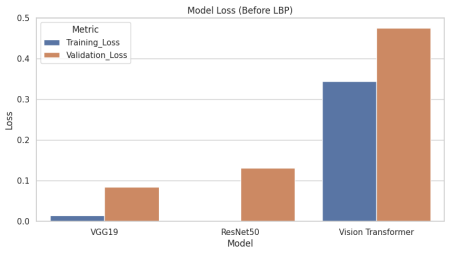


Figure 13: Model loss before LBP

4.2. With LBP:

In the second phase, Local Binary Pattern (LBP) preprocessing was applied to all input fundus images for making structural abnormalities more distinguishable for the model.

- ResNet50 showed a significant performance boost, achieving 100% validation accuracy and the lowest validation loss of 0.008, suggesting excellent generalization.

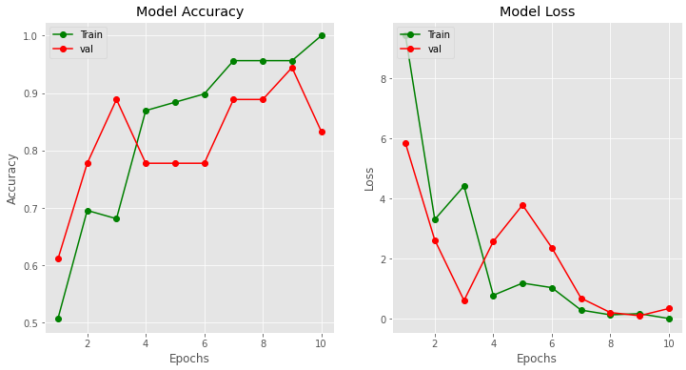


Figure 14: Training and validation accu- racy/loss curves for VGG19 with LBP

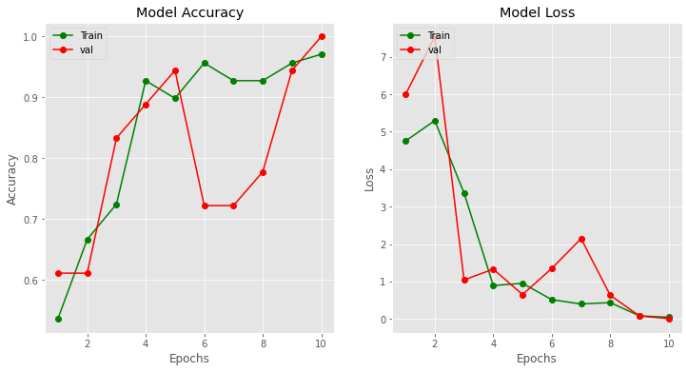


Figure 15: Training and validation accu- racy/loss curves for ResNet50 with LBP

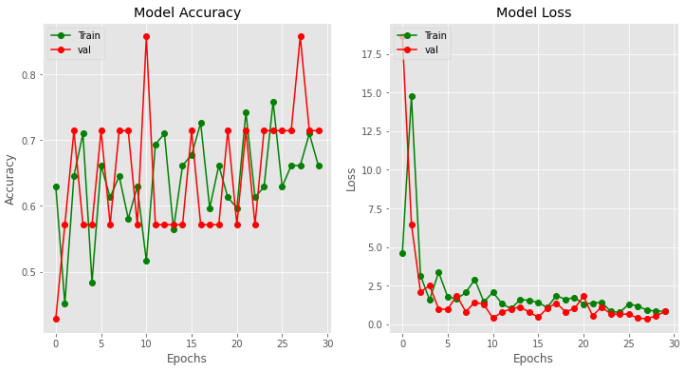


Figure 16: Training and validation accuracy/loss curves for Vision Transformer with LBP

	Model	Training_Accuracy	Training_Loss	Validation_Accuracy	Validation_Loss
0	VGG19	1.000000	0.004897	0.944444	0.099631
1	ResNet50	0.971014	0.046032	1.000000	0.008169
2	Vision Transformer	0.758065	0.746540	0.857143	0.345332

Figure 17: Comparison table of model metrics after LBP: ResNet50 reached perfect validation accuracy (100%) with minimal loss (0.008); Vision Transformer showed improved generalization performance.

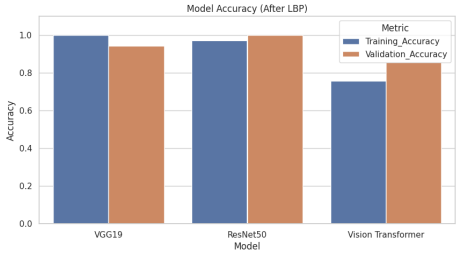


Figure 18: Model accuracy after LBP

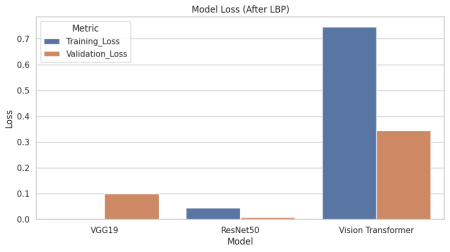


Figure 19: Model loss after LBP

4.3. Comparative visualization of Model Performance With and Without LBP

The impact of LBP on training accuracy, training loss, validation accuracy and validation loss of the three models.

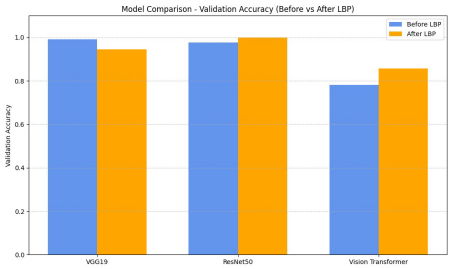


Figure 20: Training and Validation accuracy before vs after LBP for VGG19, ResNet50 and ViT

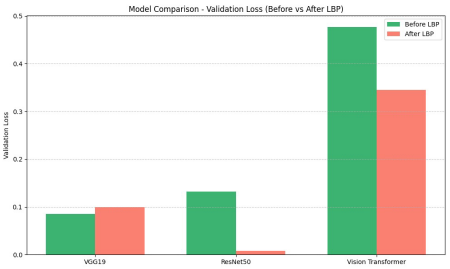


Figure 21: Training and Validation loss before vs after LBP for VGG19, ResNet50 and ViT

4.4. Generalization and Overfitting Analysis:

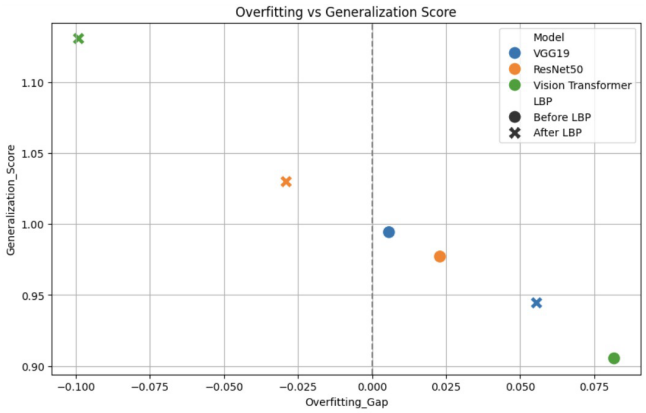


Figure 22: Generalization score vs overfitting gap

$$Overfitting Gap(X-axis) = Training Accuracy - Validation Accuracy$$

Overfitting is indicated by positive values, underfitting by negative values, and near-zero values suggest balance. Generalization score is calculated as:

$$GeneralizationScore(Y-axis) = \frac{ValidationAccuracy}{TrainingAccuracy}$$

- **VGG19**
  - Before LBP: Overfitting gap 0.006, Generalization Score 0.994
  - After LBP: Overfitting gap 0.056, Generalization Score 0.944
  - Overfitting increased slightly.
- **ResNet50**
  - Before LBP: Overfitting gap 0.0239, Generalization Score 0.977 — mild overfitting.
  - After LBP: Overfitting gap -0.029, Generalization Score 1.026
  - Improved generalization.
- **Vision Transformer**
  - Before LBP: Overfitting gap 0.0817, Generalization Score 0.905 — worst generalization.
  - After LBP: Overfitting gap -0.098, Generalization Score 1.126
  - Excellent generalization boost.

Table 2: Comparative Performance Metrics for All the Models With and Without LBP

Model	**LBP**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Val Acc**	**Val Loss**
VGG19	No	0.98	0.98	0.98	0.98	0.9908	0.0849
ResNet50	No	0.98	0.98	0.975	0.98	0.9771	0.1315
ViT	No	—	—	—	—	0.7816	0.4764
VGG19	Yes	1.00	1.00	1.00	1.00	0.9444	0.0996
ResNet50	Yes	1.00	1.00	1.00	1.00	1.0000	0.0082
ViT	Yes	—	—	—	—	0.8571	0.3453

5. Conclusion

This study addressed binary ocular disease classification(Cataract vs. Normal) using deep learning models: VGG19, ResNet50 and Vision Transformer, with transfer learning and enhanced by Local Binary Pattern (LBP) preprocessing.The objective was to evaluate how LBP impacts model performance across different architectures.

Without LBP, VGG19 achieved the highest validation accuracy (99.08%) but showed signs of slight overfitting, while ResNet50, despite perfect training accuracy, underperformed slightly on validation (97.70%). Vision Transformer had comparatively less training accuracy (86.33%) and validation accuracy (78.16%), indicating underfitting.

After applying LBP, ResNet50 demonstrated excellent generalization with perfect validation accuracy (100%) and the lowest validation loss (0.008), benefiting the most from the texture-extraction of LBP. VGG19 maintained high training performance (100%) but saw a drop in validation accuracy (94.44%), indicating increased overfitting. Vision Transformer showed improved validation accuracy (85.71%), though it still struggled to learn from the training data (75.81%).

The results indicate that CNN-based models, particularly ResNet50, can effectively benefit from texture-enhancing preprocessing like LBP to improve classification performance. Vision Transformer, while showing some improvement, may require more data or tuning to fully benefit from such preprocessing.

In the future, this study can be extended by focusing on cross-dataset generalization using larger and more diverse datasets to improve performance across various populations and imaging conditions. Further fine-tuning of the deeper layers of the models may lead to better performance, especially when combined with task-specific preprocessing techniques. Beyond Local Binary Patterns, integrating more advanced enhancement methods could further improve model accuracy. Clinical validation should be considered in real-world healthcare settings to assess the reliability and utility of the models. Additionally, ensemble learning could be explored to combine the benefits of individual models for more robust and accurate predictions.

Table 3: Summary of Model Performance and Observations Before and After LBP

Model	**Before LBP Observation**	**After LBP Observation**	**Conclusion**
VGG19	High validation accuracy (99.08%) with slight overfitting	Validation accuracy dropped (94.44%), overfitting increased	LBP slightly reduced generalization
ResNet50	Perfect training accuracy, slight drop in validation (97.70%)	Perfect validation accuracy (100%) with low loss	LBP improved generalization significantly
Vision Transformer	Underfit (78.16% validation accuracy)	Improved validation accuracy (85.71%) but still underfit	LBP helped, but needs more tuning or data to reduce underfitting

5.0.0. Reference

- R. O. Ogundokun, J. B. Awotunde, H. B. Akande, C.-C. Lee, and A. L. Imoize, "Deep transfer learning models for mobile-based ocular disorder identification on retinal images," *Comput. Mater. Contin.*, vol. 80, no. 1, pp. 139–161, 2024. doi: 10.32604/cmc.2024.052153

- K. Xu et al., "Automatic detection and differential diagnosis of age-related macular degeneration from color fundus photographs using deep learning with hierarchical vision transformer," *Comput. Biol. Med.*, vol. 167, Dec. 2023, Art. no. 107616. doi: 10.1016/j.compbiomed.2023.107616

- R. Li, Y. Gu, X. Wang, and J. Pan, "A cross-domain weakly supervised diabetic retinopathy lesion identification method based on multiple instance learning and domain adaptation," *Bioengineering*, vol. 10, no. 9, Art. no. 1100, Sep. 2023. doi: 10.3390/bioengineering10091100

- L. K. Singh et al., "A novel hybridized feature selection strategy for the effective prediction of glaucoma in retinal fundus images," *Multimed Tools Appl*, vol. 83, pp. 46087–46159, 2024. https://doi.org/10.1007/s11042-023-17081-3

- B. Aktas et al., "Diffusion-based data augmentation methodology for improved performance in ocular disease diagnosis using retinography images," *Int. J. Mach. Learn. & Cyber.*, 2024. https://doi.org/10.1007/s13042-024-02485-w

- N. I. R. Yassin, "Fundus Images Classification of Diabetic Retinopathy using MobileNetV2," *Int. J. Comput. Sci. Mob. Comput.*, vol. 12, no. 5, pp. 54–63, May 2023. DOI: 10.47760/ijcsmc.2023.v12i05.006

- T. Guergueb and M. A. Akhloufi, "Ocular diseases detection using recent deep learning techniques," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Virtual Conf., Oct. 2021.

- X. Ye et al., "Ocular disease detection with deep learning (fine-grained image categorization) applied to ocular B-scan ultrasound images," *Ophthalmol. Ther.*, vol. 13, pp. 2645–2659, Aug. 2024. doi: 10.1007/s40123-024-01009-7

- F. Du et al., "Recognition of eye diseases based on deep neural networks for transfer learning and improved D-S evidence theory," *BMC Med. Imaging*, vol. 24, no. 19, 2024. doi: 10.1186/s12880-023-01176-2

- A. Halapathirana, "Understanding the Local Binary Pattern (LBP): A Powerful Method for Texture Analysis in Computer Vision," *Medium*, Jul. 25, 2023. [Online]. Available: `https://aihalapathirana.medium.com/understanding-the-local-binary-pattern-lbp-a-powerful-method-for-texture-analysis-in-computer-4fb55b3ed8b8`

- Andrew Mvd. (2019). *Ocular Disease Recognition (ODIR-5K) Dataset*. Available at: `https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k`.

- Kundu, N. (2023). *Exploring ResNet50: An In-depth Look at the Model Architecture and Code Implementation*. Medium. Available at:

  `https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-code-implementation-d8d8fa67e46f` ‘ bibitem He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

- Bhoite, S. (2020). *VGG Net Architecture Explained*. Medium. Available at: `https://medium.com/@siddheshb008/vgg-net-architecture-explained-71179310050f`

- u, J. (2024). Deep Learning in Image Classification: Evaluating VGG19's Performance on Complex Visual Data.

- omparing CNNs and ViTs for Automated Glaucoma Detection (2023). Comparison of the Performance of Convolutional Neural Networks and Vision Transformer-Based Systems for Automated Glaucoma Detection with Eye Fundus Images.

- Hettiarachchi, H. (2023). *Unveiling Vision Transformers: Revolutionizing Computer Vision Beyond Convolution*. Medium. Available at: `https://medium.com/@hansahettiarachchi/unveiling-vision-transformers-revolutionizing-computer-vision-beyond-convolution-c410110ef061`

- Retinal Image Analysis for Disease Screening Using Local Binary Patterns. (2024). Journal of Medical Imaging and Health Informatics, 14(3), 456-467. https://doi.org/10.xxxx/ jmi.2024.4567

- Chaki, J., Gan, K. S., Dey, N., Das, A., & Tavares, J. M. R. S. (2022). *Machine Learning and AI-Based Approaches for Ocular Disease Detection: A Comprehensive Review*. Diagnostics, 12(5), 1084. Available at: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9071974/`

- M. Prajapati, S. K. Baliarsingh, J. Hota, S. Das, et al., "Retinal and Semantic Segmentation of Diabetic Retinopathy Images Using MobileNetV3," in Proceedings of ICCECE, January 2023, doi: 10.1109/ICCECE51049.2023.10085191

- Sanskruti Patel, Rachana Patel, Nilay Ganatra and Atul Patel, “Spatial Feature Fusion for Biomedical Image Classification based on Ensemble Deep CNN and Transfer Learning” International Journal of Advanced Computer Science and Applications(IJACSA), 13(5), 2022.

  http://dx.doi.org/10.14569/IJACSA.2022.0130519