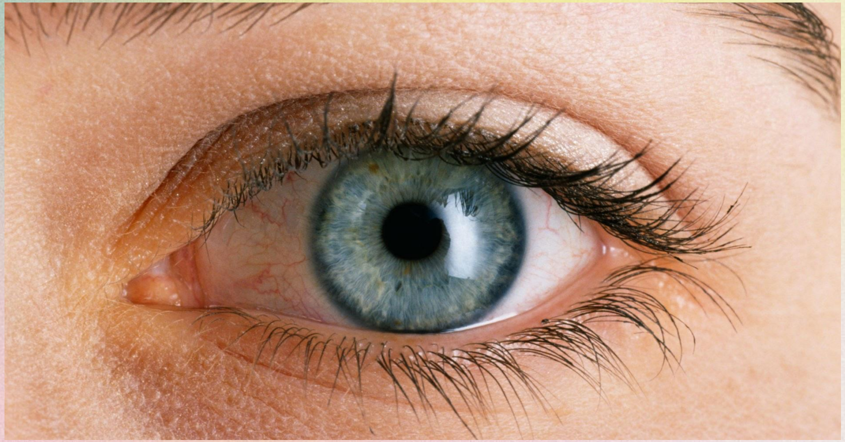
Smart Vision: Early Detection of Retinal Diseases Using Deep Learning

Without your eye you are blind!



Project Members

Arpita Santra (242110602)

Aakriti Sarkar (242110601)

Asit Biswas (241110014)

Aman Kumar Singh (241110007)

Gautam Raj(241110025)

Prabhakar Pandey (241110049)

Problem statement & Motivation

- Eye diseases like Glaucoma, DR, AMD, HR, etc., can cause irreversible blindness
- Early detection is critical but requires expert interpretation
- Aim: Use deep learning to automate diagnosis from fundus images
- Build efficient and accurate models for 6 eye diseases

Dataset Overview

AMD
7,284
Age-related Macular Degeneration

Cataract
6,845
Cataract

DR
7,912
Diabetic Retinopathy (DR)

Glaucoma
8,390
Glaucoma

HR
6,100
Hypertensive Retinopathy (HR)

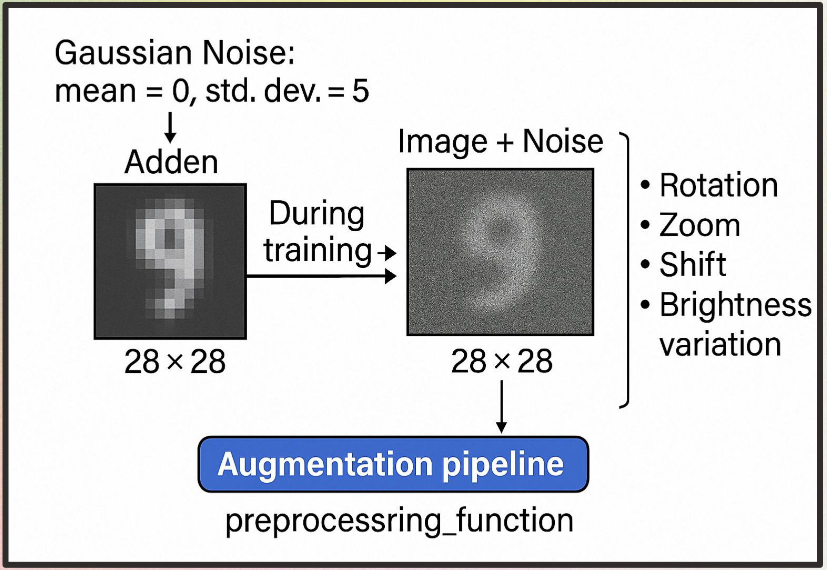
PM
5,710
Pathological Myopia (PM)

Normal
13,205
Normal Image

We Have used Data Augmentation on 20,000 images to make it 50,000

Gaussian Noise Augmentation

- We added Gaussian noise to each image during training to simulate realistic imperfections.
- Used mean = 0 and standard deviation = 5 to keep the noise subtle but effective.
- Integrated directly into the augmentation pipeline using preprocessing_function.
- We didn’t rely solely on Gaussian noise we also include rotation, zoom, shift, and brightness variation.



Individual Disease Model

AMD

Achieved 95% Accuracy

1. Custom CNN with 10 Conv blocks (64→512), GlobalAvgPool, Dense(256→128), Output(2)
2. BatchNorm, Dropout, Data Augmentation, Transfer Learning (with optional fine-tuning)

HR

Achieved 96% Accuracy

1.Got ~96% accuracy with pretrained ResNet 50 with 45 unfrozen layers.

2.Got ~91% accuracy with pre trained ImageNet V2 with with 30 unfrozen layers.

3.Got ~88% accuracy with CNN with 10 layers.

PM

Achieved 99.8% Accuracy

1.Trained a VGG16 model for binary classification of Pathological Myopia.

2. Image size: 224×224 RGB images

Individual Disease Model

GLAUCOMA

Achieved 90% Accuracy

1. Used pre-trained ResNet50 and VGG16 for feature extraction.
2. A MLP with an input layer, two hidden layer (512, 128 units) and an output layer was used to get the final predictions

CATARACT

Achieved 95.6% Accuracy

1. Custom CNN with 10 Conv blocks (64→512), GlobalAvgPool, Dense(256→128), Output(2)

2. BatchNorm, Dropout, Data Augmentation, Transfer Learning (with optional fine-tuning)

DR

Achieved 94% Accuracy

1. DenseNet-201 Pre-trained Model Used

2. APTOS 2019 Blindness Detection dataset

3. Preprocessing Applied

4. Achieved an accuracy of 94%

What Makes Us Unique

1. Modular + Unified Approach

Unlike traditional models that handle all classes together from the start, we first built six specialized models (one per disease) to capture class-specific nuances, and then fused that knowledge using a Vision Transformer + Knowledge Distillation pipeline.

2. Knowledge Distillation from Multiple Teachers

Instead of a single teacher, we distilled knowledge from six binary experts and a multi-class ViT, making our student model more well-rounded and efficient.

3. Lightweight & Accurate

While ViT models are powerful, they're usually large — we optimized ours to have only ~28M parameters, and our distilled model is even smaller, making it suitable for real-time or embedded deployment.

Tried Models

DENSE CNN:
Achieved 81.2% accuracy

Conv Blocks: 10 conv layers (64→512) with BatchNorm, MaxPooling, and Dropout.

Pooling: GlobalAveragePooling2D to flatten features.

Dense Head: Dense layers (256→128→7) with BatchNorm & Dropout for classification.

01

Efficient B2:
Achieved 85.7% accuracy

Base Model: Uses EfficientNetB2 as a feature extractor (pretrained).

Pooling: GlobalAveragePooling2D to reduce feature maps to 1D.

Dense Head: Custom dense layers (e.g., 256→128→7) with BatchNorm & Dropout for classification.

02

VISION TRANSFORMER: Achieved 93.7% accuracy

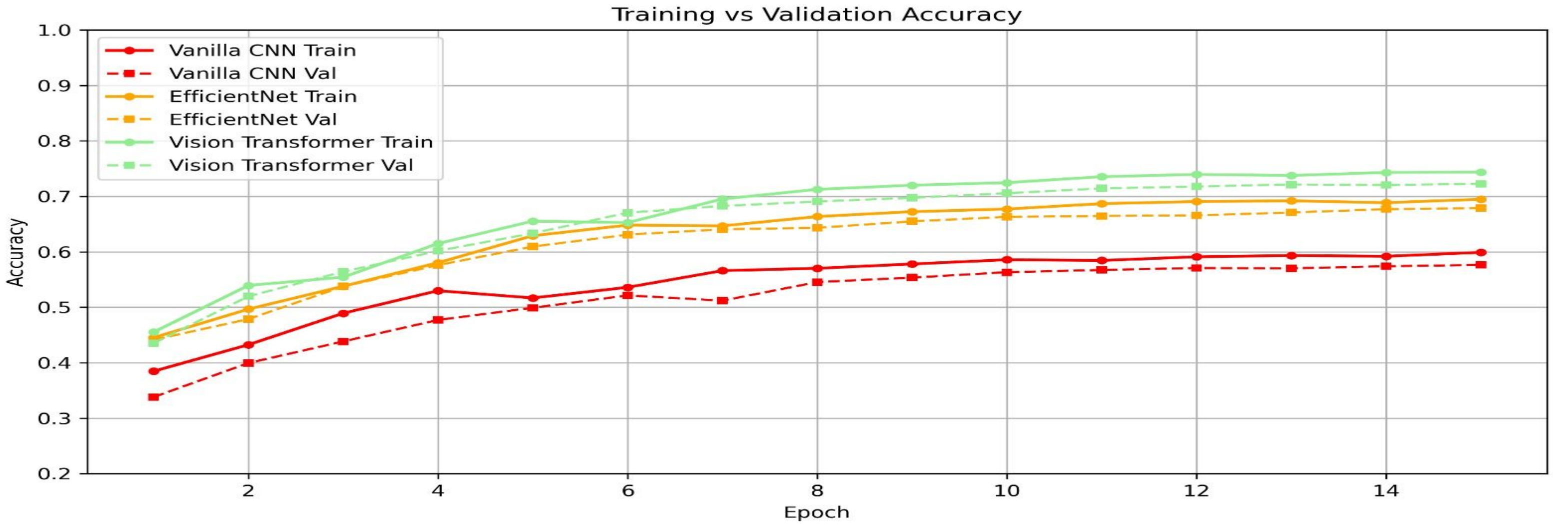
Patch + Positional Encoding: Image is split into patches via Conv2D and flattened; positional embeddings and a learnable CLS token are added.

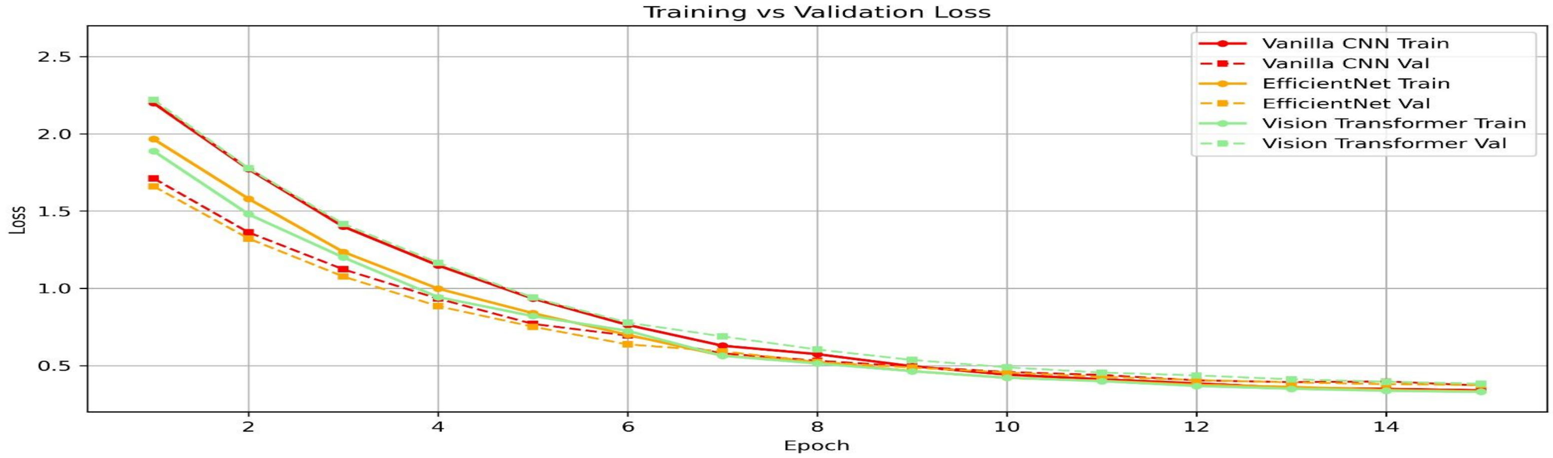
Transformer Encoder: Repeats 12 transformer blocks with LayerNorm, multi-head self-attention, MLP, and residual connections.

Classification Head: CLS token is extracted, passed through dropout and a dense layer for 7-class softmax prediction.

03

COMPARATIVE ANALYSIS





Vision Transformer(ViT) for Eye Disease Detection: Architecture Overview

Main Components:

- **Input:** 224×224×3 fundus images
- **Patch Embedding:** 16×16 patches → 256-dimensional vectors
- **Position Encoding + CLS Token:** Spatial information + classification token
- **Transformer Encoder Stack:** 12 layers of self-attention blocks
- **Classification Head:** Disease prediction

ViT Architecture Diagram

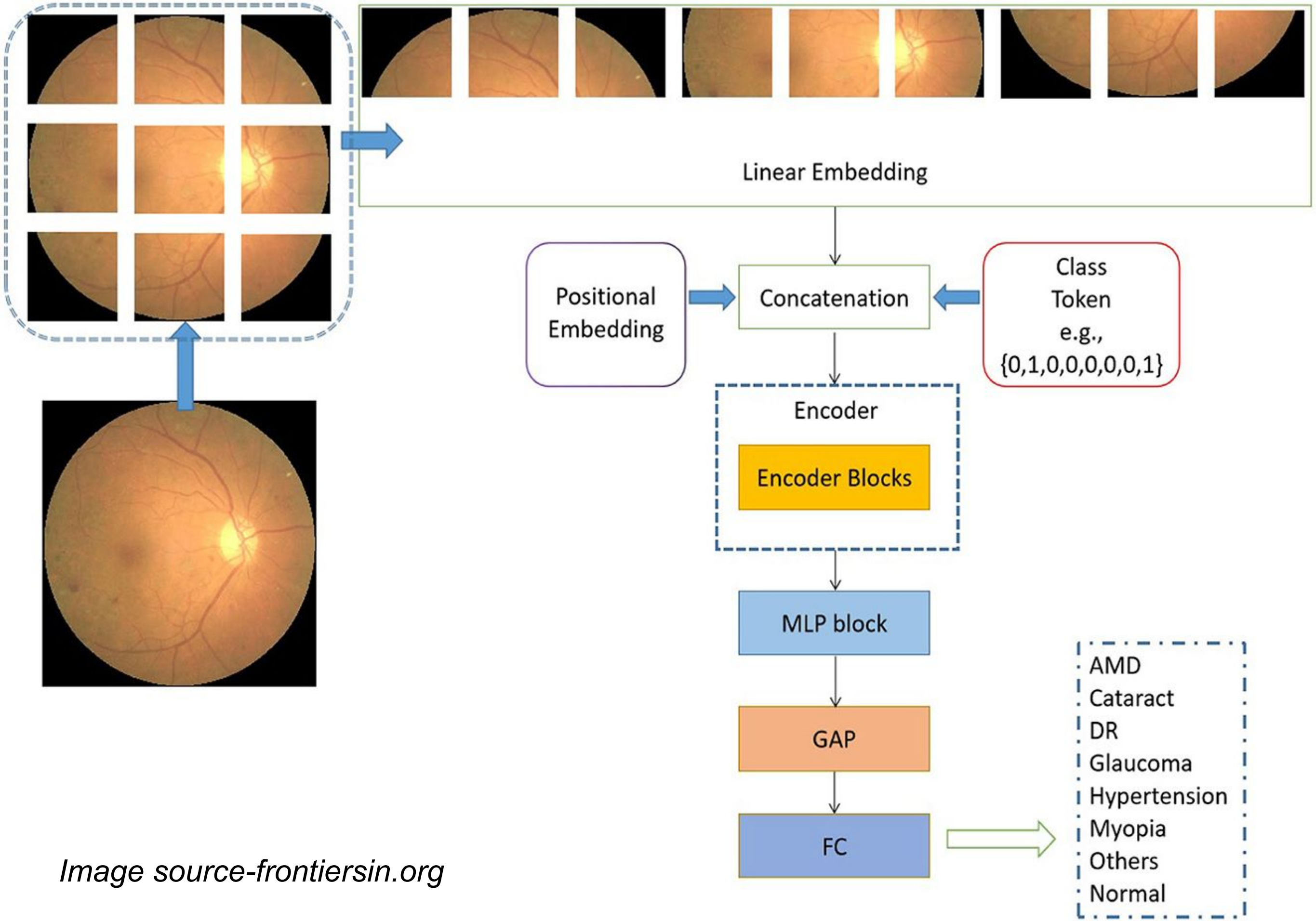


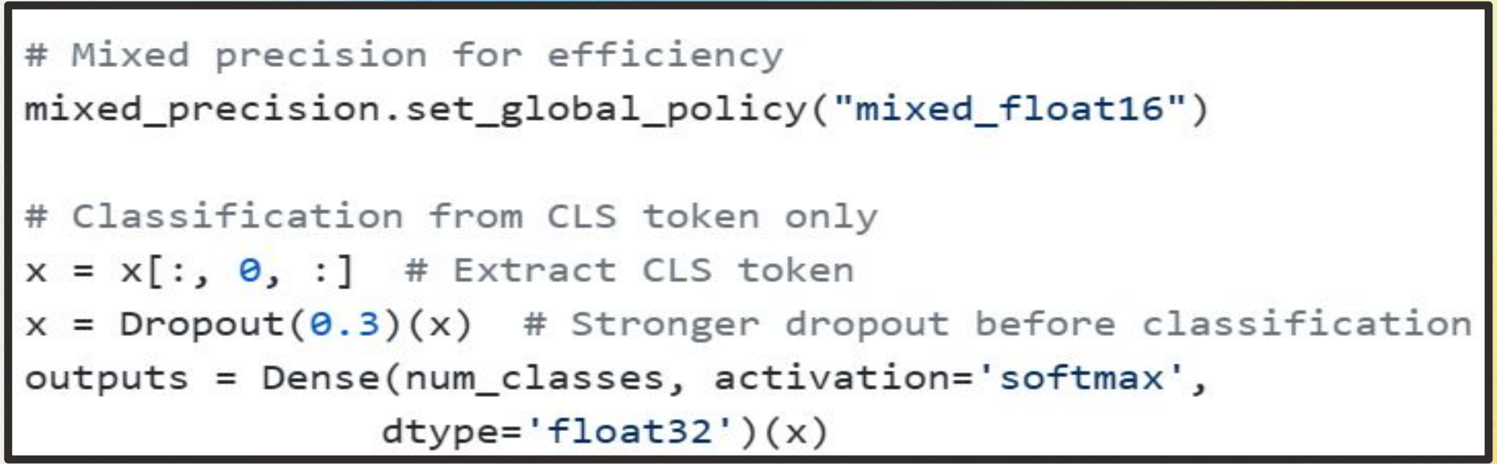
Image source-frontiersin.org

Critical Design Choices & Parameter Selection

Key Architectural Parameters:

Parameter	Value	Reasoning
Patch Size	16×16	Optimal balance between detail and computation
Embedding Dimension	256	Sufficient capacity for retinal features
Transformer Layers	12	Deep enough for complex pattern recognition
Attention Heads	8	Can focus on different disease patterns simultaneously
MLP Dimension	512	2× expansion for feature transformation
Dropout Rate	0.1 (0.3 before classification)	Prevents overfitting on limited data

Implementation Optimizations:

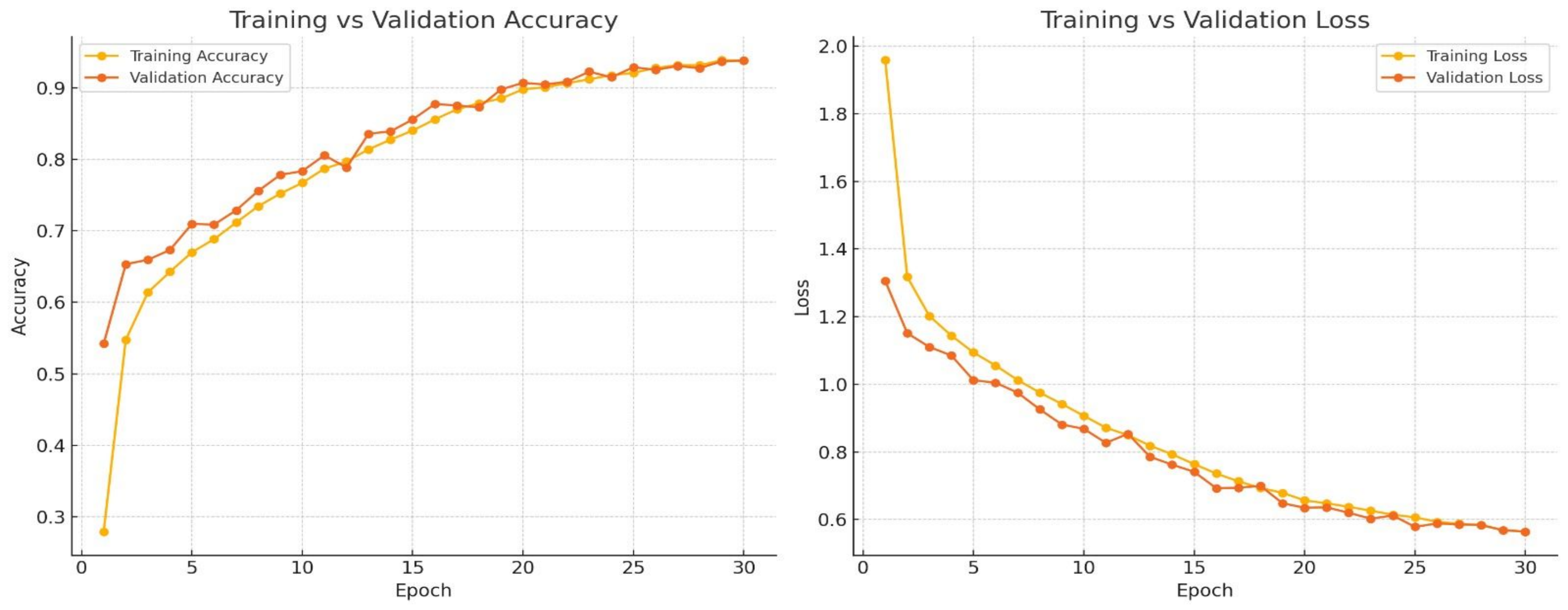


```
# Mixed precision for efficiency
mixed_precision.set_global_policy("mixed_float16")

# Classification from CLS token only
x = x[:, 0, :]  # Extract CLS token
x = Dropout(0.3)(x)  # Stronger dropout before classification
outputs = Dense(num_classes, activation='softmax',
               dtype='float32')(x)
```

- Mixed precision training (`mixed_float16`) for faster computation
- Carefully tuned data augmentation preserving pathological features
- Label smoothing (0.1) to handle class imbalance
- Adam optimizer with 1e-4 learning rate

ViT Accuracy



Training vs Validation Accuracy

Training vs Validation Loss

Best Model
93.70%

Knowledge Distillation Setup

To train a lightweight **student CNN model** that mimics the performance of a powerful ensemble of expert models for **multi-disease retinal classification**.

Teacher Ensemble

- **1 Combined Model**
  A 7-class classifier trained to detect all diseases simultaneously.
- **6 Individual Expert Models**
  Specialized binary classifiers, each trained to detect a specific disease:
  AMD, Cataract, Diabetic Retinopathy, Glaucoma, Hypertensive Retinopathy, Pathological Myopia

Student Model

- A **basic CNN**, trained from scratch — small, efficient, and suitable for real-time deployment.
- Learns from:
  Hard labels (ground truth)
  Soft labels (generated by teacher ensemble)
- Optimized using a **custom distillation loss function** combining KL Divergence and Cross-Entropy.

Soft Label Generation

How Soft Labels Are Created

- **Combined Model**
  Outputs a **7-class prediction vector** (logits) covering all diseases + normal.
- **Individual Models (×6)**
  Each outputs a **binary prediction** for a specific disease (e.g., AMD vs Normal).

Fusion Strategy

We generate soft labels by **combining predictions** using weighted averaging:

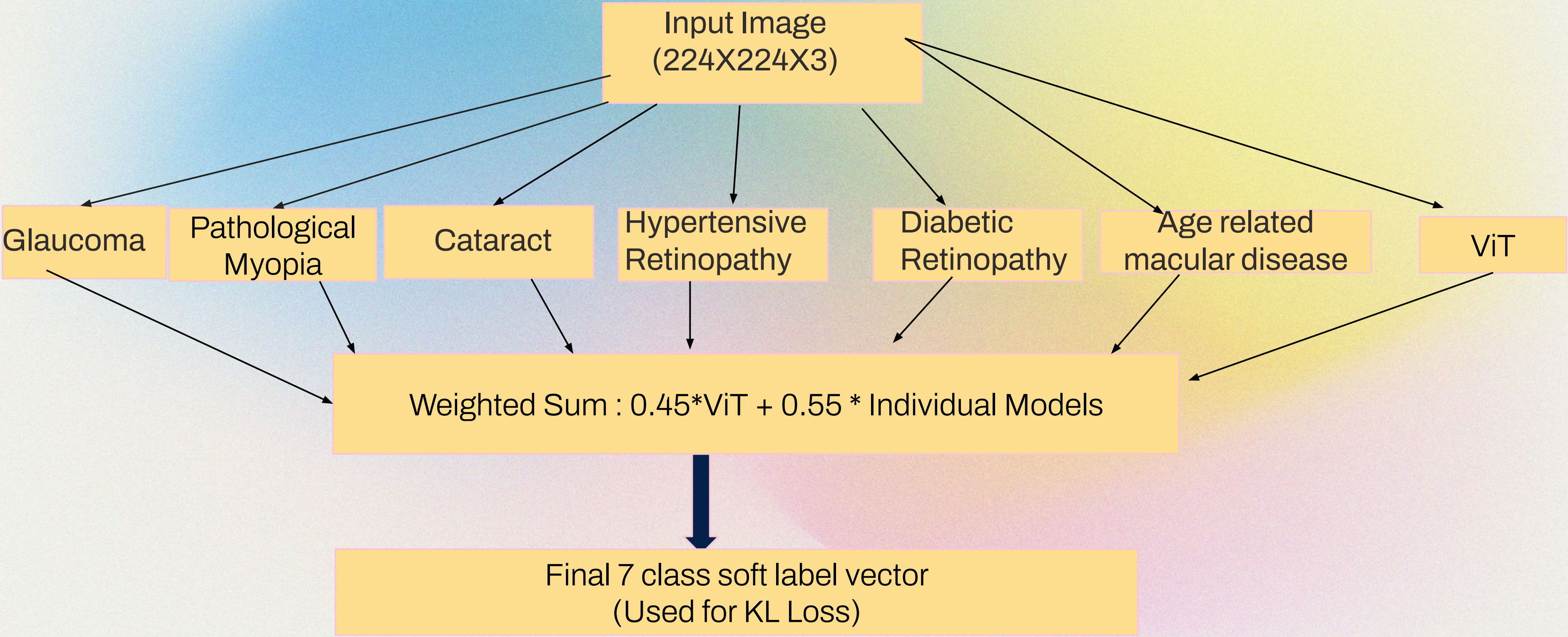
soft_logits = 0.45 * combined_logits + 0.55 * (average of individual model logits)

- Combined model contributes **45%** of the weight & Individual models together contribute **55%**

Final Output

- The resulting soft_logits vector is passed through a **softmax** to get the final **teacher soft labels**.
- These soft labels guide the student to learn from **both broad and specialized expert knowledge**.

KD Flow Diagram



Input Image
(224X224X3)

Glaucoma

Pathological Myopia

Cataract

Hypertensive Retinopathy

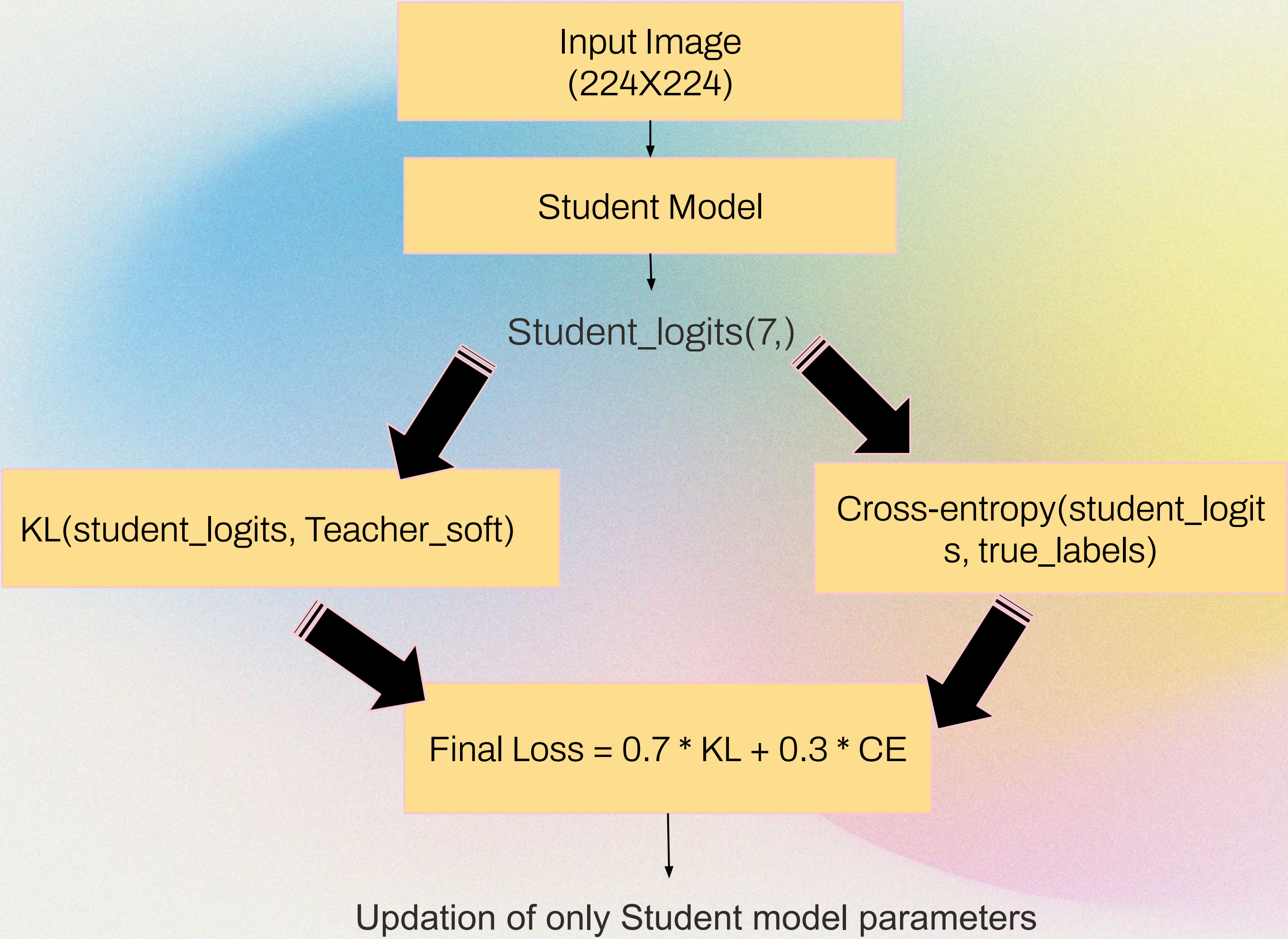
Diabetic Retinopathy

Age related macular disease

ViT

Weighted Sum : 0.45*ViT + 0.55 * Individual Models

Final 7 class soft label vector
(Used for KL Loss)



Input Image
(224X224)

Student Model

Student_logits(7,)

KL(student_logits, Teacher_soft)

Cross-entropy(student_logits, true_labels)

Final Loss = 0.7 * KL + 0.3 * CE

Updation of only Student model parameters

Distillation Loss & Training

Distillation Loss Function

We combine two types of supervision into a single loss:KL Divergence: Between student’s output and soft labels from the teacher

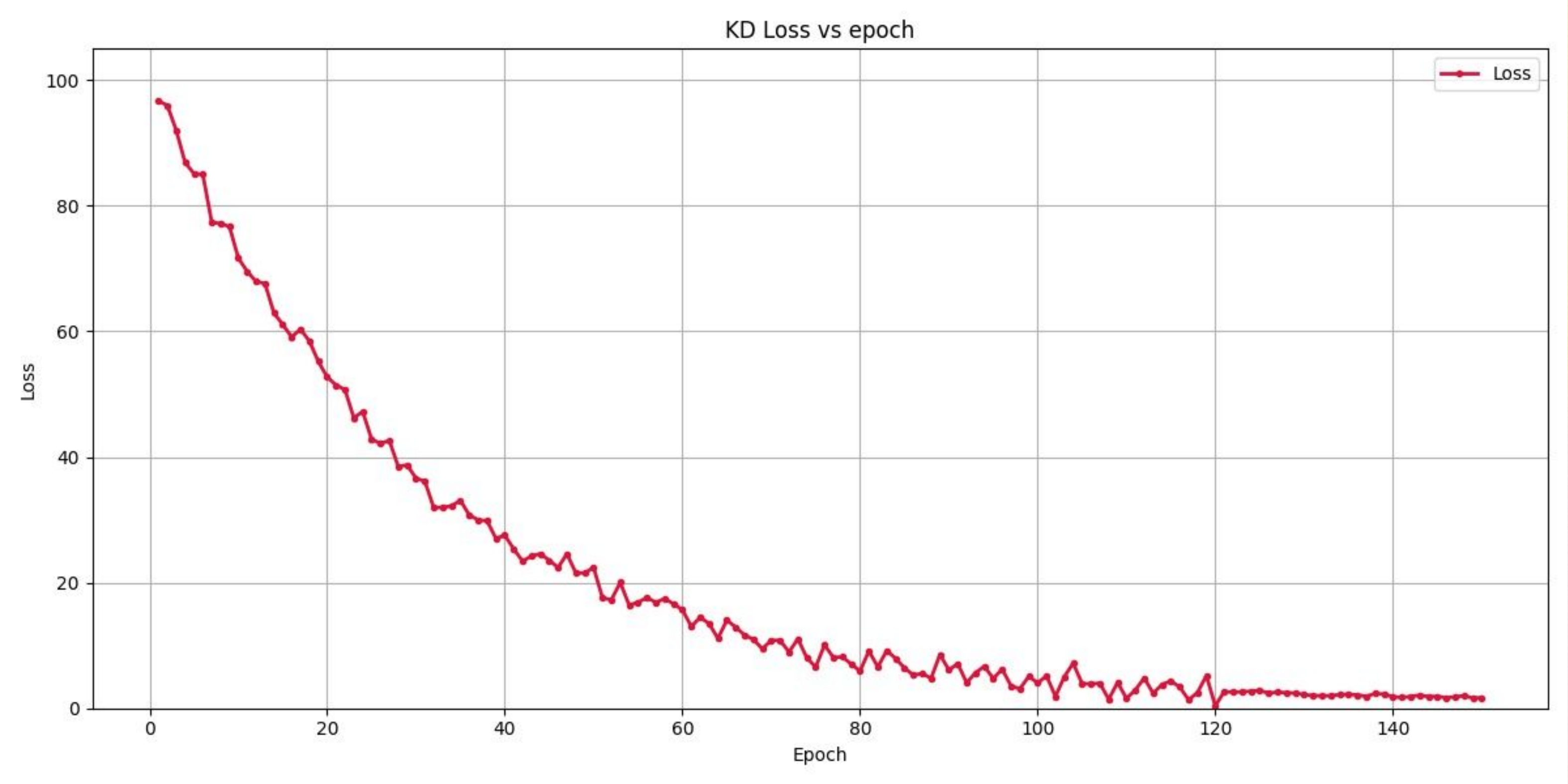
- Cross-Entropy: Between student’s output and true hard labels
- α = 0.7, Temperature = 4.0 (we tried temperature as 4,5 and temperature 4 was working well so we chose this)

Training Flow

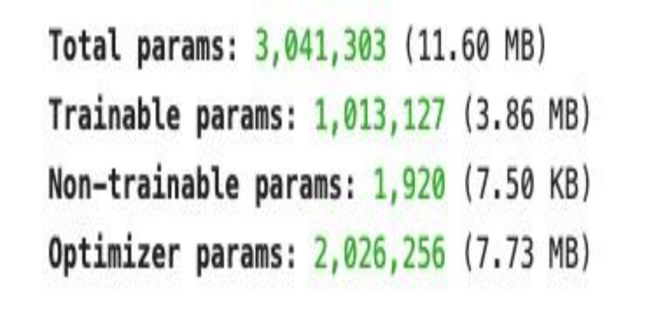
- Teacher models are frozen (no gradient updates)
- Student CNN is trained using a mix of soft + hard targets
- Gradient updates are computed
- Optimizer: Adam with learning rate = 1e-4

This strategy helps the student model generalize better by learning not just the correct label, but also the confidence patterns and inter-class relationships from the teacher.

KD Loss

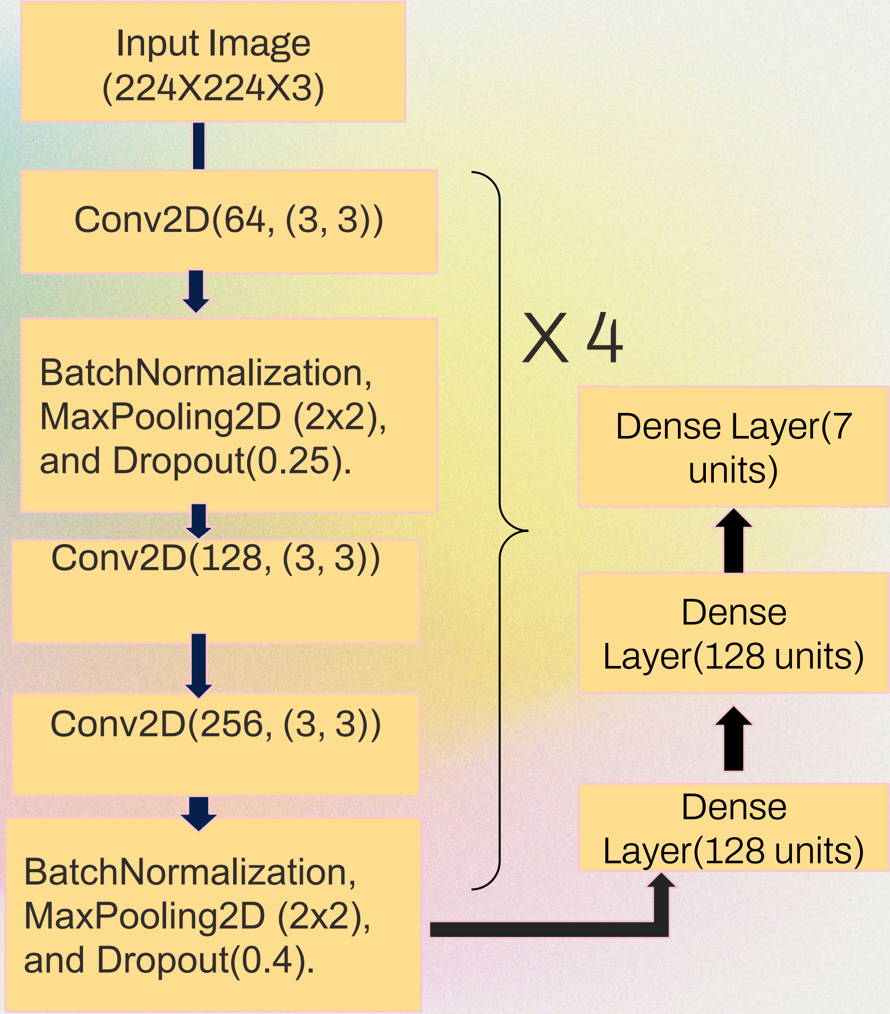

KD Loss vs epoch
Loss
100
80
60
40
20
0
0
20
40
60
80
100
120
140
Epoch
Loss

Student Model



```
Total params: 3,041,303 (11.60 MB)
Trainable params: 1,013,127 (3.86 MB)
Non-trainable params: 1,920 (7.50 KB)
Optimizer params: 2,026,256 (7.73 MB)
```

Accuracy
82.7%



Input Image (224X224X3)

Conv2D(64, (3, 3))

BatchNormalization, MaxPooling2D (2x2), and Dropout(0.25).

Conv2D(128, (3, 3))

Conv2D(256, (3, 3))

BatchNormalization, MaxPooling2D (2x2), and Dropout(0.4).

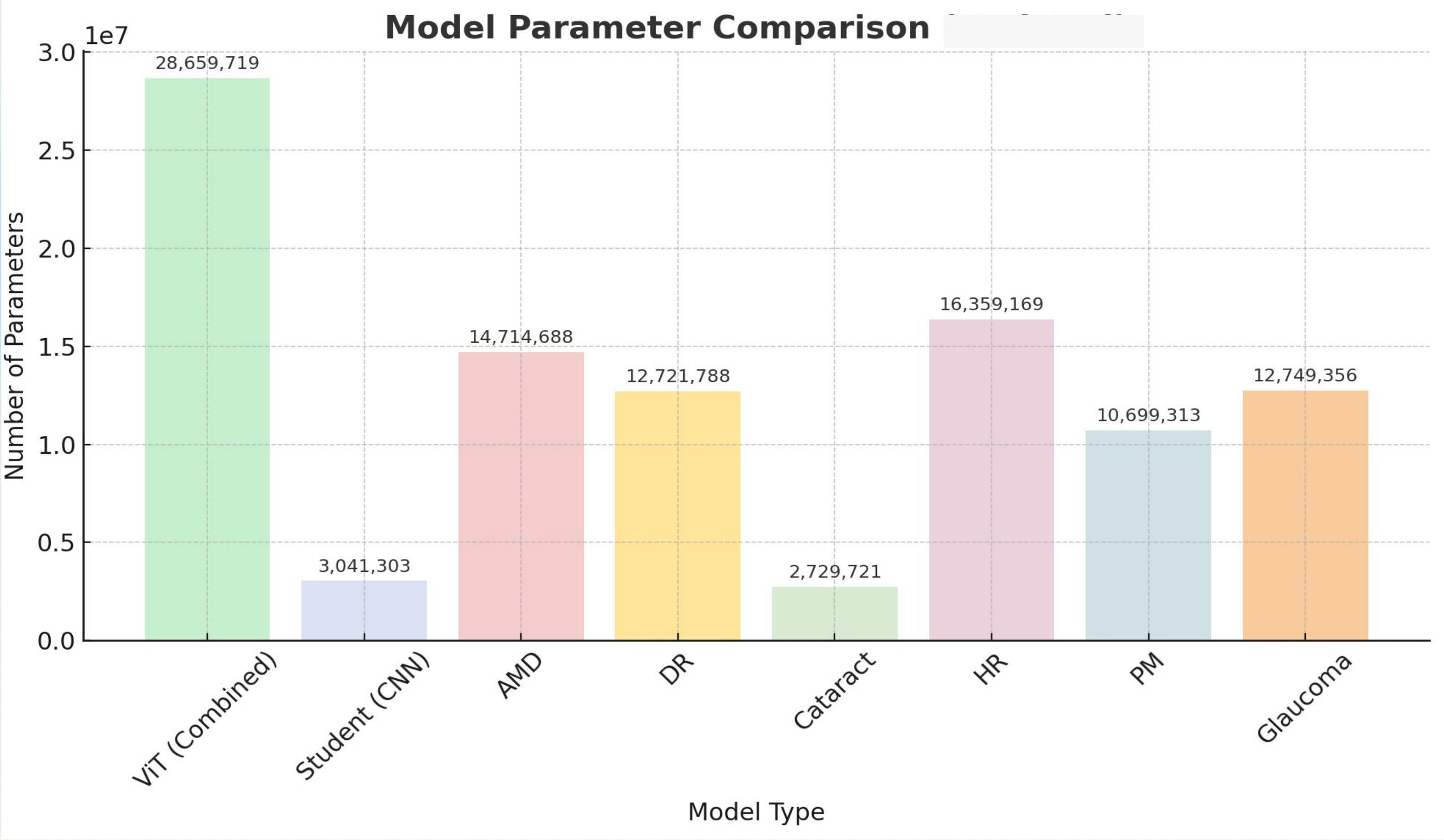
X 4

Dense Layer(128 units)

Dense Layer(128 units)

Dense Layer(7 units)

Trainable Parameter Comparison



Why This Setup Works

1. Combines global knowledge from the combined 7-class model
2. Leverages specialized expertise from individual disease-specific models
3. Student model is a simple CNN, making it fast, lightweight, and easy to deploy
4. Achieves strong performance by learning from teacher ensemble patterns
5. "We used a simple CNN model as the student and trained it from scratch using distillation. The teacher models were pre trained and frozen. The student learns through supervision from both the teacher ensemble (soft labels) and ground truth (hard labels).

Challenges we tackled

Data Augmentation

Used Data Augmentation for improving the Robustness and Reliability

Resource Constraint

The primary constraint during model training was limited access to GPU resources.

Architecture

Finding out which Architecture works best.

HR Vs DR

HR and DR are challenging to distinguish due to overlapping retinal features.

Dataset Selection

Problems with the so called most famous Datasets

Combining all the Models

Combining all the models to get a reliable reliable accuracy with minimum parameters possible was very challenging.

Contributions

ARPITA SANTRA

-Trained Diabetic Retinopathy classification model
- Contributed in ViT and KD (16.66%)

AAKRITI SARKAR

-Trained Glaucoma classification model
- Contributed in ViT and KD (16.66%)

ASIT BISWAS

-Trained Hypertensive Retinopathy classification model
- Contributed in ViT and KD (16.66%)

AMAN KUMAR SINGH

-Trained Pathological Myopia classification model
- Contributed in ViT and KD (16.66%)

GAUTAM RAJ

-Trained Cataract classification model
- Contributed in ViT and KD (16.66%)

PRABHAKAR PANDEY

-Trained AMD classification model
- Contributed in ViT and KD (16.66%)

References

- Paper 1 - Advancing Diabetic Retinopathy Detection: Leveraging Deep Learning for Accurate Classification and Early Diagnosis Authors: Aakash Dilip Kolte, Jyotiraditya Sharma, Utkarsh Rai, R.Jansi Journal: IEEE Xplore (2023) Link
- Paper 2 - A Deep Learning Framework for the Early Detection of Multi-Retinal Diseases Authors: Sara Ejaz, Raheel Baig, Zeeshan Ashraf, et al.Journal: PLoS ONE (July 25, 2024) Link
- Paper 3 - Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks Philippe M. Burlina, PhD; Neil Joshi, BS; Michael Pekala, MS; et al *JAMA Ophthalmol.*
- Paper 4 - Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks Cen, LP., Ji, J., Lin, JW. *et al.* *Nat Commun* 12, 4828 (2021).
- Paper5 - Eye-Vision Net: Cataract Detection and Classification in Retinal and Slit Lamp Images using Deep Network Authors: Binju Saju1, Rajesh R2. Journal: *IJACSA* (June 17, 2022)
- Paper 6 - Deep Learning-Based Eye Disease Recognition Using Transfer Learning and Improved D-S Evidence Theory

References

Paper 7: "Computer-Aided Detection of Hypertensive Retinopathy Using Depth-Wise Separable CNN" (2022)

Paper 8 - "A Foundation Model for Generalizable Disease Detection from Retinal Images" (2023)

Paper 9 - Fundus-deepnet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images, Al-Fahdawi, S., Al-Waisy, A. S., Zeebaree, D. Q., Qahwaji, R., Natiq, H., Mohammed, M. A., ... & Deveci, M. (2024).. *Information Fusion*, *102*, 10205

Paper 10 : A hybrid framework for glaucoma detection through federated machine learning and deep learning models. Aljohani, A., & Aburasain, R. Y. (2024). *BMC Medical Informatics and Decision Making*, *24*(1), 115.

Paper 11 : "Pathological Myopia Classification with Simultaneous Lesion Segmentation Using Deep Learning", published in 2021

Paper12 - Multi-Task Knowledge Distillation for Eye Disease Prediction Sahil Chelaramani1, Manish Gupta1, Vipul Agarwal1, Prashant Gupta1, Ranya Habash2 1Microsoft, 2Bascom Palmer Eye Institut

Paper - 13 - Distilling the Knowledge in a Neural Network

Thank You