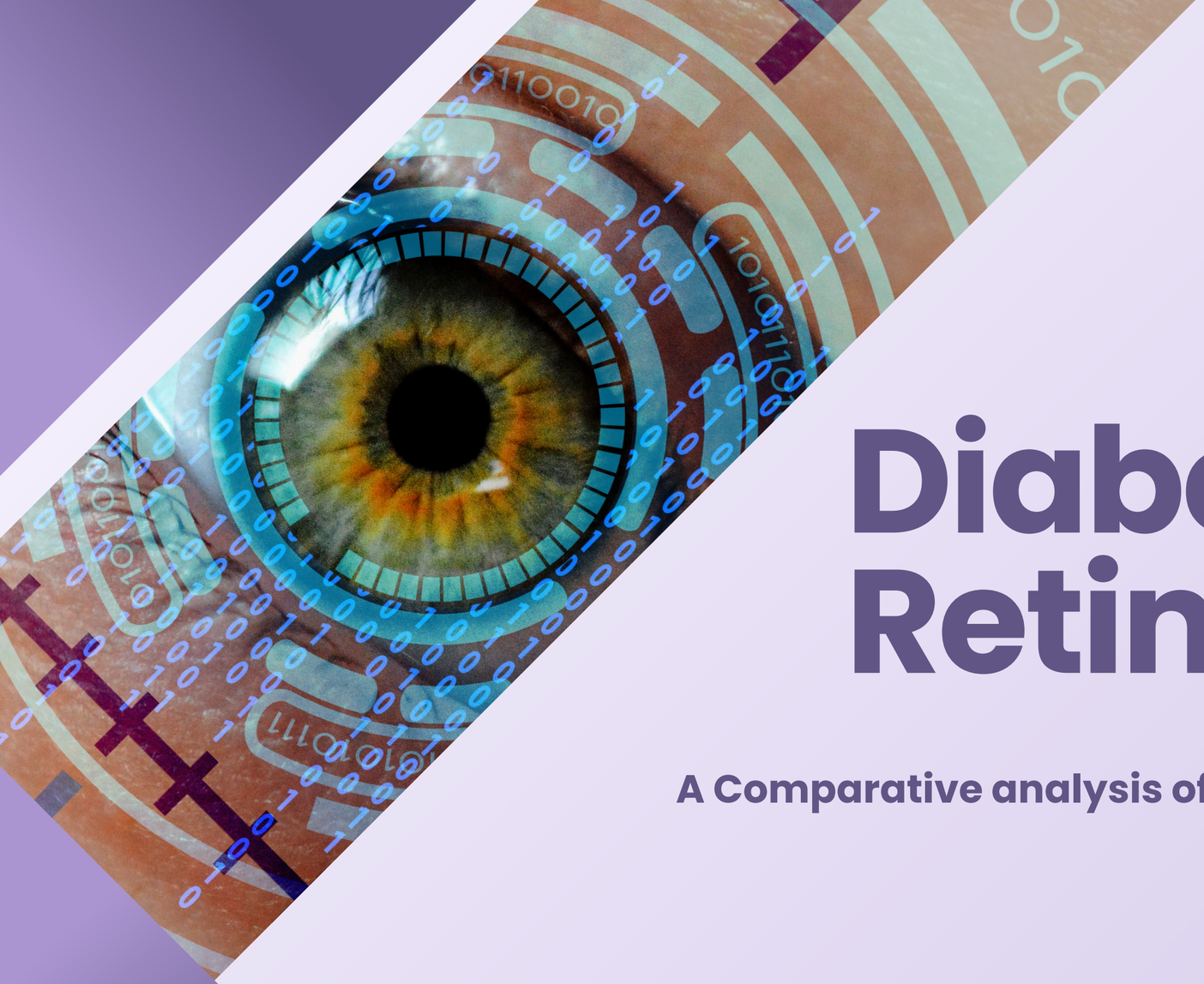


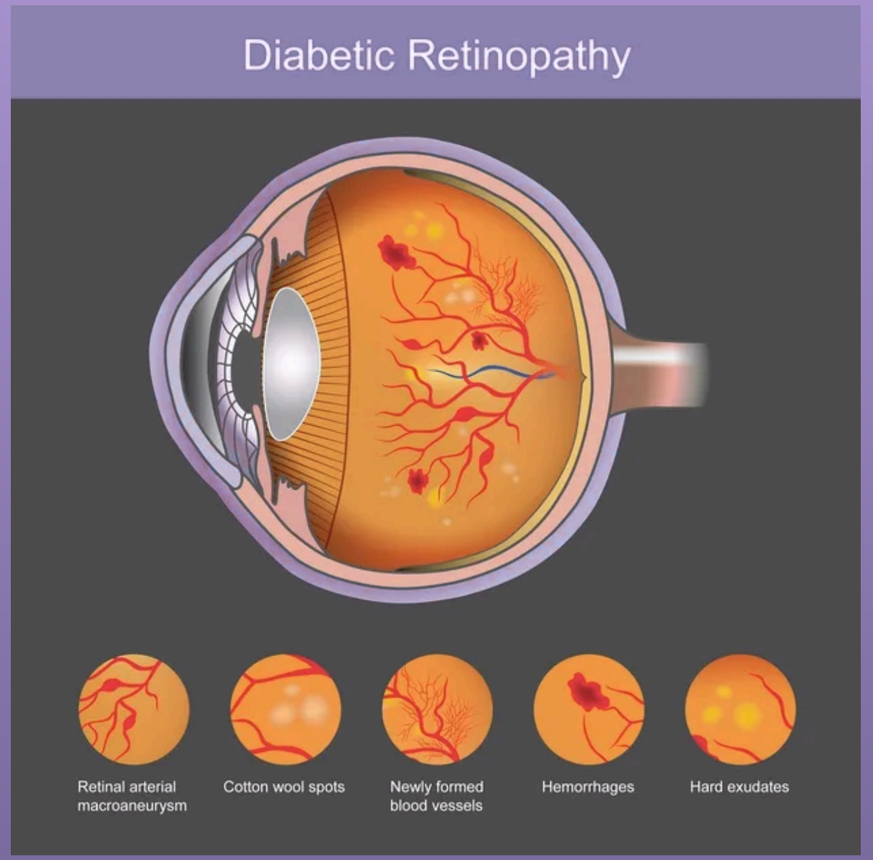
Diabetic Retinopathy

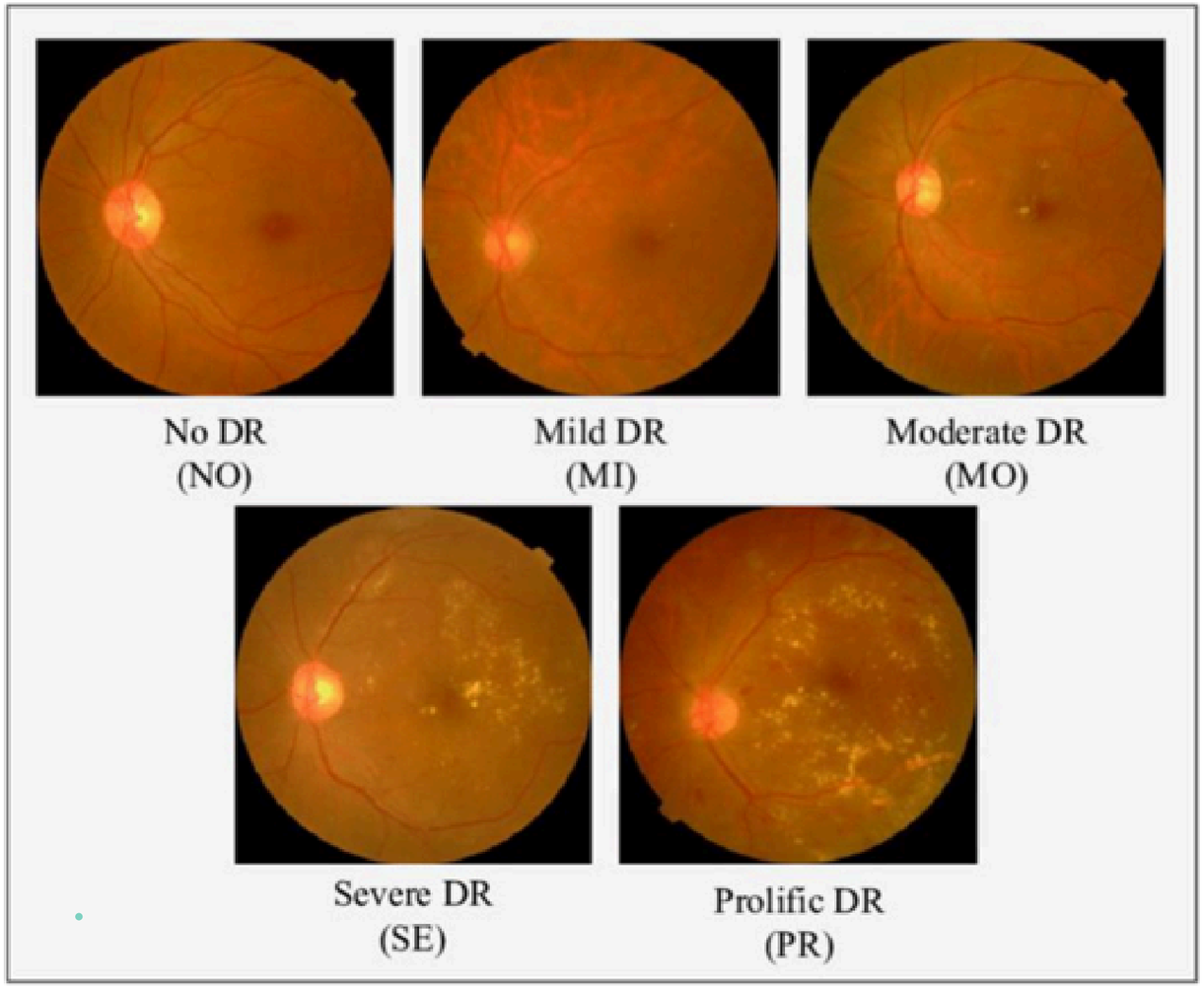
A Comparative analysis of CNN models and Vision Transformer.

Team Members:
PARNAVI SHARMA : E22CSEU1315
SHIVANSHU GARG : E22CSEU1296
DIVYANSH GUPTA : E22CSEU1301

What is Diabetic Retinopathy?

- Diabetic retinopathy is an eye condition caused by damage to the blood vessels in the retina due to high blood sugar levels in people with diabetes.
- It is a serious microvascular complication of diabetes mellitus of the retinal vasculature that causes vision loss and blindness unless treated.





Diabetic Retinopathy (DR) occurs through a sequence of five steps:

01 No DR - perfect vision

02 Mild nonproliferative retinopathy is the initial phase characterized by the occurrence of micro-aneurysms only.

03 Moderate nonproliferative retinopathy occurs as the disease progresses, leading to the distortion and swelling of blood vessels, resulting in the impaired transportation of blood.

04 Severe non-proliferative retinopathy manifests when the blockage of blood vessels increases, causing a reduced blood supply to the retina. This prompts signals for the growth of new blood vessels.

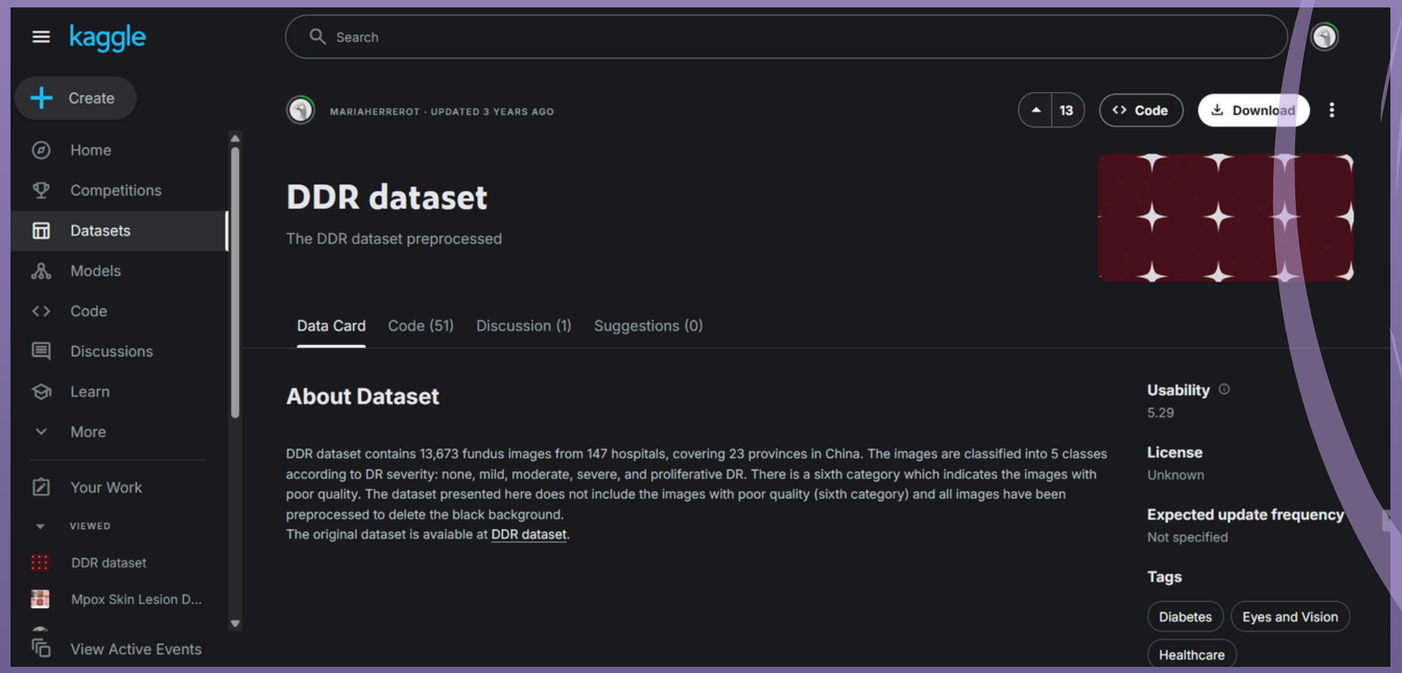
05 Proliferative diabetic retinopathy denotes the advanced stage characterized by the release of growth factors by the retina, which triggers the development of new blood vessels.

PROJECT OBJECTIVES

- This project presents a comparative study of deep learning models for computer-aided DR detection and severity grading using the DDRDataset. We analyze the efficiency of three models, i.e., a Convolutional Neural Network (CNN), VGG-19, and Vision Transformer (ViT), to evaluate their performance in DR detection and grading. The approach used involves transfer learning and data augmentation to present strong and accurate models.

- Our experimental result reveals that CNN structures exhibit high generalization with higher validation accuracy compared to more complex architectures like ViT, which exhibit faster convergence with more overfitting. VGG-19 acts fairly by being well-balanced with regard to accuracy and complexity. The study attests to the potential of deep learning-driven DR detection systems, marrying diagnostic performance with model efficiency, and supports the deployment of scalable, AI-driven screening instruments.

ABOUT THE DATASET

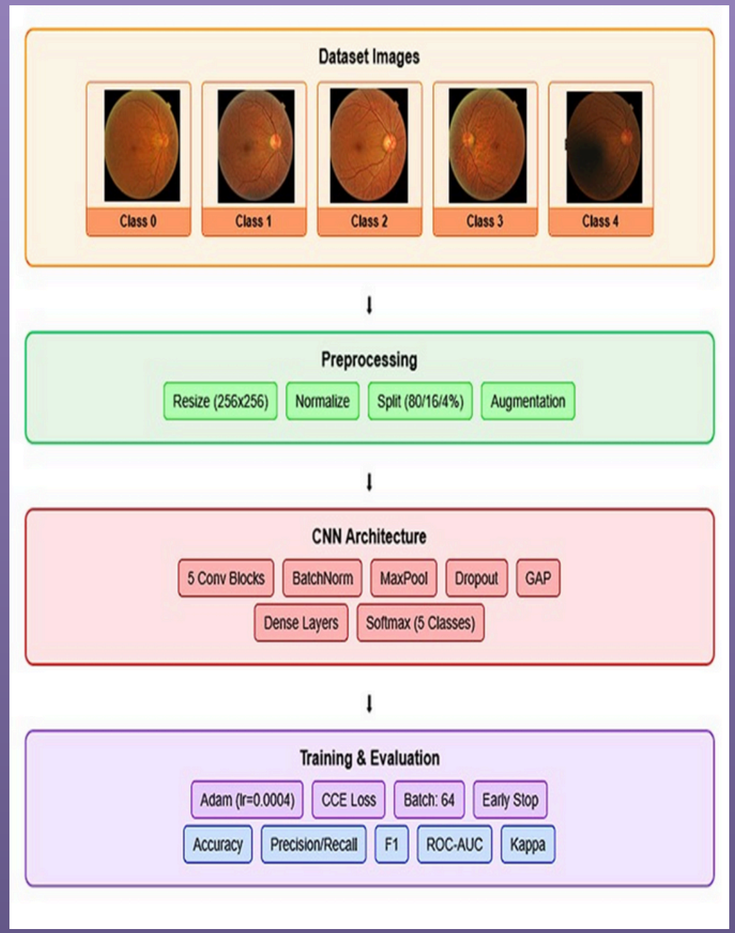


- The DDR (Diabetic Retinopathy Recognition) Dataset was used for training, validation, and testing purposes. It consists of high-resolution retinal images labeled based on the severity of diabetic retinopathy (e.g., No DR, Mild, Moderate, Severe, Proliferative).
- Total Images: 12K color fundus images
- Images are labeled into 5 diabetic retinopathy (DR) severity levels:
- 0 – No DR, 1 – Mild, 2 – Moderate, 3 – Severe, 4 – Proliferative DR
- Folder Structure: DDRdataset contains a folder DR_grading which further contains two folders → DR_grading (containing image files) and DR_grading.csv (Contains image filenames and their corresponding DR class labels).

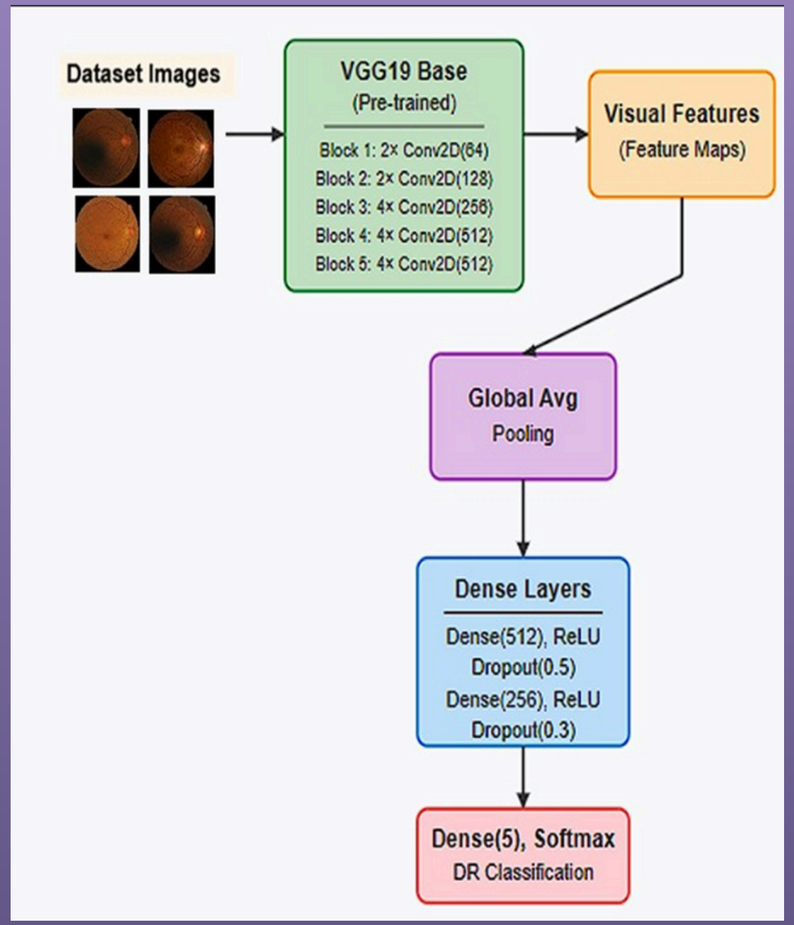
METHODOLOGY

We trained two CNN models (a custom basic CNN and VGG19) and Vision Transformer (ViT base patch 16) on our dataset and compared the results.

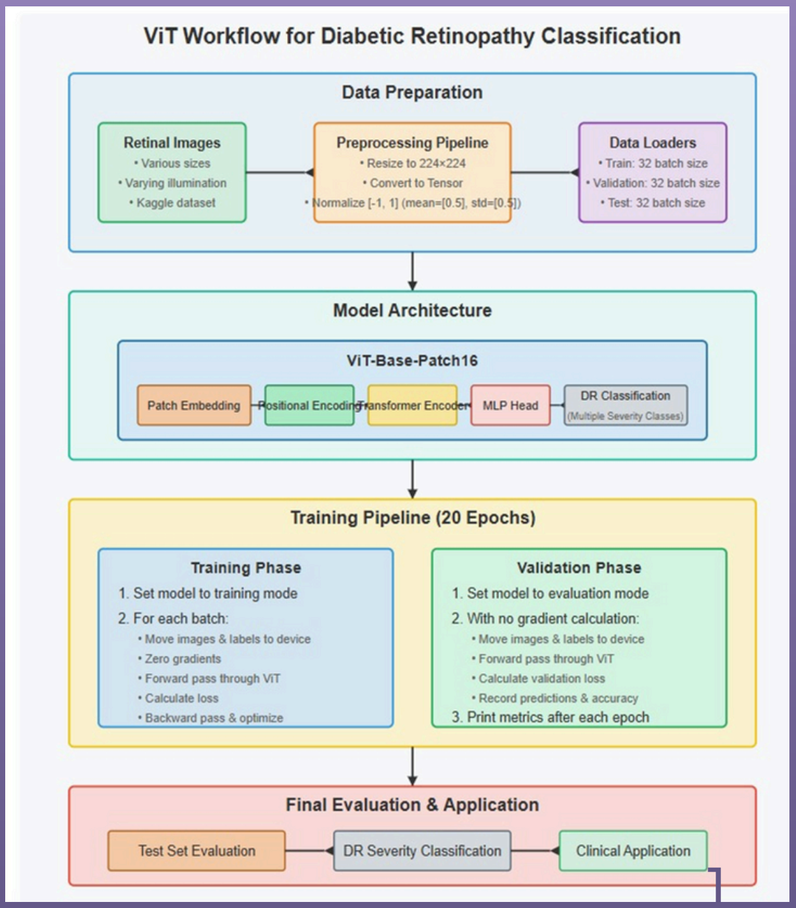
CNN WORKFLOW



VGG19 WORKFLOW



ViT WORKFLOW



Custom Basic CNN

Model Architecture:

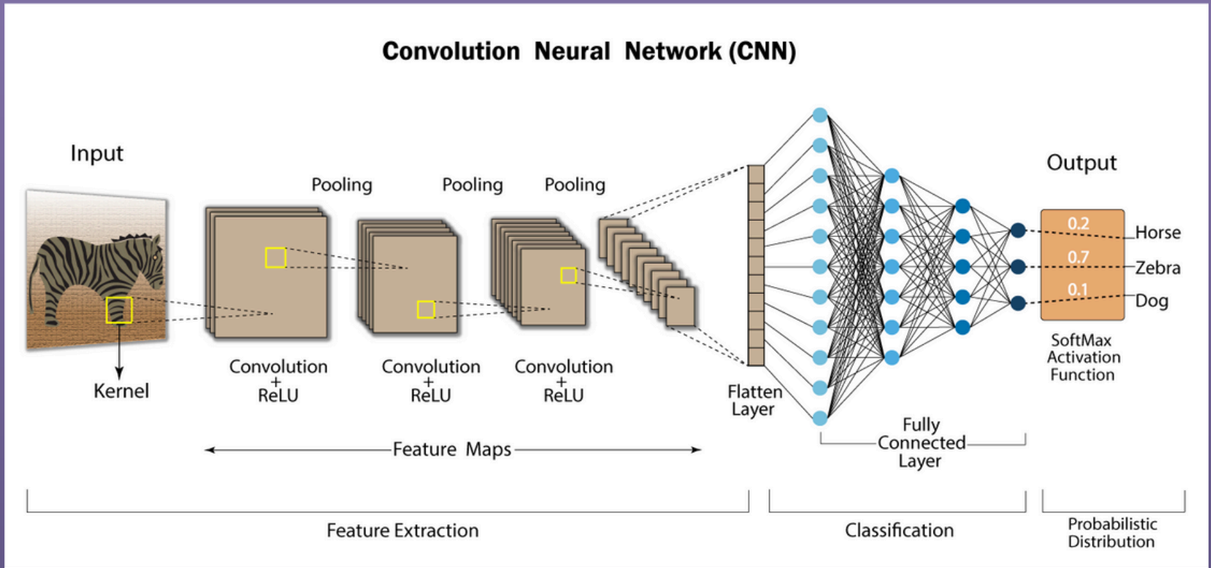
- Consists of 5 convolutional blocks (Conv2D + ReLU + Batch Normalization + MaxPooling) followed by Global Average Pooling
- Fully connected layers with Dropout (0.5), final SoftMax layer for multi-class prediction (5 DR grades)

Training Details:

- Optimizer: Adam (lr = 0.0004), Loss: Categorical Cross entropy, Batch size: 64, Epochs: 20, Early stopping & LR scheduler used
- Input images resized to 256×256 & normalized; data augmented using ±15° rotation & 10% shifts for better generalization

Performance & Evaluation:

- •Achieves ~65-70% accuracy; lightweight and computation-efficient
- •Evaluated using metrics like Accuracy, Precision, Recall, F1, AUC, Cohen's Kappa; visualized via loss/accuracy curves, confusion matrix



VGG19

Model Architecture (Transfer Learning – VGG19):

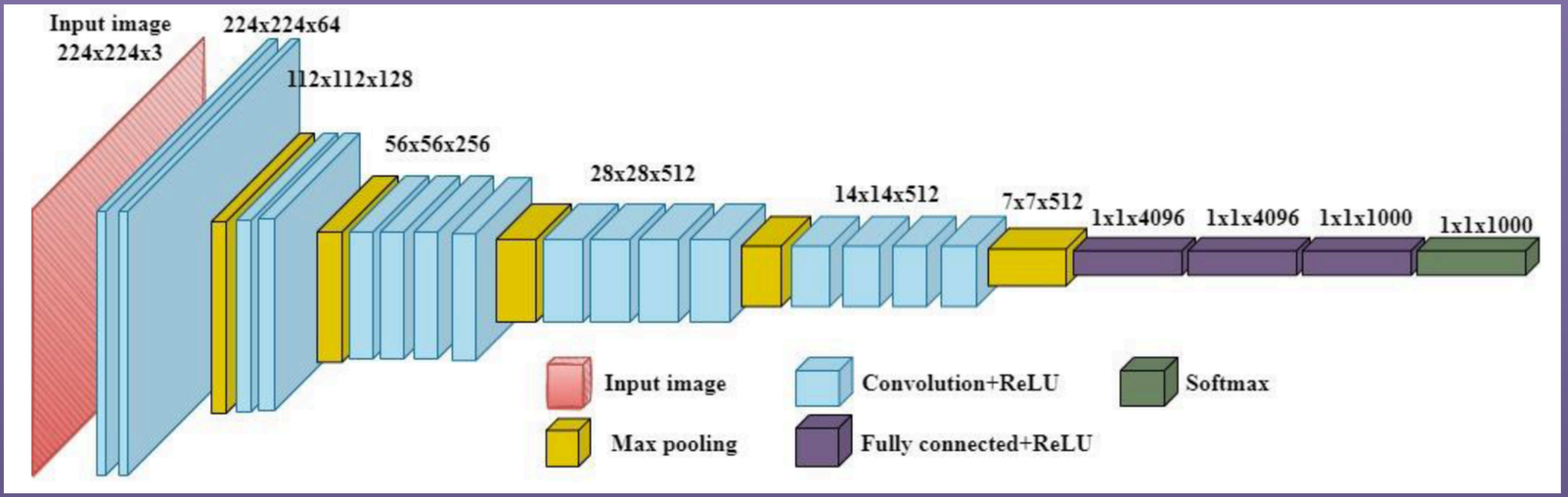
- Uses pre-trained VGG19 (trained on ImageNet) as a feature extractor; original top layers removed
- Final layers added: Global Average Pooling → Dense(512, ReLU) + Dropout(0.5) → Dense(256, ReLU) + Dropout(0.3) → SoftMax (5 classes)

Training Details:

- Input images resized to 256×256×3; initially froze all VGG19 layers, later unfroze last 8 for fine-tuning
- Optimizer: Adam (low LR), Loss: Categorical Cross entropy; Early stopping & checkpointing applied to prevent overfitting

Performance & Evaluation:

- Accuracy: ~75–80%; benefits from faster convergence and strong pre-learned visual features
- Limitations: Higher memory usage & longer training time than custom CNN; better suited for deeper feature extraction



ViT Base Patch 16

Model Architecture (Vision Transformer - ViT):

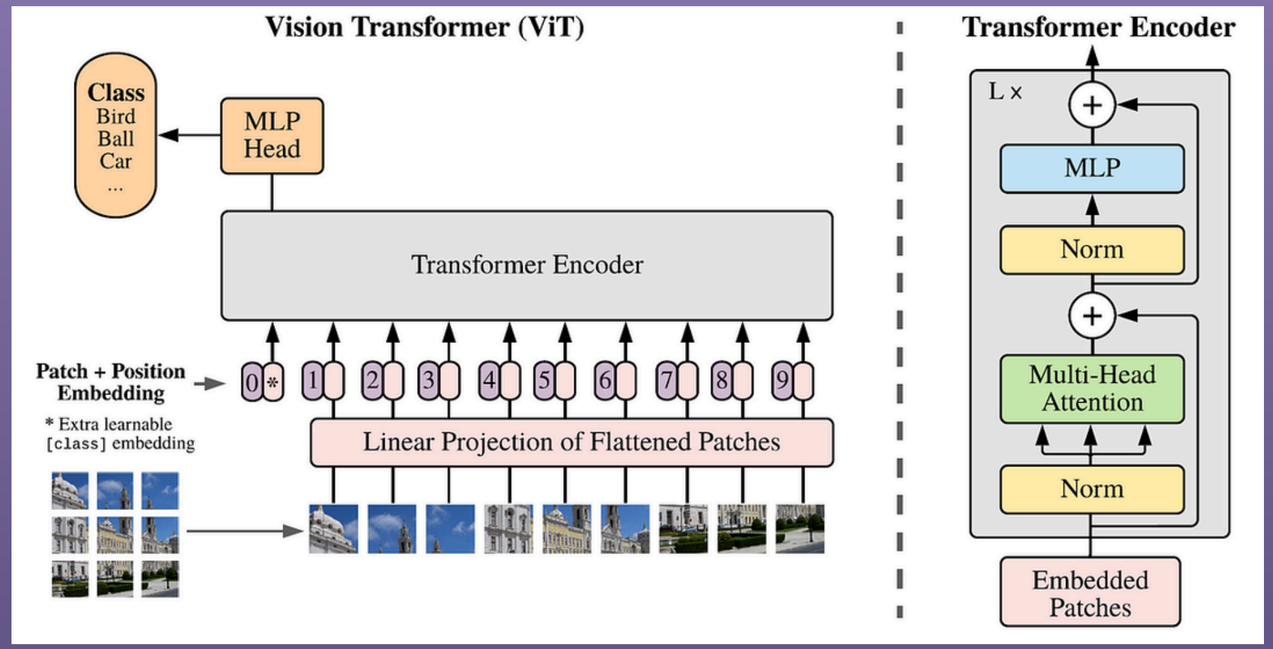
- Uses ViT-Base-Patch16-224 pretrained on ImageNet; image split into 16×16 patches (14×14 = 196 patches)
- Each patch is embedded into a 768-D vector; [CLS] token prepended; positional embeddings added
- Sequence passed through 12 Transformer encoder blocks using Multi-Head Self-Attention (MHSA) and Feed-Forward Networks (FFN)
- Final classification done via a linear layer on [CLS] token embedding → SoftMax over 5 DR classes

Training Details:

- Input resized to 224×224 and normalized to [-1, +1] (mean=0.5, std=0.5)
- Original ViT classification head replaced for DR; trained with categorical cross-entropy loss and fine-tuned on retinal images
- Residual connections + Layer Norm used for stability and faster convergence

Performance & Evaluation:

- Accuracy: ~75-80%; strong at capturing global relationships and fine-grained retinal features
- Limitations: Requires heavy compute, high memory, and longer training/inference time compared to CNNs



RESULTS

TABLE I: PERFORMANCE COMPARISON BASED ON TRAINING METRICS

Metric	Model 1 (Basic CNN)	Model 2 (VGG-19)	Model 3 (ViT-Base)	Performance Comparison
Final Training Loss	0.40	0.21	0.03	ViT lowest by 0.18 (vs VGG-19) and 0.37 (vs CNN)
Final Validation Loss	0.78	0.85	1.17	Basic CNN lowest by 0.07 (vs VGG-19) and 0.39 (vs ViT)
Final Training Accuracy	0.86	0.92	0.98	ViT highest by 0.06 (vs VGG-19) and 0.12 (vs CNN)
Final Validation Accuracy	0.64	0.75	0.76	ViT highest by 0.12 (vs Basic CNN) and 0.01 (vs VGG19)
Training-Validation Accuracy Gap	0.10	0.18	0.23	Basic CNN shows least overfitting
Loss Convergence Rate	Moderate	Fast	Very Fast	ViT converges fastest to minimum training loss
Validation Loss Stability	Moderate fluctuation	Moderate fluctuation	High fluctuation	Basic CNN most stable
Epochs to Reach 70% Val. Accuracy	7	5	2	ViT fastest to reach acceptable accuracy
Parameter Efficiency*	High (0.141)	Low (0.005)	Medium (0.009)	Basic CNN most parameter-efficient

GRAPHS

Fig. 8. Custom Basic CNN model: Training and Validation Loss graph

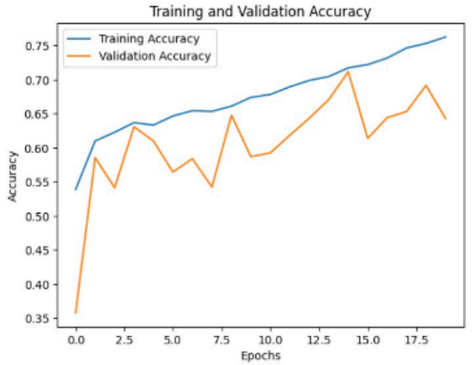


Fig. 9. Custom Basic CNN model: Training and Validation Accuracy graph

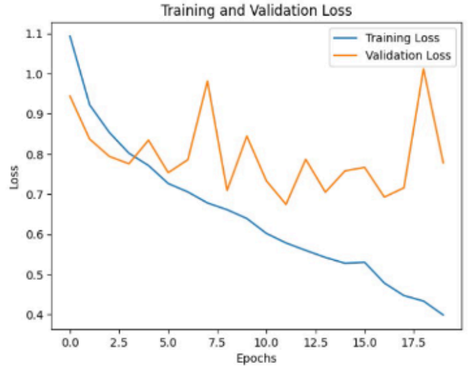


Fig. 10. VGG19 model: Training and Validation Loss graph

Fig. 11. VGG19 model: Training and Validation Accuracy graph

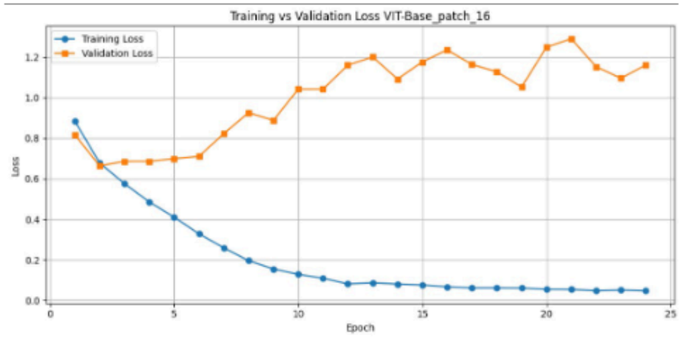


Fig. 18. Train VS Validation Loss : ViT

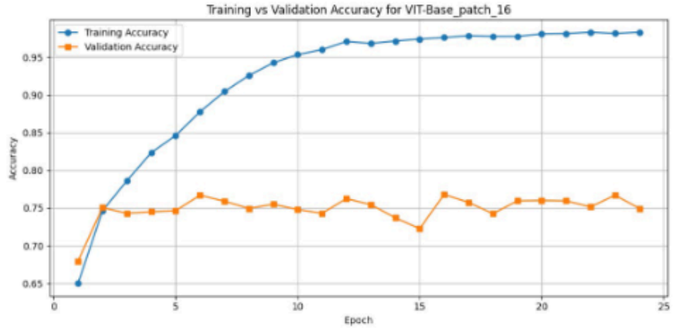


Fig. 19. Train VS Validation Accuracy : Vi

Basic CNN and VGG19

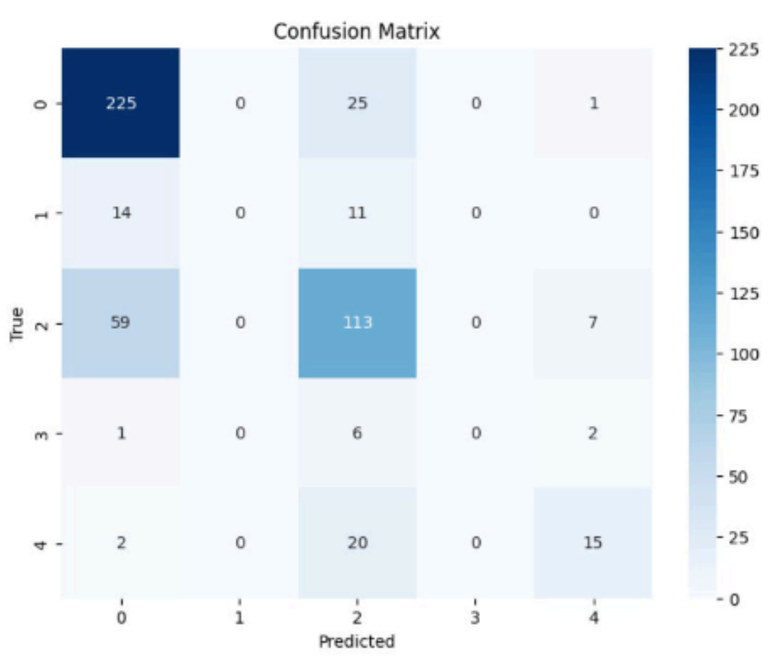


Fig. 12. Confusion Matrix of Basic CNN model

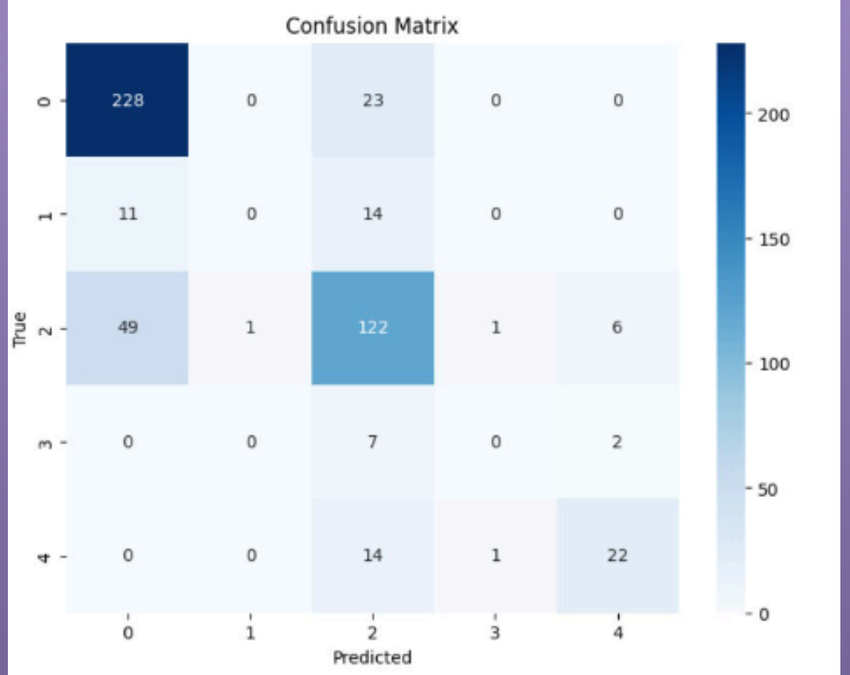


Fig. 13. Confusion Matrix of VGG19 model

	precision	recall	f1-score
0	0.75	0.90	0.82
1	0.00	0.00	0.00
2	0.65	0.63	0.64
3	0.00	0.00	0.00
4	0.60	0.41	0.48

Fig. 14. Precision, Recall and F1 score of Basic CNN model

	precision	recall	f1-score
0	0.79	0.91	0.85
1	0.00	0.00	0.00
2	0.68	0.68	0.68
3	0.00	0.00	0.00
4	0.73	0.59	0.66

Fig. 15. Precision, Recall and F1 score of VGG19 model

TABLE I
PERFORMANCE COMPARISON BETWEEN CUSTOM CNN AND VGG19 MODELS

Metric	Custom CNN	VGG19
Test Accuracy	0.7046	0.7425
Test Loss	0.8077	0.7234
Cohen's Kappa	0.4822	0.5553
Quadratic Weighted Kappa	0.6467	0.7384

THANK YOU