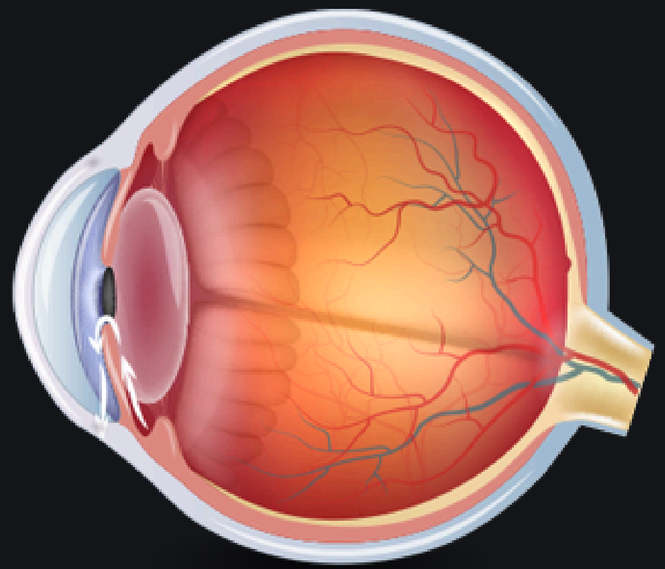
SWEG-Net : Deep Learning Model for Glaucoma Classification

Under the Guidance of : Professor Soumen Bag

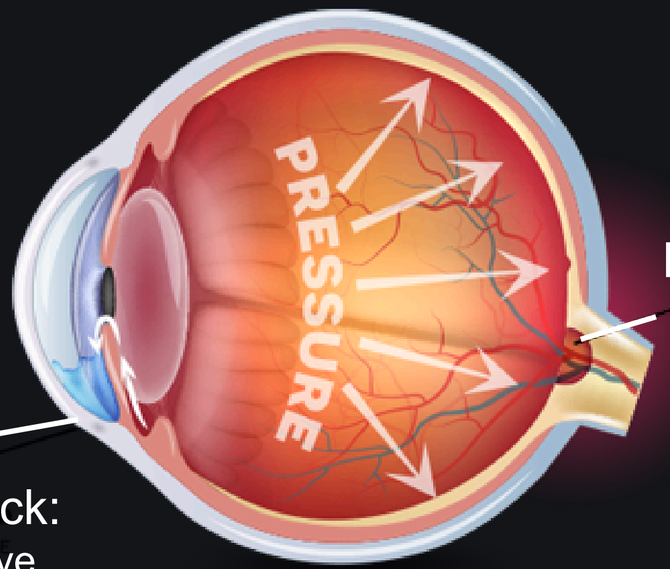
Team Members:
Kamal Anand (21JE0442)
Parth Pandya (21JE0635)
Yashwant Khare (21JE1073)

SWEGNet

INTRODUCTION



Normal Eye



Glaucoma

Glaucoma is a leading cause of irreversible blindness. Early detection is crucial for treatment and vision preservation.

Fundus imaging is a non-invasive, cost-effective screening method.

Our goal: Contribute in development and application of deep learning model to classify glaucoma severity from fundus images.

SWEGNet

Why Deep learning?

- Manual diagnosis is subjective from doctor to doctor, time-consuming, and resource-intensive, requires the expertise of trained ophthalmologists, and is not scalable.
- Deep learning automates feature extraction and classification, removing the need for hand-crafted features and reducing human bias.
- Deep learning has demonstrated superior performance in detecting diseases such as diabetic retinopathy, pneumonia, and breast cancer from radiological images, often comparing to or surpassing expert-level diagnosis.
- Can generalize well with large datasets and proper augmentation, making them suitable for real world application.

SWEGNet

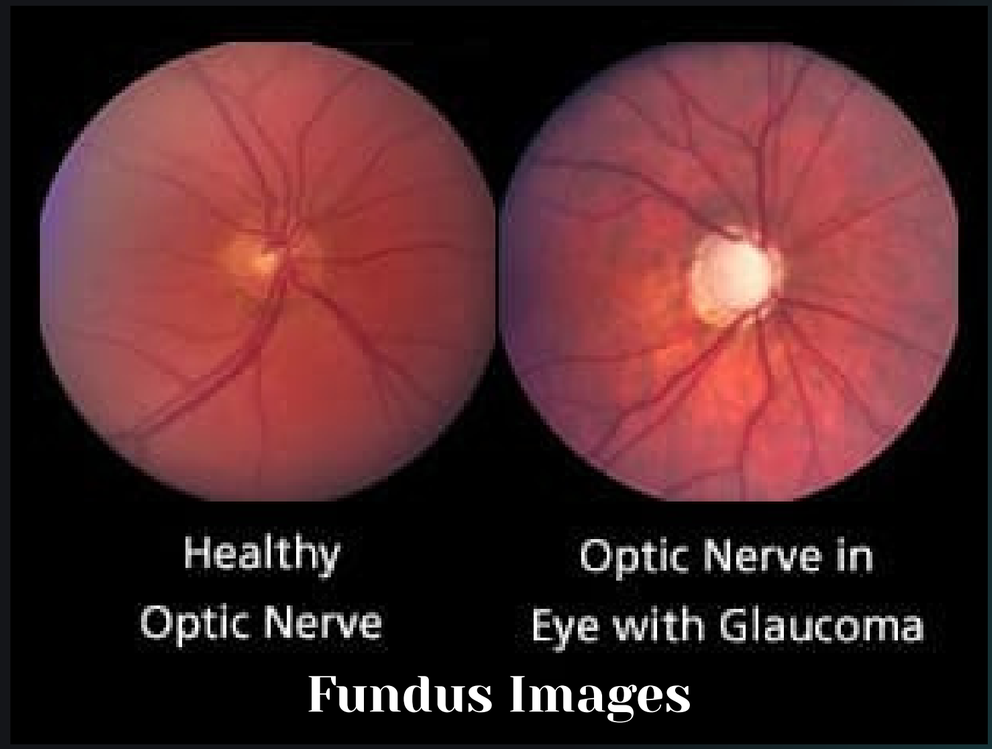
Dataset Used

LMG Dataset-
This is a publiciy availible dataset made by combination of the two:

- HDV1 - Harvard Dataverse version fundus images
- RIMONE dataset

Pre-Processing

- Resizing
- Normalization
- Augmentation (flip, rotate, contrast)



Fundus Images

DataSet 3 Class	Normal	Early	Advanced	Total
HDV1	788	289	467	1544
Rim-One	14	12	14	40
LMG (Combined)	802	301	481	1584

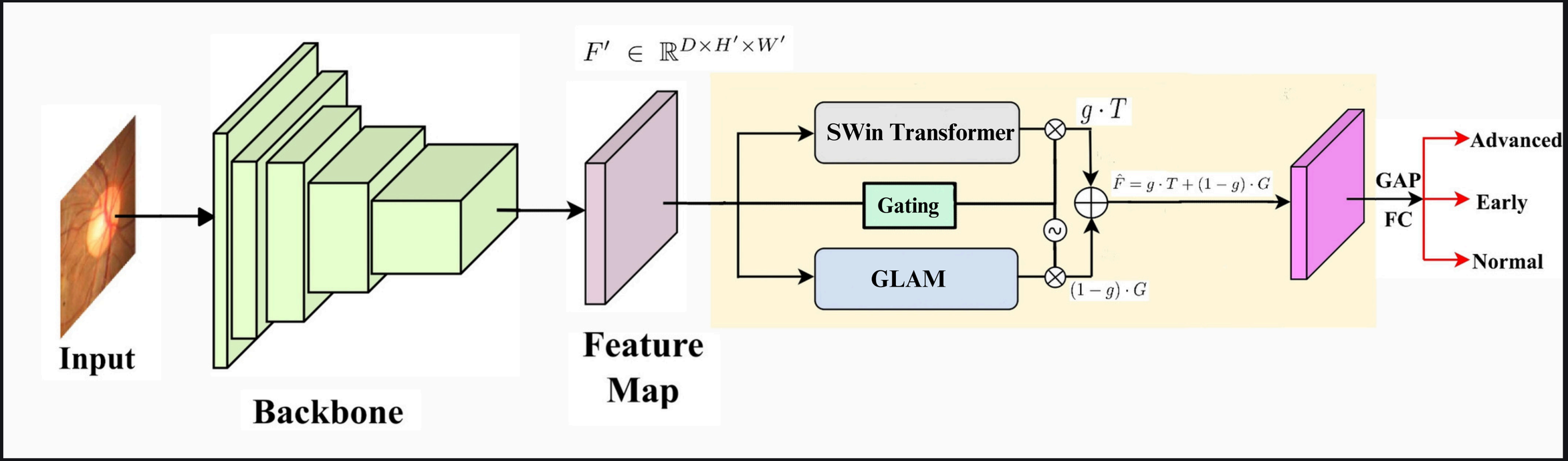
DataSet 2 Class	Normal	Glaucoma	Total
Rim-One R1	118	50	168
Rim-One R2	255	200	455
Rim-One Extended	373	250	623

SWEGNet

Our proposed Model:

SWEGNet: SWin-EfficientNet-Enhanced-GLAM Network, with Learnable Loss

Architecture Overview



SWEGNet

Backbone CNN

GLAM

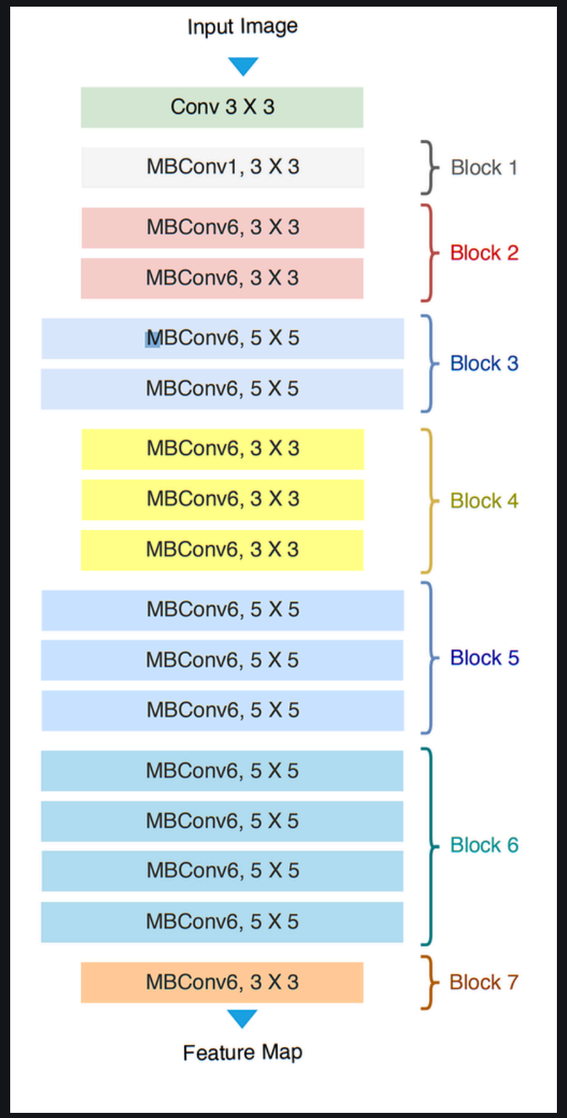
SWin

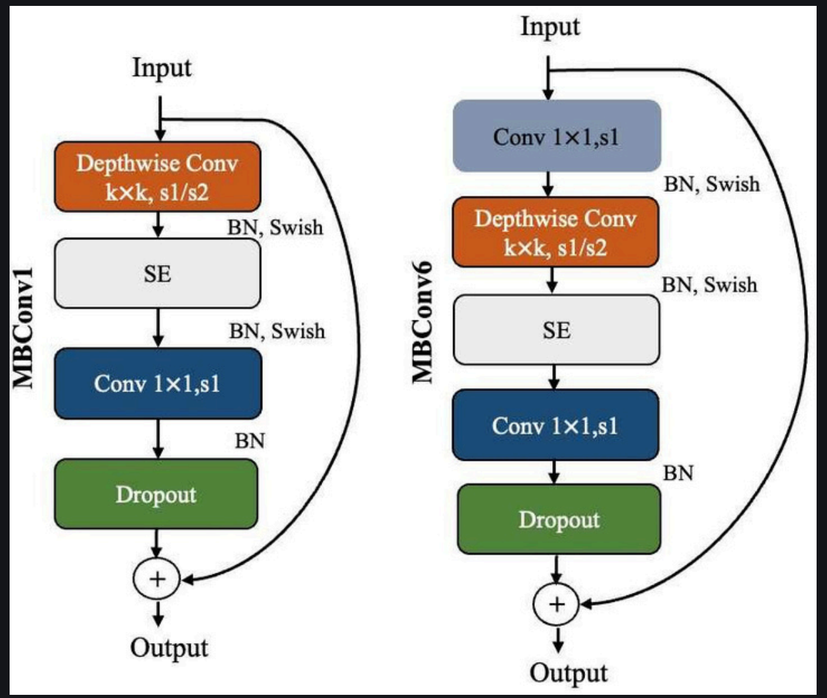
Gating Mechanism

Learnable Loss Function

EfficientNetB0 as Convolutional Neural Network Backbone :

Extracts high-level spatial features from input fundus images.
We Evaluated multiple CNN architectures and found that this is a suitable backbone for our model for its balance between accuracy and computational efficiency.





SWEGNet

Performance Comparison

Different CNN Backbones

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
Densnet169	80.26%	81.39%	80.26%	80.73%	93.01%
Densnet121	79.61%	78.57%	79.61%	78.83%	91.71%
Resnet18	81.23%	80.54%	81.23%	80.80%	93.14%
Resnet50	82.20%	82.05%	82.20%	81.96%	93.63%
MobileNet	79.29%	80.40%	79.29%	79.76%	92.45%
Inception	82.20%	81.22%	82.20%	81.28%	92.01%
EfficientNetB0	**83.50%**	85.20%	83.50%	84.05%	93.50%

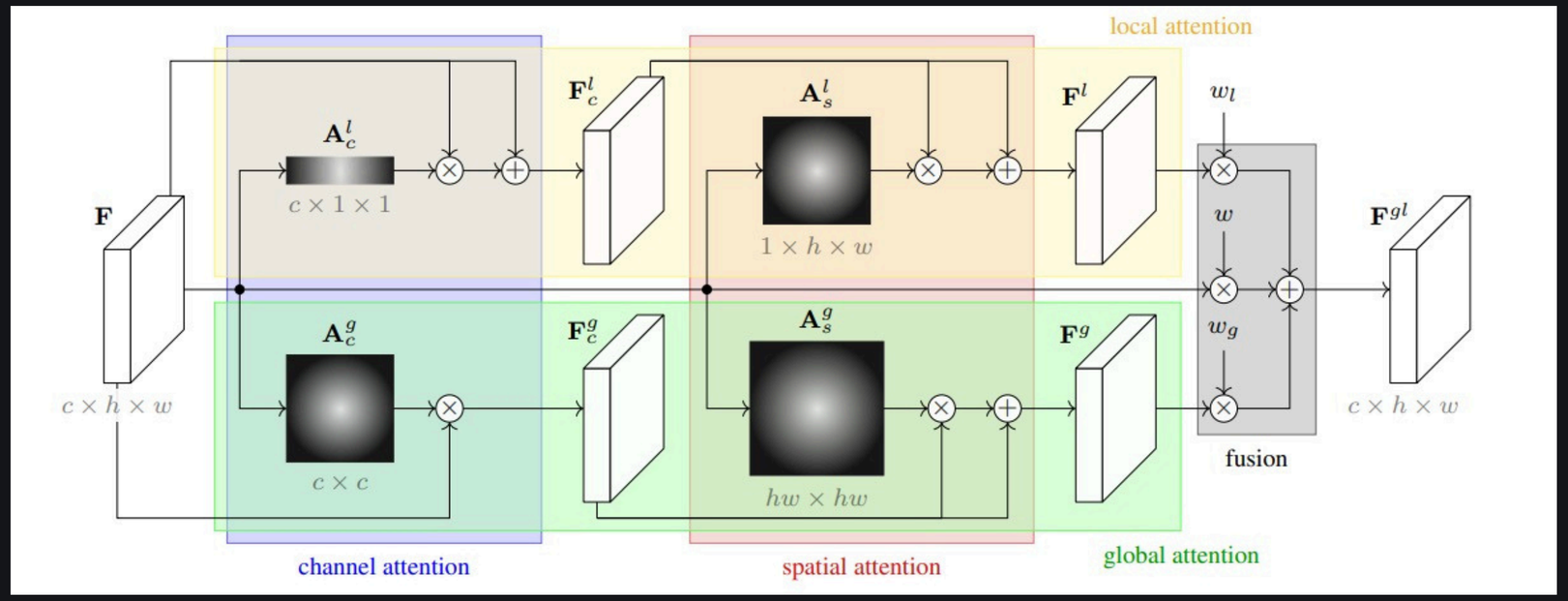
SWEGNet

BackBone CNN | **GLAM** | SWin | Gating Mechanism | Learnable Loss Function

Branch 1: Global-Local Attention Module Pathway

One of the 2 branches after the CNN is the GLAM Module (Global-Local Attention Module), It emphasizes important spatial regions using the concept of attention. Then Combines the channel and spatial attention mechanisms together.

This will help enhance the local texture and the structure relevant to glaucoma progression. by helping the model to focus on optic disc and surrounding nerve fiber layers.



SWEGNet

Performance Comparison

EfficientNetB0 With Attentions

HDV1

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
EfficientNetB0 + CBAM	79.61%	82.58%	79.61%	80.47%	91.90%
EfficientNetB0 + GC	82.52%	82.68%	82.52%	82.52%	93.62%
EfficientNetB0 + TCA	83.17%	83.60%	83.71%	83.20%	94.04%
EfficientNetB0 + CAB	81.23%	81.71%	81.23%	81.40%	92.67%
EfficientNetB0 + GLAM	**83.82%**	83.74%	83.82%	83.78%	93.88%

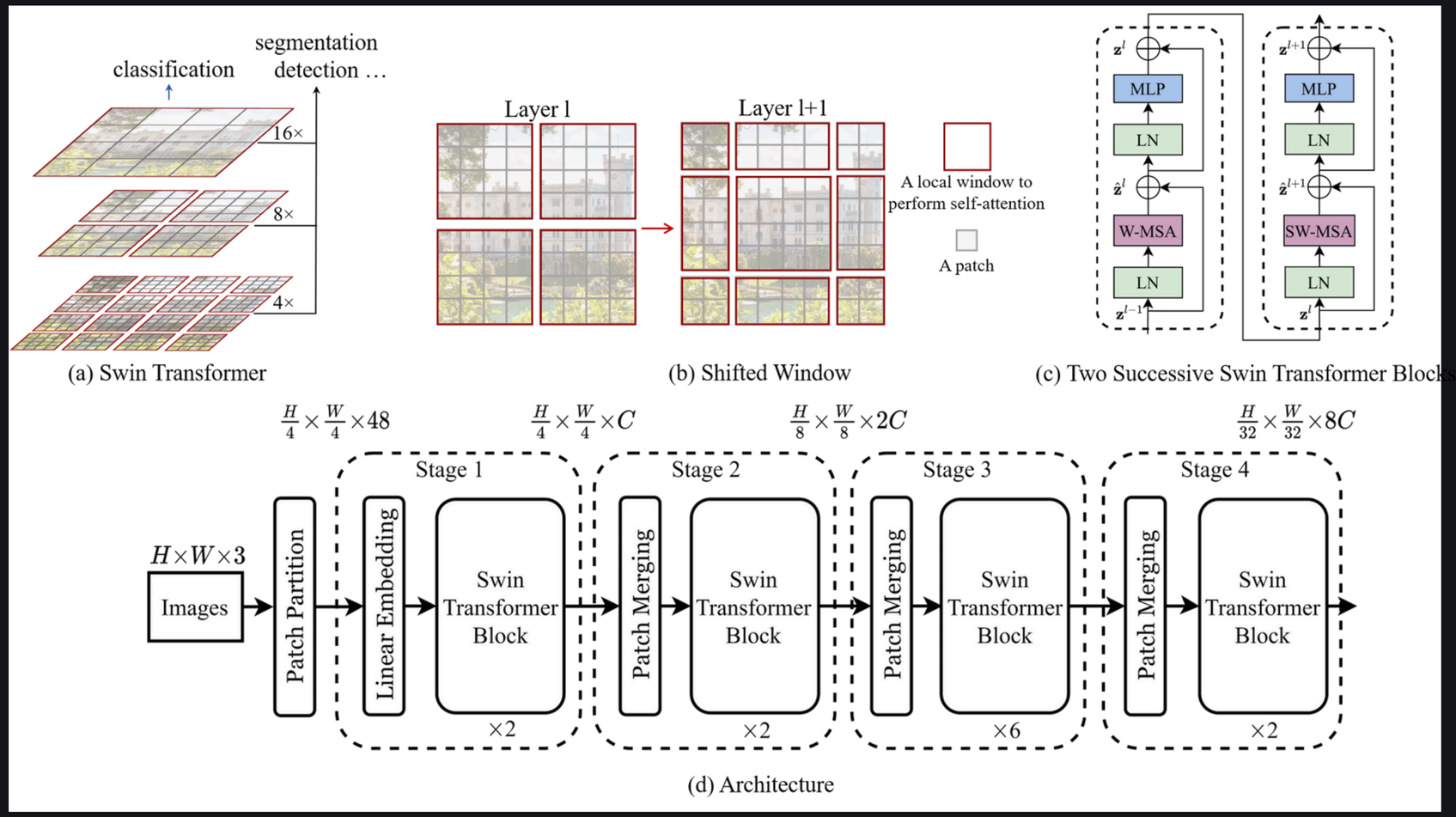
SWEGNet

BackBone CNN | GLAM | **SWin** | Gating Mechanism | Learnable Loss Function

Branch 2: Shifted Window Image Transformer

The second branch parellal to GLAM after the CNN backbone.
It divides feature maps into windows and applies self-attention. While Efficiently capturing global contextual dependencies across the image.

It is Ideal for modeling long-range relationships in fundus images where disease cues may be spatially distant.



SWEGNet

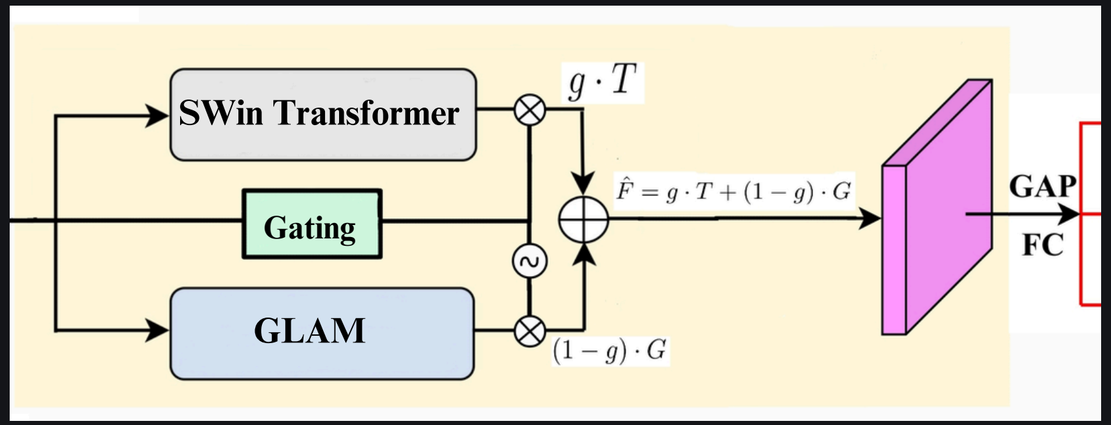
Performance Comparison

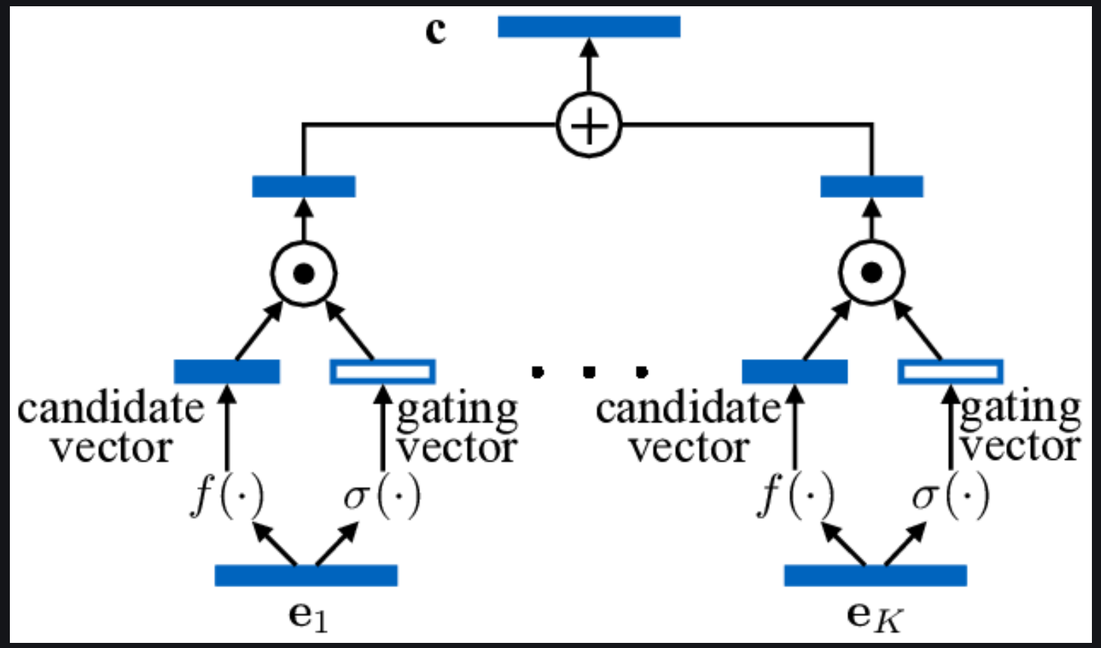
Different Transformers

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
PvtV2	77.35%	77.73%	77.35%	75.52%	91.98%
Twins Svt	82.20%	81.845	82.20%	81.94%	93.09%
mobilevit_s	79.61%	82.99%	79.61%	80.78%	93.99%
CROSSVIT_9_240	82.52%	81.94%	82.52%	82.14%	93.59%
cait_s24_224	77.02%	81.50%	77.02%	78.45%	91.51%
deit_base_patch16_224	82.52%	81.65%	82.52%	81.53%	92.95%
SwinV2	**82.84%**	83.15%	82.84%	82.90%	92.40%

Gated Fusion & Classification

Outputs from SWIN and GLAM pathways are fused via a gated mechanism. This gating layer learns to weigh contributions from both branches.
Then the Final 3-class classification is done through a fully connected layer with softmax activation function.





Learnable Loss Function

A Combination of Focal Loss and Cross Entropy Loss was used in our model.
In this Focal Loss reduces the impact of well-classified examples, focusing learning on harder, misclassified samples. It also plays an important role in handling class imbalance.

Cross Entropy provides a stable gradient flow and is effective for general classification, with smooth curves and easy convergence. Their weighted combination helps balance robustness and sensitivity to minority classes.

The weights for each loss component are treated as learnable parameters and optimized during training with backpropogation. This adaptive approach enables the model to self-tune its loss contribution based on data dynamics, improving convergence and generalization.

$$L_{\mathrm{fl}}(\widehat{y}, y) = \begin{cases} -\alpha(1-\widehat{y})^{\gamma}\log\widehat{y}, & y = 1, \\ -(1-\alpha)\widehat{y}^{\gamma}\log(1-\widehat{y}), & \text{otherwise.} \end{cases}$$

Focal Loss

$$L = -\sum_{k=1}^{K} y_k \log(p_k)$$

Cross-entropy Loss

SWEGNet

Performance Comparison

Loss Function

HDV1

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
CrossEntropy	84.79%	85.58%	84.79%	84.88%	93.82%
Weighted CrossEntropy	81.23%	84.47%	81.23%	82.27%	94.08%
Focal Loss	86.08%	85.56%	86.08%	85.58%	92.50%
Correntropy Loss	84.14%	84.75%	84.14%	84.37%	93.20%
Serial (Focal + Dice)	85.11%	84.55%	85.11%	84.55%	93.84%
Parallel (Focal + Dice + Hinge)	86.41%	85.62%	86.41%	85.52%	93.76%
Learnable Loss	**86.73%**	86.92%	86.73%	86.52%	93.58%

SWEGNet

Performance Comparison

Other CNN Backbones With SWEGNet

	CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
HDV1	Densnet169	82.20%	82.19%	82.20%	82.19%	91.78%
	Densnet121	85.43%	86.12%	85.43%	85.63%	93.05%
	Resnet18	85.11%	84.77%	85.11%	84.65%	93.79%
	Resnet50	82.85%	82.04%	82.85%	82.30%	90.53%
	MobileNet	80.58%	81.29%	80.58%	80.83%	89.27%
LMG	Densnet169	79.49%	78.44%	79.49%	78.44%	90.12%
	Densnet121	79.18%	78.98%	79.18%	79.07%	89.73%
	Resnet18	82.96%	82.65%	82.96%	82.51%	91.98%
	Resnet50	80.13%	80.08%	80.13%	80.08%	88.71%
	MobileNet	80.13%	79.31%	80.13%	79.67%	88.75%

SWEGNet

Performance Comparison

Ablation Studies

HDV1

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
EfficientNetb0 + GLAM	83.50%	83.00%	83.50%	83.08%	94.46%
EfficientNetb0 + Swin	82.52%	81.94%	82.52%	82.14%	93.59%
EfficientNetb0 + GLAM +Swin	85.44%	85.31%	85.44%	83.99%	96.21%
EfficientNetb0 + GLAM +Swin Learnable Loss	85.76%	85.43%	85.57%	85.53%	90.35%
EfficientNetb0 + GLAM +Swin Learnable Loss (Scheduler)	86.73%	86.92%	86.73%	86.52%	93.58%

SWEGNet

Performance Comparison

Ablation Studies

LMG

CNN BackBone	Accuracy (%)	Precision	Recall	F1 Score	AUC Score
EfficientNetb0 + GLAM	84.54%	83.55%	84.54%	83.33%	93.69%
EfficientNetb0 + Swin	82.02%	82.82%	82.02%	82.30%	93.50%
EfficientNetb0 + GLAM +Swin	85.49%	84.61%	85.49%	84.81%	93.05%
EfficientNetb0 + GLAM +Swin Learnable Loss	86.08%	85.57%	86.08%	85.58%	92.50%
EfficientNetb0 + GLAM +Swin Learnable Loss (Scheduler)	86.12%	85.51%	86.12%	85.49%	94.99%

SWEGNet

Performance Comparison

2 Class Classification

CNN BackBone	Accuracy (%)	F1 Score	AUC Score
Baseline Model	87.16%	89.65%	85.76%
SWEGNet	**92.00%**	91.97%	96.43%

SWEGNet

Performance Comparison

State of the Art Models

	CNN BackBone	Accuracy (%)	F1 Score	AUC Score
HDV1	4-Layer CNN	75.64%	75.84%	89.00%
	Customized CNN	78.45%	79.01%	90.97%
	ResNet-50-GAB	82.75%	82.75%	92.59%
	ResNet-50-GAB+CAB	82.75%	82.75%	92.59%
	SWEGNet (Ours)	**86.73%**	86.52%	93.58%
LMG	4-Layer CNN	76.93%	76.02%	90.12%
	Customized CNN	77.35%	75.89%	89.69%
	ResNet-50-GAB	81.97%	81.32%	92.44%
	ResNet-50-GAB+CAB	82.59%	82.42%	92.47%
	SWEGNet (Ours)	**86.12%**	85.49%	94.99%

SWEGNet

Performance Comparison

State of the Art Models

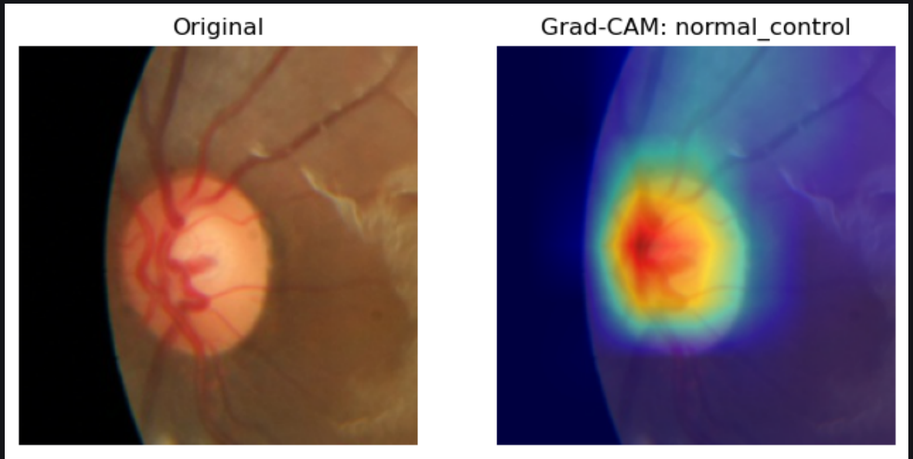
2 Class

CNN BackBone	Accuracy (%)	F1 Score	AUC Score
4-Layer CNN	80.21%	83.70%	79.07%
Customized CNN	79.14%	82.66%	78.18%
ResNet-50-GAB	88.23%	90.23%	87.31%
ResNet-50-GAB+CAB	89.30%	91.22%	88.42%
SWEGNet (Ours)	**92.00%**	91.97%	96.43%

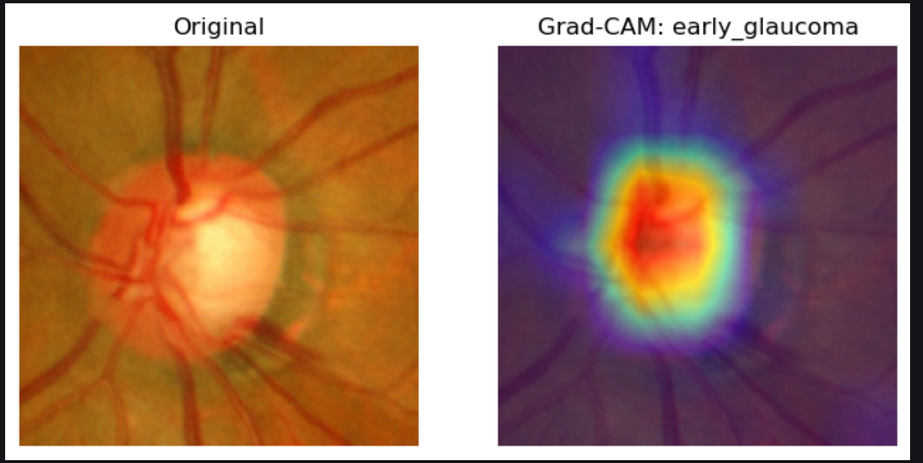
SWEGNet

GradCam Images

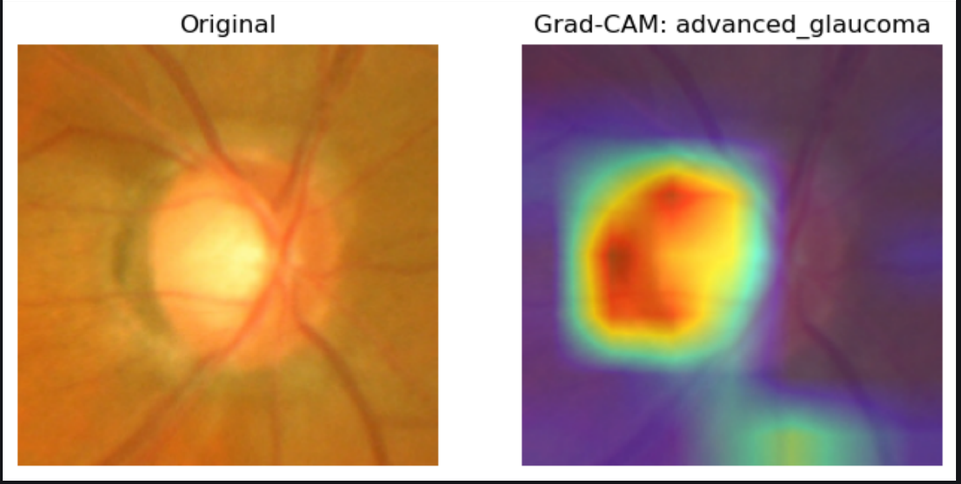
GradCAM heatmaps showing model attention for each class.



Normal



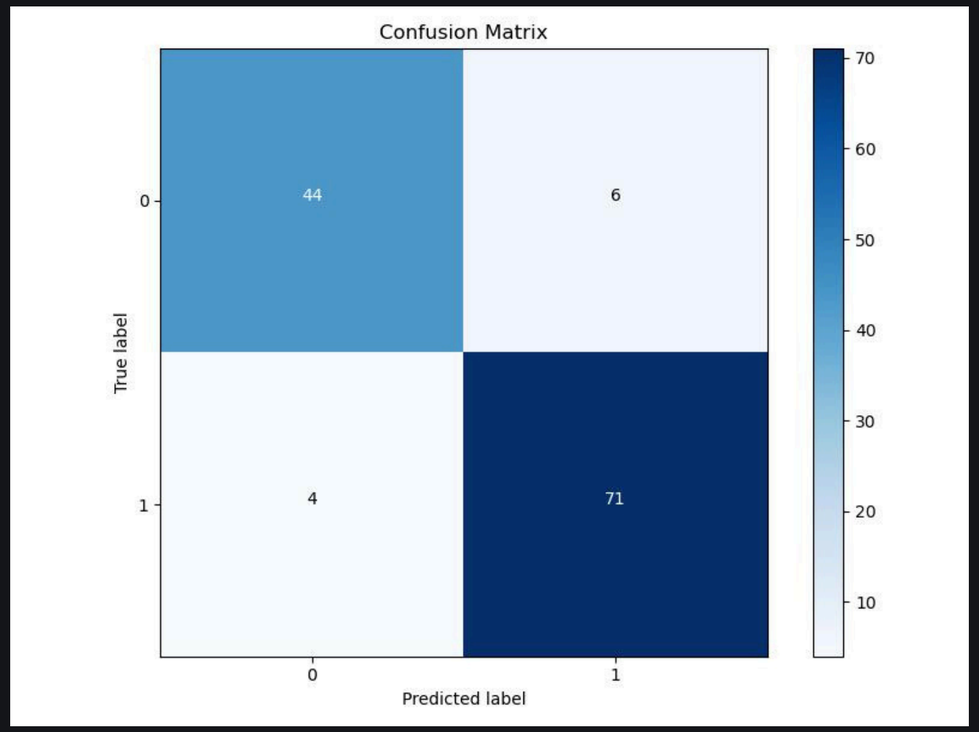
Early



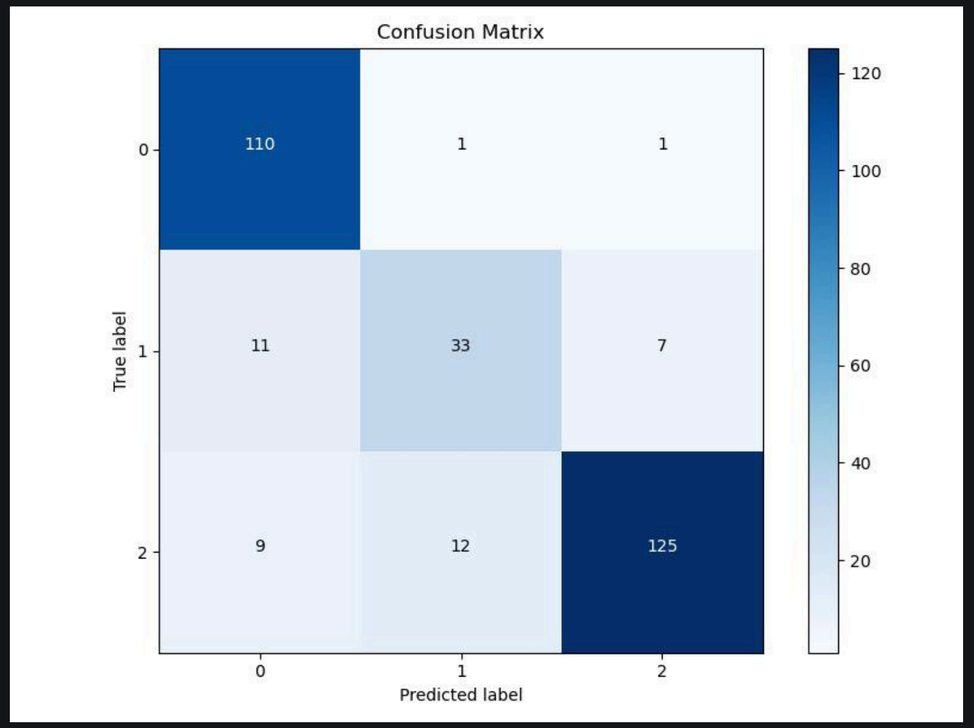
Advanced

SWEGNet

Normalized Confusion Matrices



Normalized confusion matrix for our SWEGNet on the RIMONE Extended test sett

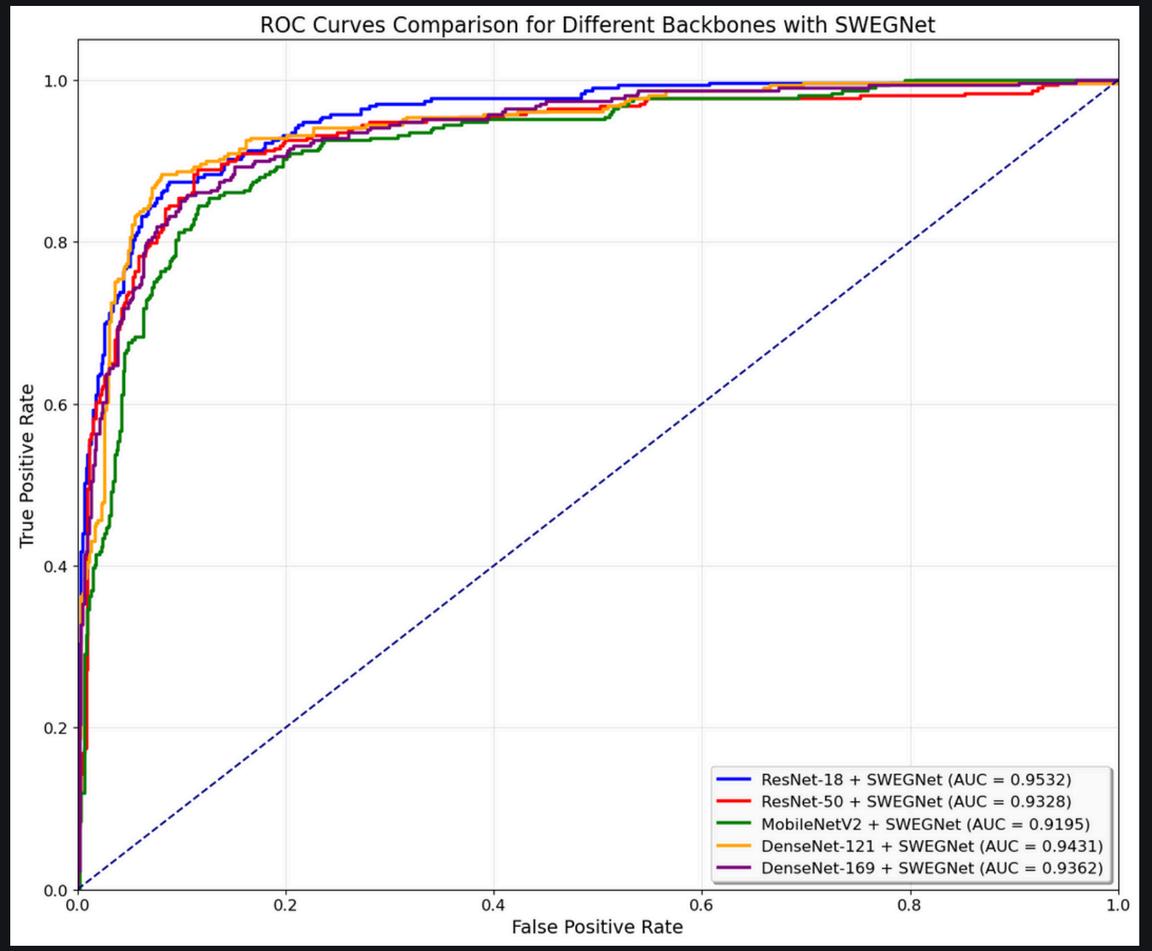


Normalized confusion matrix for our SWEGNet on the HDV1 test set.

SWEGNet

AUROC Curves

ROC Curves per class and macro averages



SWEGNet

Summary & Key Insights

The combination of EfficientNet as the CNN backbone, with SWIN Transformer and GLAM Attention operating in parallel, achieved the highest performance among all tested configurations.

The adaptive loss function, significantly improves performance and boosts model stability. Gated fusion mechanism effectively integrates complementary features from both global and local attention paths.

Overall, our architecture is a strong candidate for aiding automated glaucoma screening in real-world settings with accuracy of 86.73% in 3 class and 92.00% in 2 class classification.

SWEGNet

Thank You!

Kamal Anand(21JE0442) Parth Pandya(21JE0635) Yashwant Khare(21JE1073)