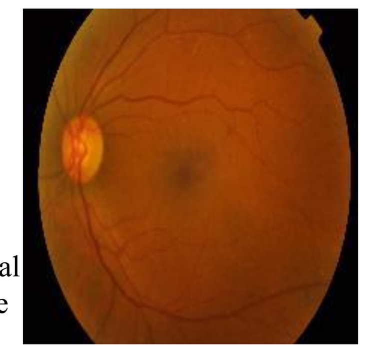
CS 230 Final Project: Diabetic Retinopathy Detection with Convolutional Neural Networks (CNNs): Fine-Tuning ResNet-50, InceptionNet and AlexNet

Lara Seyahi
seyahi@stanford.edu
Stanford University
Department of Computer Science

Introduction & Problem

Diabetic Retinopathy (DR) is a condition caused by damage to small blood vessels that arises from the chronic effects of diabetes. It is one of the leading causes of blindness worldwide. The number of people with DR will grow from 126.6 million in 2010 to 191.0 million by 2030 [1, 2, 3].

DR + Retinal Image

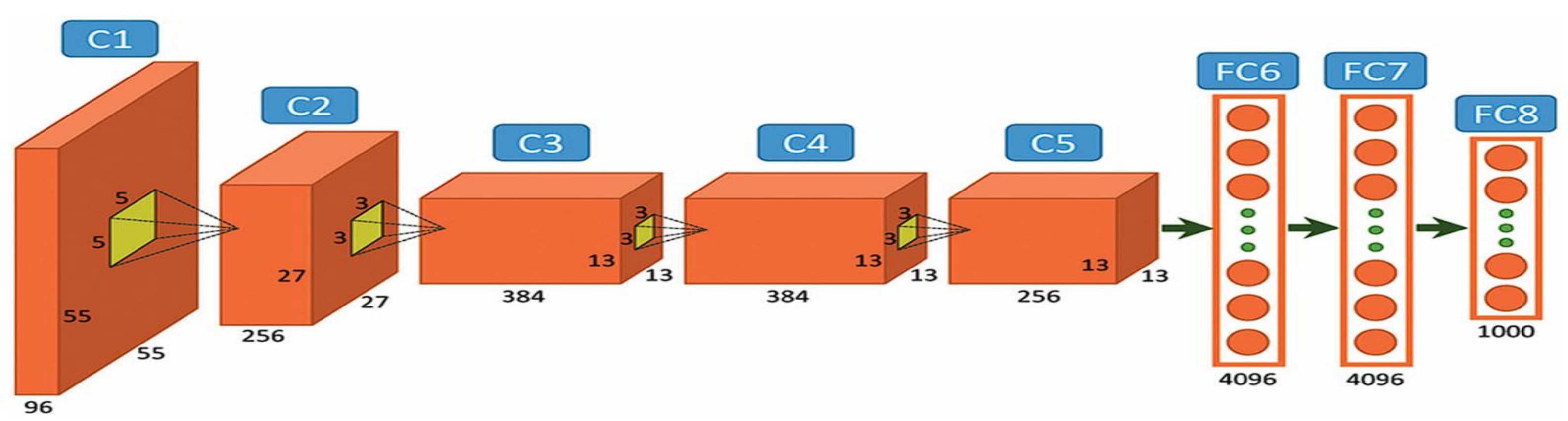


DR - Retinal Image

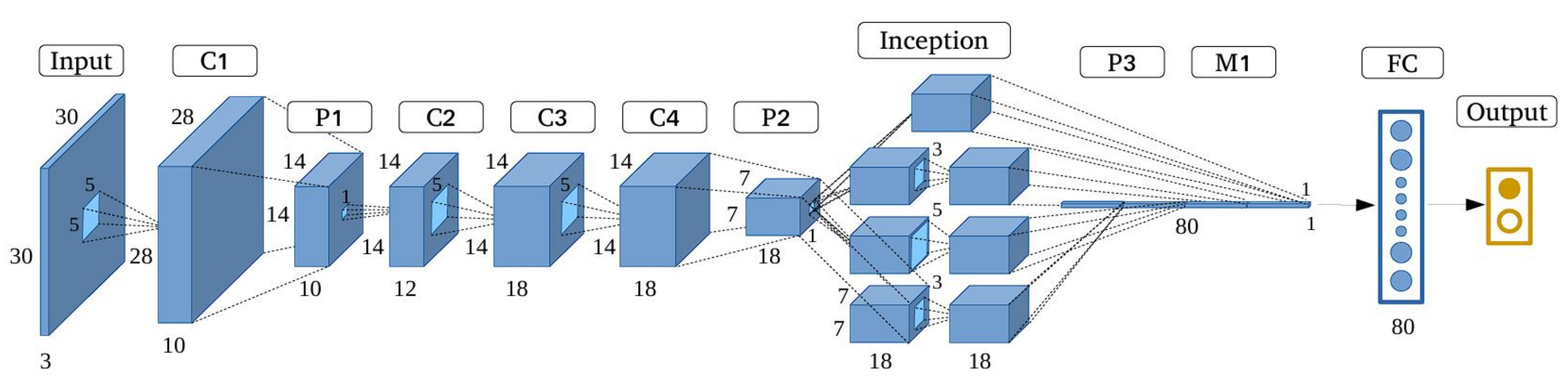
Timely detection of DR is crucial, as early intervention can significantly reduce the risk of severe vision loss. This project aims to develop an AI-based tool for automated DR detection using Convolutional Neural Networks (CNNs), which can improve patient outcomes and reduce the global burden of diabetes-related blindness.

Background: Pre-trained CNN models

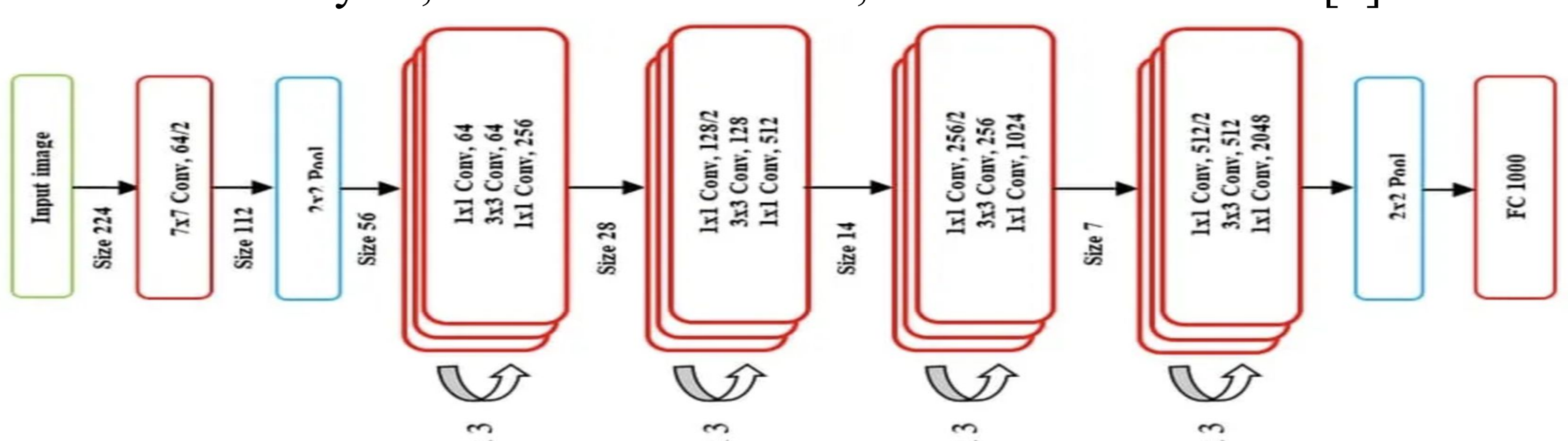
AlexNet is composed of five convolutional layers and three fully connected layers. It uses ReLU activation to speed up training and max pooling for down-sampling [4].



InceptionNet uses Inception modules, which apply multiple convolutional filter sizes (1x1, 3x3, and 5x5) in parallel within the same layer, improving the representation of the input data. It employs 1x1 convolutions and global average pooling [5].



ResNet-50 has 50 layers and introduces residual learning via residual blocks, which allow inputs to bypass certain layers, addressing the vanishing gradient problem in deeper networks. Each block typically includes three convolutional layers, batch normalization, and ReLU activation [6].



Methods

Baseline Models

- **Binary Classification**

AlexNet, ResNet-50, and InceptionNet were fine-tuned. Training was conducted for 5 epochs with a batch size of 32. All layers preceding the final classification layers were frozen. Optimization was performed using the Adam algorithm. **Binary Cross Entropy Loss** is used as the loss function.
ResNet-50 & InceptionNet: The top layers were removed by setting the *include_top=False* parameter in TensorFlow. An average pooling layer, originally part of the removed top layers, was added back. The final layer was replaced with a single unit with **sigmoid** activation function.
AlexNet: The number of output classes in the last layer was reduced from 1000 to 2.

- **Multiclass Classification**

ResNet-50: was selected due to its high performance in the binary classification task. Each experiment was trained for 10 epochs. Final layer was replaced with a fully connected layer containing 5 classes, with **softmax** activation function. **Categorical Cross Entropy loss** function is used:

$$logloss = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\log(p_{ij})$$

M: # of classes
N: # of samples

Early Stopping callback was added to stop training when the validation loss stopped improving.

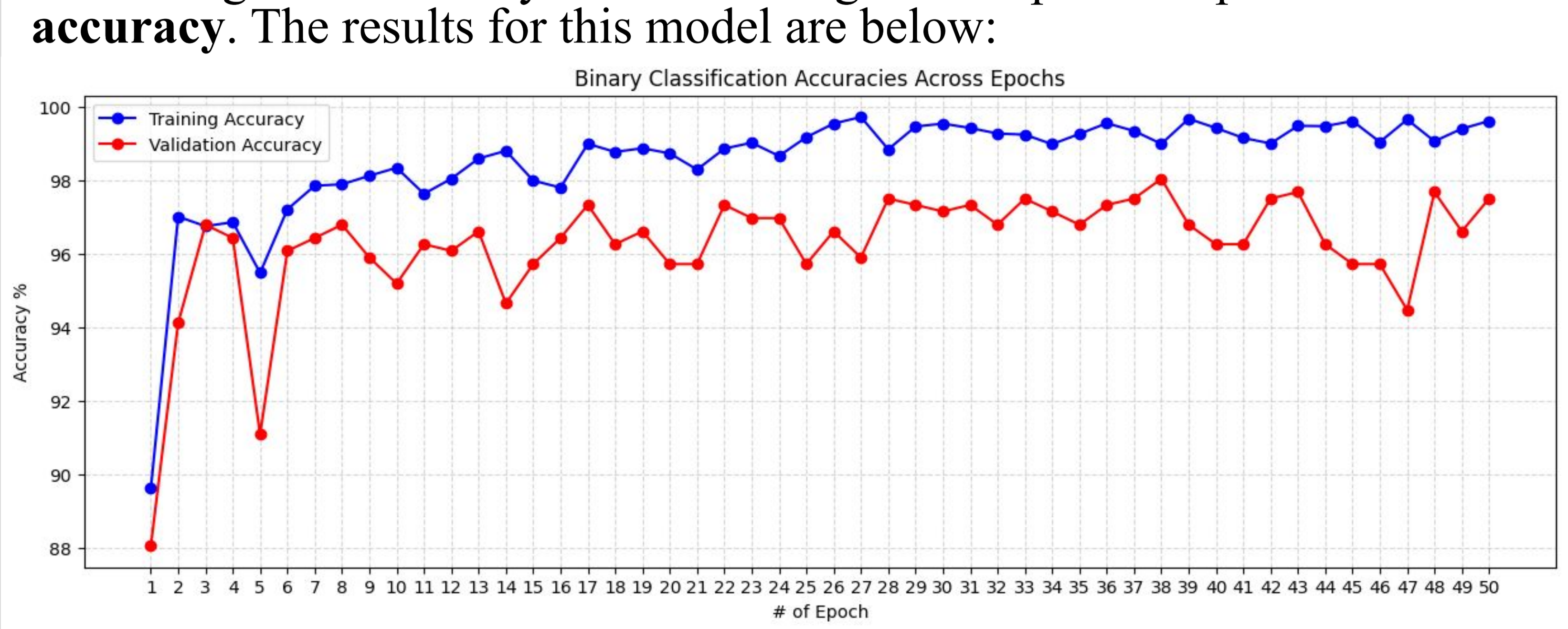
Binary Classification Experiments & Analysis

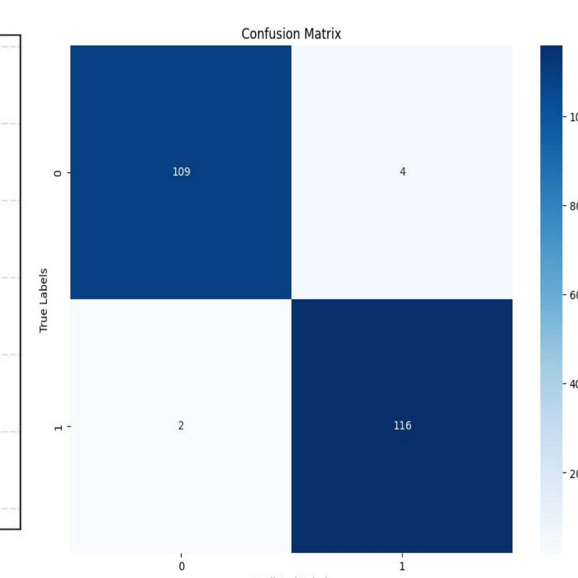
Input: Retinal images
Output: DR + (1), DR - (0)

Model	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
ResNet-50	0.1378	95.43%	0.1405	95.90%
InceptionNet	0.1602	95.40%	0.3245	84.14%
AlexNet	0.2045	95.32%	0.1728	94.65%

Table 1: Baseline Model Results (Epoch 5) for Binary Classification

Fine-Tuning Models:
ResNet-50: A fully connected (FC) layer with 128 units and ReLU activation before the final classification layer was added (modified baseline). Further experiments were done:
Experiment A: Adding another dense layer (128 units with ReLU)
Experiment B: Increasing FC layer size to 256 units
Experiment C: Adding a dropout layer (p = 0.5)
The modified baseline achieved the highest validation accuracy. Fine-tuning this mode; further by unfreezing the last 10 layers and training for 50 epochs improved results further: **97.4% test accuracy**. The results for this model are below:





InceptionNet: In the baseline model, overfitting was encountered.

- A **dropout layer** with dropout probability of 0.5 was added as a regularization technique:

Accuracy: 94.29%, Val Accuracy: 85.74% *overfitting continues*

- A **dense layer with 128 units with ReLu activation** below the dropout layer:

Accuracy: 94.51%, Val Accuracy: 91.27% *increased complexity solved the problem of overfitting*

- Increasing the number of units in the dense layer from 128 to 256:

Accuracy: 95.58\%, Val Accuracy: 85.74% *slightly increased test accuracy, increased overfitting*

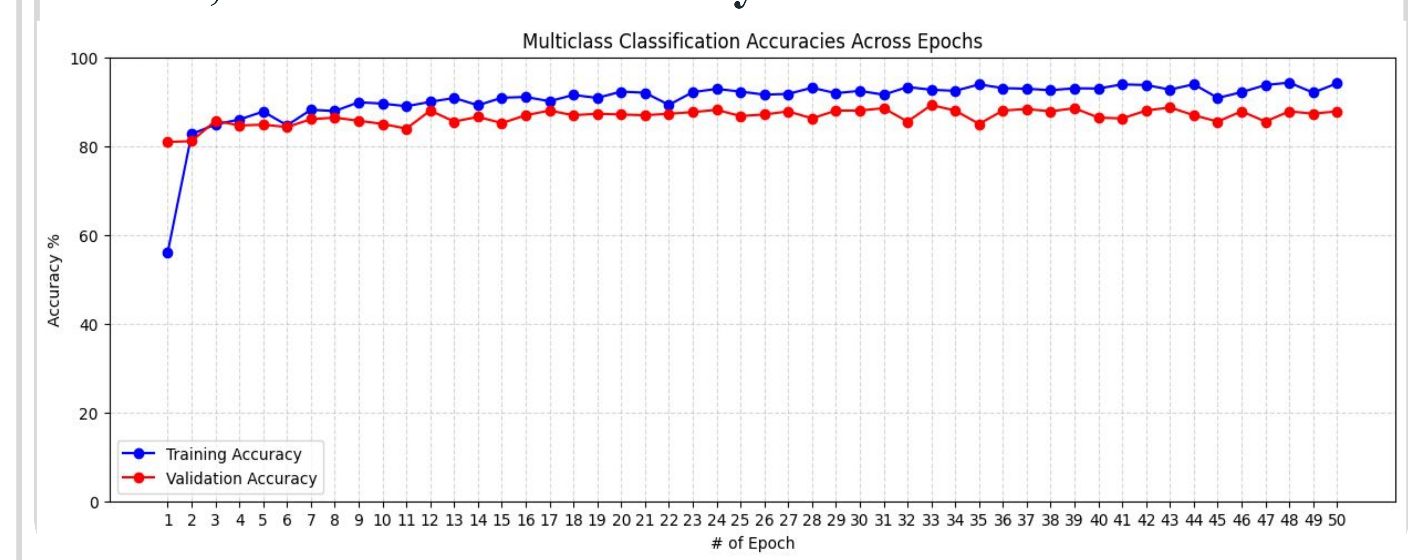
AlexNet: Adding a fully connected layer with 128 units using ReLu activation function resulted in improved performance (modified baseline). The 3 experiments applied to the ResNet-50 model were also applied. Similarly, The modified baseline achieved the best performance.

Multiclass Classification Experiments & Analysis

ResNet-50 was fine-tuned with the following experiments:

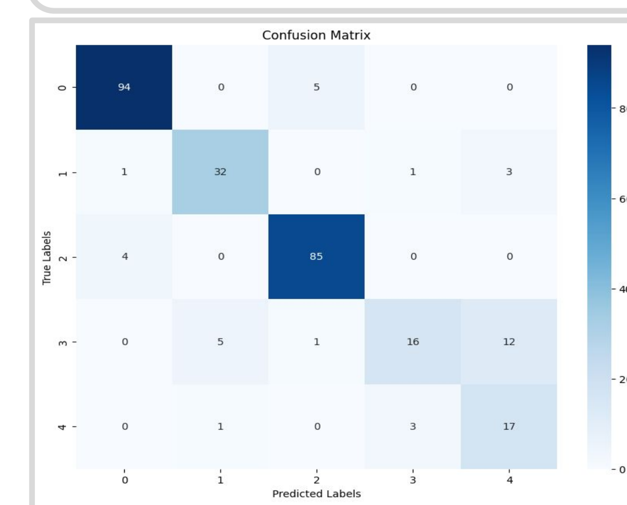
Experiment	Accuracy	Loss	Validation Accuracy	Validation Loss
Baseline Model	0.7900	0.5493	0.7261	0.7069
Baseline with Dropout & Dense	0.7111	0.7698	0.5053	1.8318
Augmentation (Flip Only)	0.8799	0.2823	0.8640	0.3445
Augmentation (All Techniques)	0.9060	0.2590	0.8604	0.3344
Fine-tuning Last 10 Layers	0.8728	0.2864	0.7297	2.2479

Input: Retinal Images
Output: Healthy (0), Mild DR (1), Moderate DR (2), Proliferate DR (3), Severe DR (4)
Baseline Model: Moderate accuracy, but faced overfitting
Increased Model Complexity: Adding dense with 256 units and a dropout layer increased overfitting and decreased performance.
Data Augmentation: Applying random flips, brightness/contrast adjustments, and random cropping on underrepresented classes improved performance
Freezing Layers: Freezing the last 10 layers led to overfitting and reduced validation accuracy.
Using **the baseline model with all data augmentation techniques** and training for 50 epochs yielded the best results, with **85% test accuracy**. The results for this model:



Conclusions

Binary classification: ResNet-50 achieved the highest accuracy after fine-tuning. InceptionNet faced overfitting, AlexNet performed well with minimal changes.
Multiclass classification: ResNet-50's performance improved through fine-tuning.
These findings suggest that ResNet-50 can be potentially used as a diagnostic tool for Diabetic Retinopathy.
For future work, models like DenseNet, larger datasets, and advanced transfer learning techniques will be explored.



Multiclass Classification Final Model Confusion Matrix

References
[1] Fung TH, Patel B, Wilmot EG, Amoaku WM. Diabetic retinopathy for the non-ophthalmologist. Clin Med (Lond). 2022 Mar;22(2):112-116. doi: 10.7861/clinmed.2021-0792. PMID: 35304370; PMCID: PMC8966825.
[2] Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. Indian J Ophthalmol.2012;60:428.
[3] Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35:556–64.
[4] "AlexNet: A Revolutionary Deep Learning Architecture." *Viso.ai*, Viso Technologies, https://viso.ai/deep-learning/alexnet/.
[5] "GoogleNet Explained: The Inception Model That Won ImageNet." *Viso.ai*, Viso Technologies,https://viso.ai/deep-learning/googlenet-explained-the-inception-model-that-won-imagenet/.
[6] "ResNet: Residual Neural Network." Viso.ai, Viso Technologies, https://viso.ai/deep-learning/resnet-residual-neural-network/.