Diabetic Retinopathy Detection

Group 13

Problem Description

Diabetic retinopathy occurs when the damaged blood vessels leak blood and other fluids into the retina, causing swelling and blurry vision.

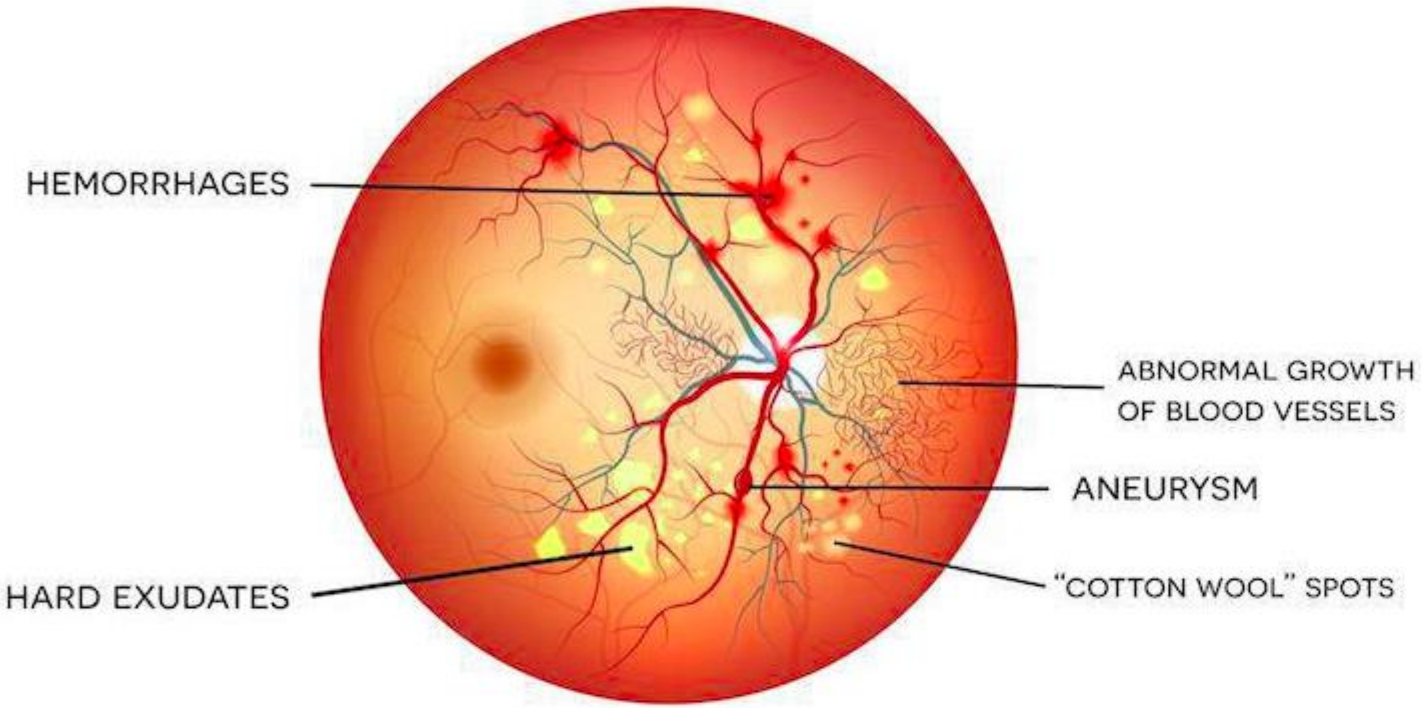
There are 5 stages of DR: **No, Mild, Moderate, Severe, Proliferative**

Ophthalmologists analyze these retina images to look for markers that give evidence for DR.

- 2D Fundus images
- OCT scans

Retina Fundus

This is a schematic showing the key features that signify Diabetic Retinopathy. We want to train a Neural network to look for these features.



Retina Fundus

This is a retina of a subject with PDR. We can observe features like exudates, wool spots, haemorrages etc. While subtle features like microaneurymsns are hard to observe.

Optical Coherence Tomography(OCT)

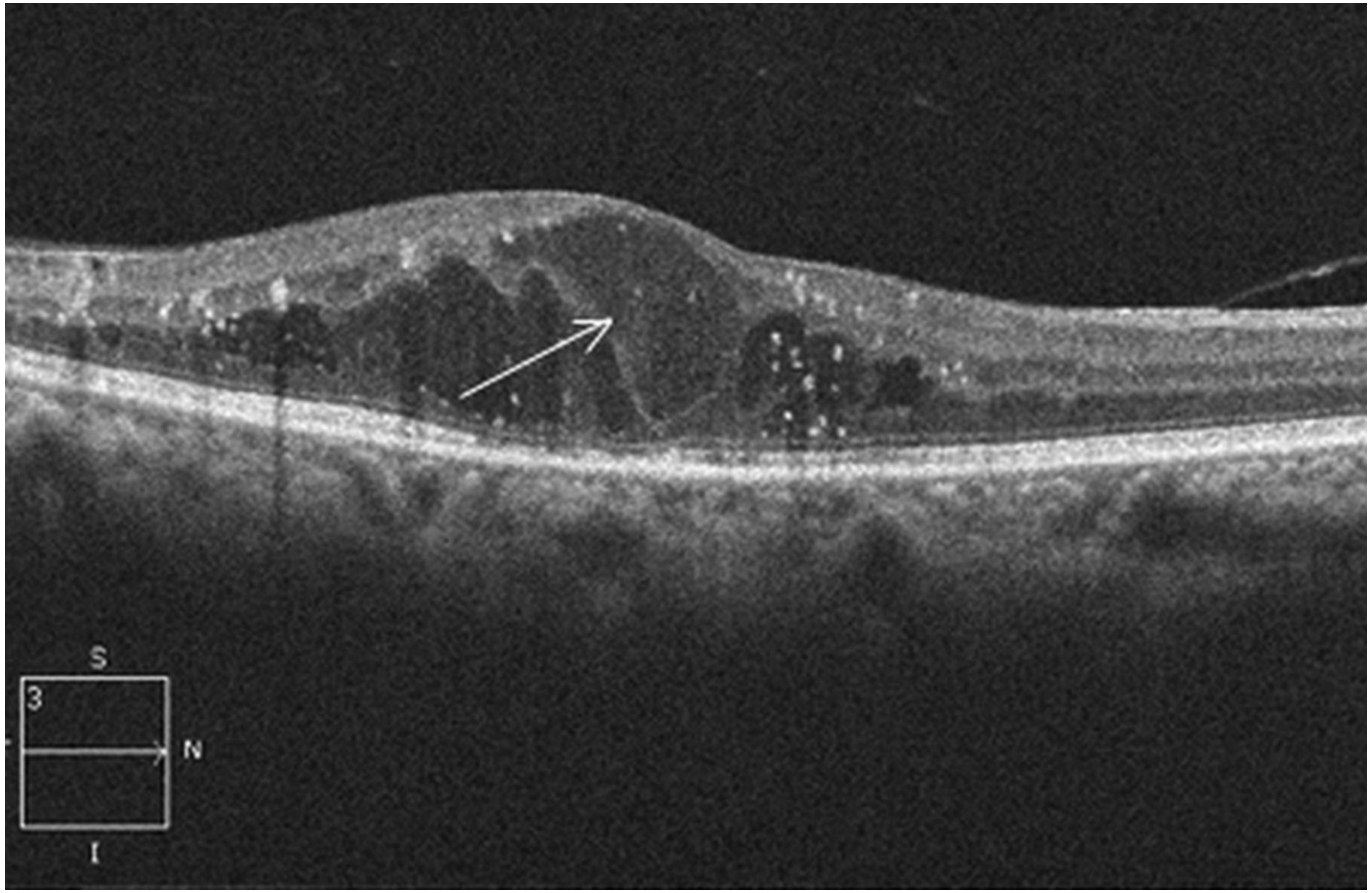
This is an imaging techniques that takes **cross-sectional** images of retina. It has an advantage over fundus images because of its depth information.

Ophthalmologists look for these layers and measure their thickness, curvature etc to diagnose DR.

Decreasing retinal thickness and optical reflectivity are significant biomarkers for detecting DR changes with OCT

Optical Coherence Tomography(OCT)

OCT appearance of diffuse retinal thickening (DRT) in diabetic patients.
Segment different layers and find key features that discriminate between DR and non DR.



Approaches Studied

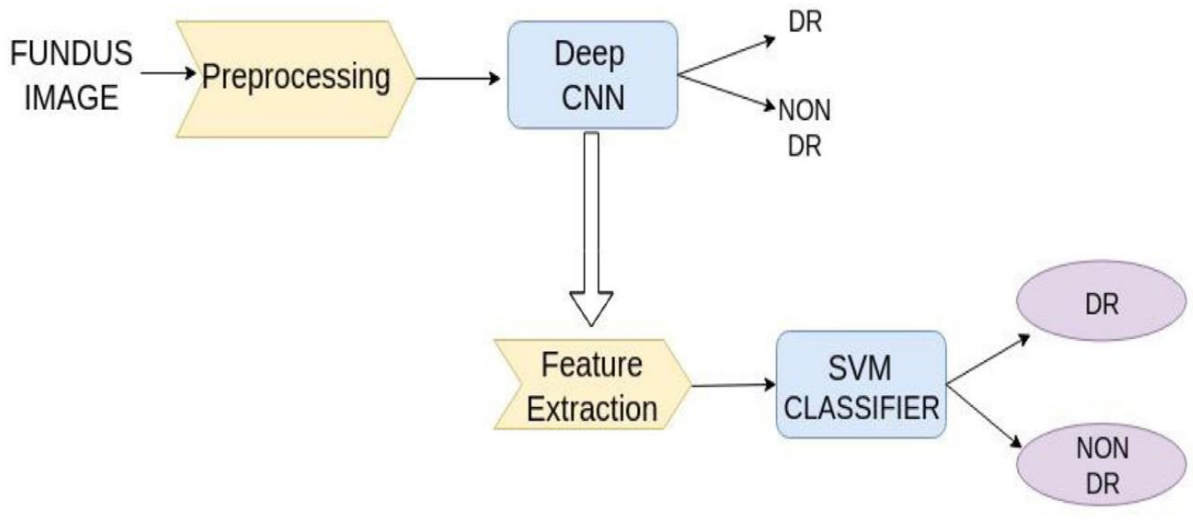
- CNN to extract features and SVM to classify images
- ResNet-50, GooGleNet, AlexNet for n-ary classification of fundus images
- Using CNN along with RAM for discriminative **localisation** and better **visual** explanation.
- Using Segmentation and DFCN(Autoencoders) for OCT images.

Hybrid Deep learning Model for classification

- EYEPACS dataset from Kaggle is used.
- Images are rotated, mirrored and color corrected for data augmentation.
- Accuracy of CNN + SVM similar to CNN + Softmax , increase in specificity and sensitivity in case of hybrid model.
- Less Misclassification.

Model	Sensitivity	Specificity
CNN + Linear SVM	0.93	0.85
CNN + Softmax	0.87	0.67

Hybrid Deep learning Model for classification



- CNN similar to VGGNet used for automatic feature extraction.
- 1024 sized vector provided to SVM for classification.
- SVM with linear kernel was used.

Model Description

- Input Size 512*512, 3*3 kernel size
- Number of filters double after every 2 conv layers from 32 to 512
- MaxPooling is done after every 3rd layer over a 2*2 window with stride 1
- SGD is used with 0.4 dropout for training and regularization
- GridSearch is used for hyperparameter tuning and 7 fold cross validation is used for evaluating the model

CNN+RAM for localization/visualization

- This paper proposes a deep learning-based method for **interpretable** Diabetic Retinopathy detection.
- Deep learning-based feature extraction works better than conventional methods like Hough transform etc.
- Using CNN with only convolutional and pooling layers and not using any dense layers which decreases the parameters a lot.

CNN+RAM for localization/visualization

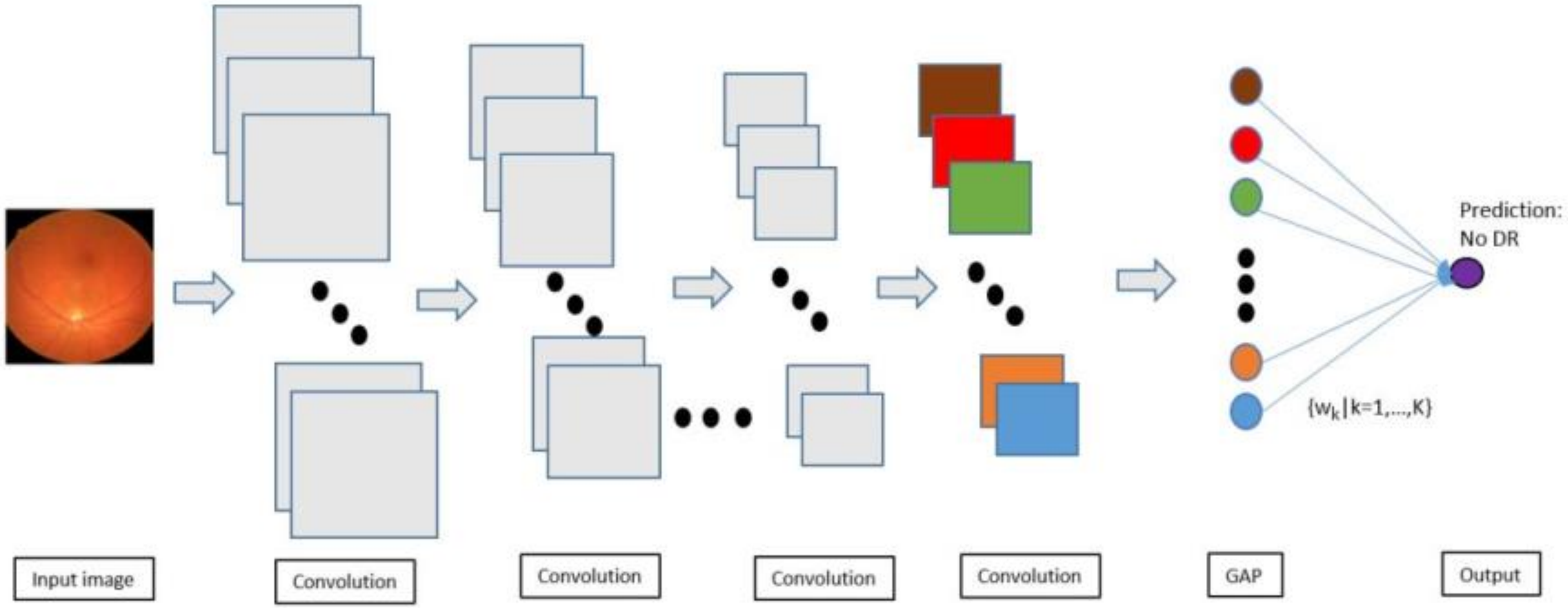
RAM

- By adding the Regression Activation Map (**RAM**) after the global averaging pooling layer of the convolutional networks (CNN) we can highlight the regions of interest through which we identified the severity of the disease.

- RAM of an input image to localize the discriminative regions or point out the regions of interest towards the regression outcomes

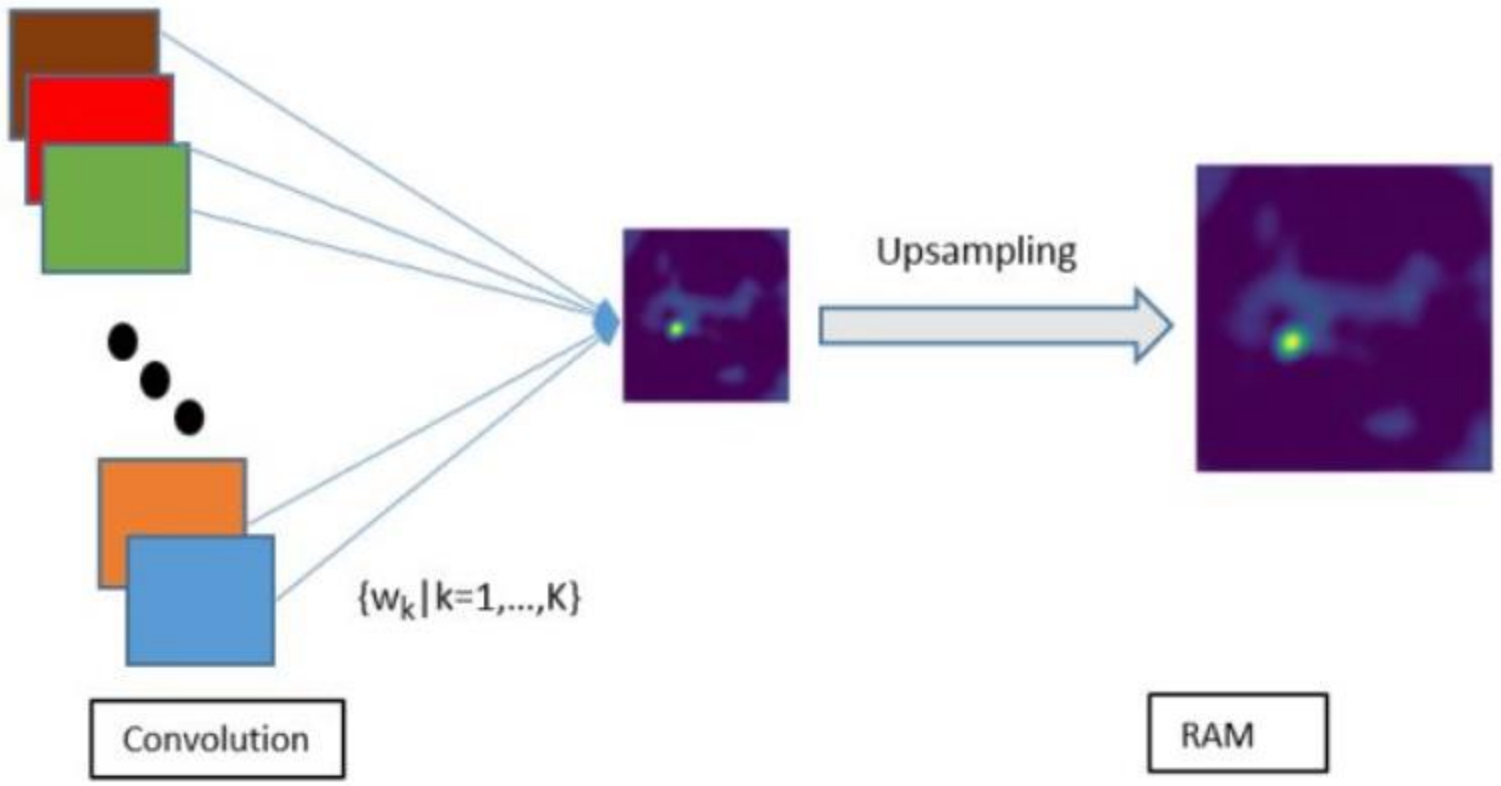
CNN+RAM for localization/visualization

Architecture



CNN+RAM for localization/visualization

Architecture



GAP layers perform an extreme type of dimensionality reduction, where a tensor with dimensions h×w×d is reduced in size to have dimensions 1×1×d.

Activation maps in the final convolutional layer before GAP layer acts as a detector of different local pattern in the image.

CNN+RAM for localization/visualization

Methodology

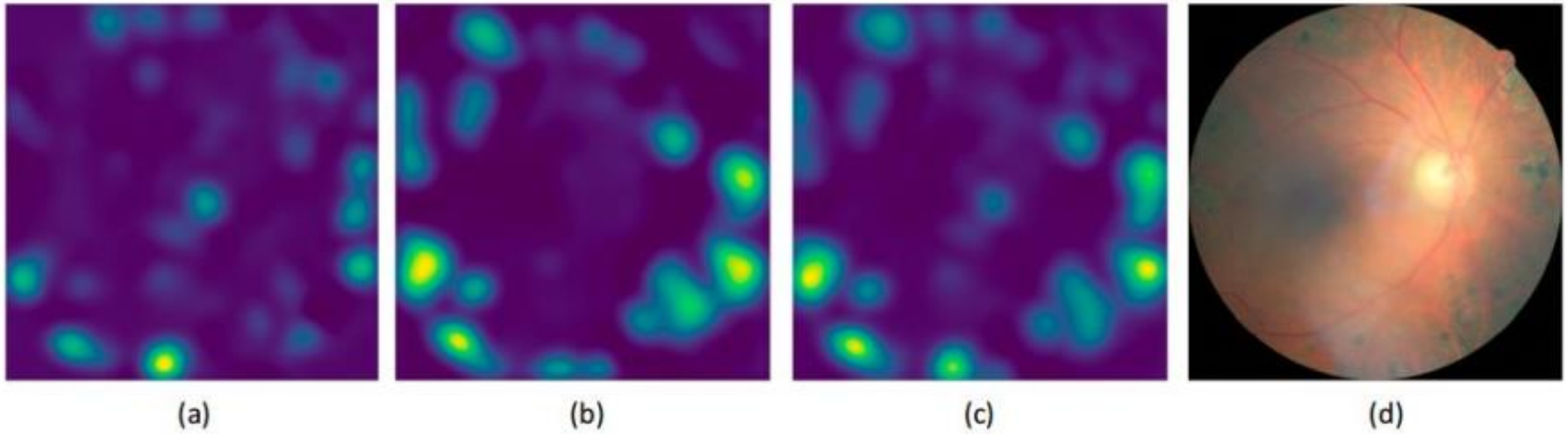
- Pretraining the weights and biases using smaller image 128 pixels on a smaller network. Medium network trained on 256 pixel images. Finally the large network trained on 512p.
- Data augmentation was used (translation, rotation, flipping etc.)
- Leaky ReLu and trained using SGD with momentum.
- Loss function is Mean squared error since regression problem.
- Thresholds to discretize regression values to obtain integer levels.

CNN+RAM for localization/visualization

Results

Performance statistics of the benchmark and our approaches on test dataset.

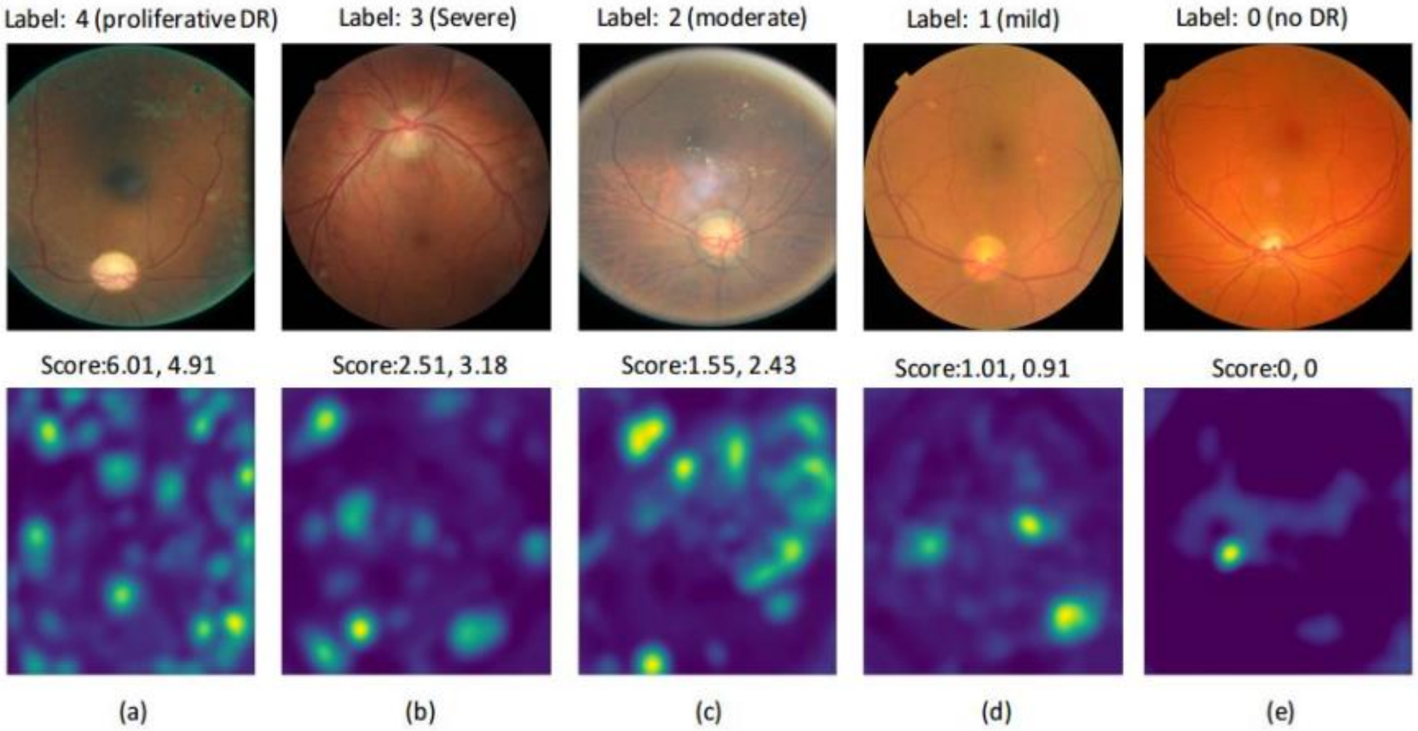
	Baseline	Ours
Kappa score (Public Leaderboard)	0.8542	0.85034
Kappa score (Private Leaderboard)	0.8448	0.8412
Parameter # (net-5)	12.4M	9.7M
Training time (second/epoch)	422.1	367.3
Parameter # (net-4)	12.5M	9.8M
Training time (second/epoch)	451.7	398.2
RAM	No	Yes



Taking the average of RAM of 128 and 256 pixel images gives a better RAM which represents the ground truth better.

CNN+RAM for localization/visualization

Results and Conclusion



We can see that the RAM provides a overall good method to display Region of Interest and also indicating the degree of DR.

CAD for OCT images

This system does the automatic diagnosis in three steps:

1. **Segmenting** and Localizing an OCT image into 12 layers.
2. Individual layer is characterized by a 3 features: **Reflexivity**, **Curvature** and **Thickness** in terms of their Cumulative Probability distributions.
3. A pre trained Deep fusion classification network(**DFCN**) finds the most discriminating features.

CAD for OCT images

- Fundus images are more common at present for automated detection but they have a significant drawback. These images have no information about **depth** and interactions b/w different layers
- Thinning retinal thickness and decreasing optical reflectivity are significant biomarkers for detecting DR changes with OCT.

CAD for OCT images

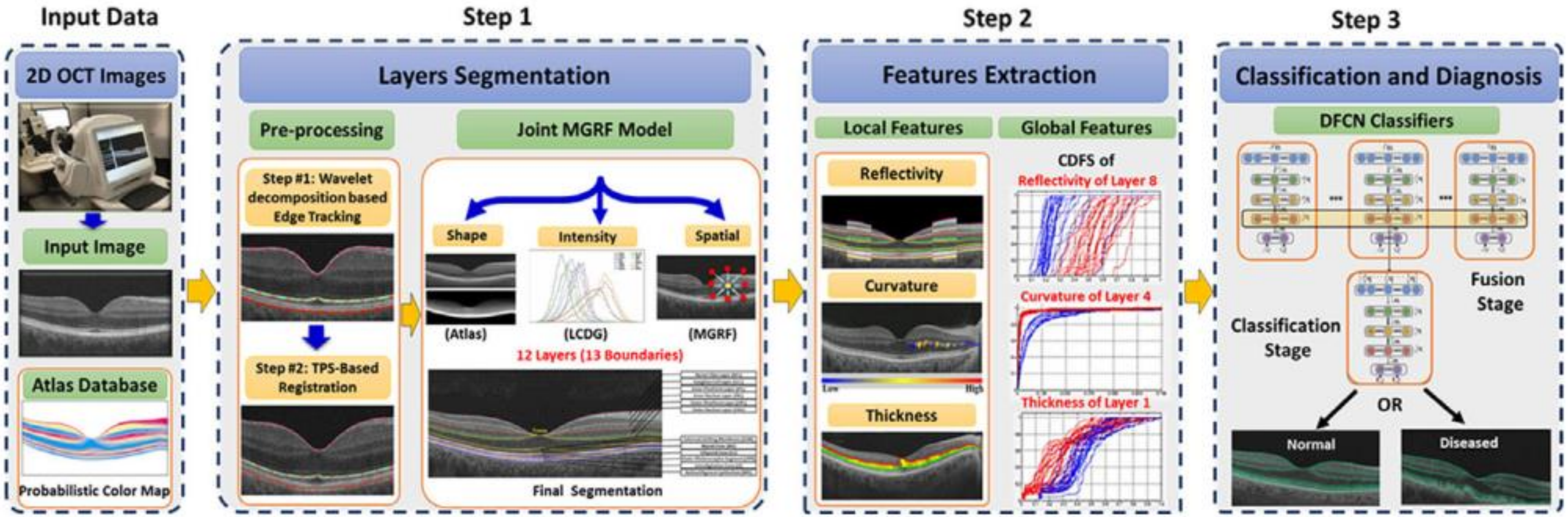


FIG. 1. Basic steps of the proposed diabetic retinopathy (DR) detection. [Colour figure can be viewed at wileyonlinelibrary.com]

CAD for OCT images

Segmentation

- Most early methods work only on images with high SNR and on normal retinas.
- This paper introduce segmentation into 12 layers and extracting 3 quantitative discriminative features.
- This papers uses a novel probabilistic model of input OCT images and output region maps of layers segmented.
- This model combines image intensities, prior knowledge of shapes of regions and local spatial dependencies between region labels.

- An input OCT image, **g**, co-aligned to a given training database, and its region map, m, are described with a joint probability model **P(g,m) = P(g|m)P(m)**

CAD for OCT images

Segmentation

Adaptive Shape Prior: Created using 12 scans and segmented by retina specialists.

- The pixel-wise labels across a stack of the **co-aligned** maps lead to the prior shape probabilities, of the typical retina.
- An image to segmented is aligned using **optimised TPS** that works by identifying control points (Eg. Fovea and points along this layer) and aligning the input image appropriately.

For input image, histogram of pixel-wise intensities is created and approximated. This is combined with **MGRF** which combines the dependency of a pixel with neighbouring pixels.

CAD for OCT images

Feature Extraction

- Reflectivity Curvature and Thickness of **each retinal layer** is extracted from each segmented OCT image.
  - The reflectivity is obtained from two regions per scan, comprising the **thickest** portions of the retina on the **nasal** and **temporal** sides of the foveal peak.
  - The curvature of a retinal layer combines all Menger curvature values calculated for each point across the layer after local weighted polynomial smoothing of the surface.
  - The thickness of a retinal layer is calculated from streamlines between the two surfaces of the layer.

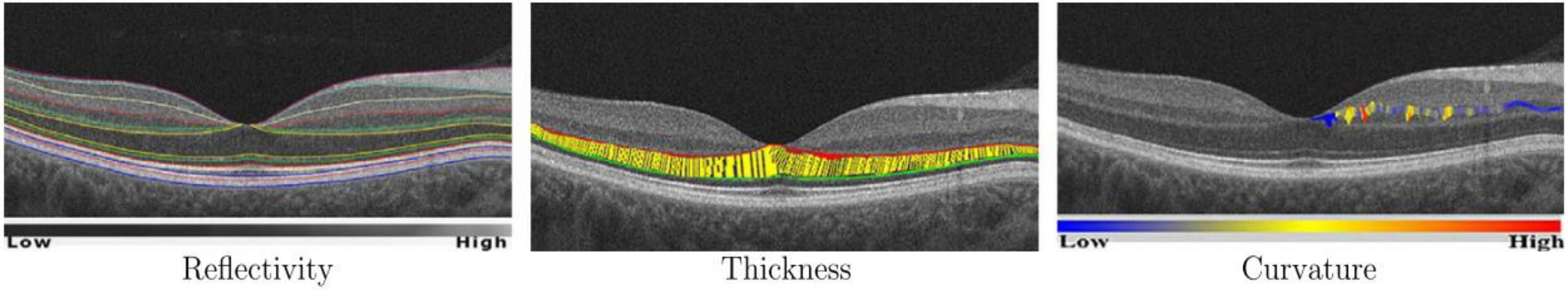
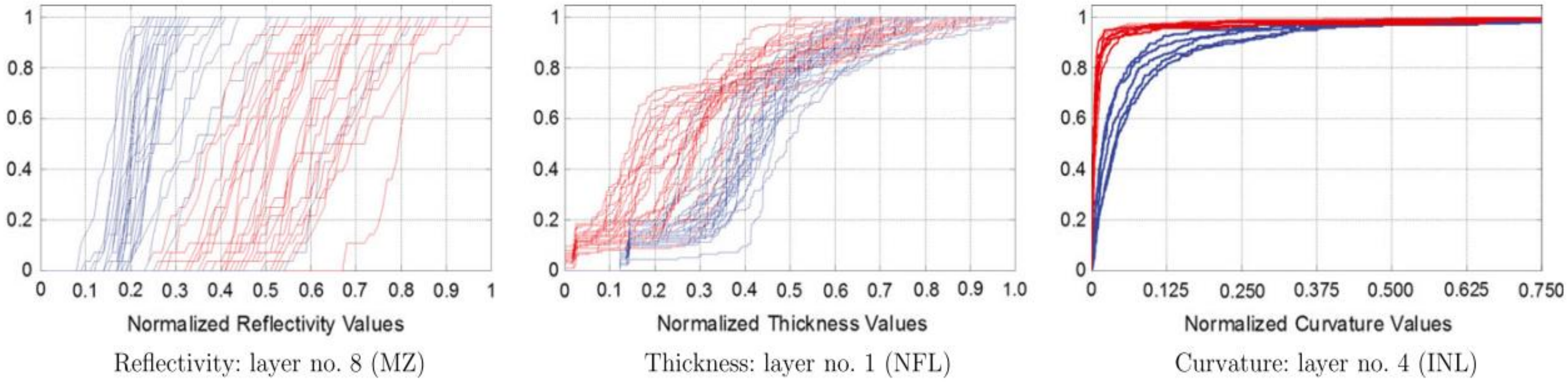


FIG. 3. Features for classification (the bottom gray and color bars encode the reflectivity and curvature values, respectively; the thickness is evaluated from streamlines (yellow) between the upper (red) and bottom (green) borders of a layer). [Colour figure can be viewed at wileyonlinelibrary.com]



CAD for OCT images

Architecture

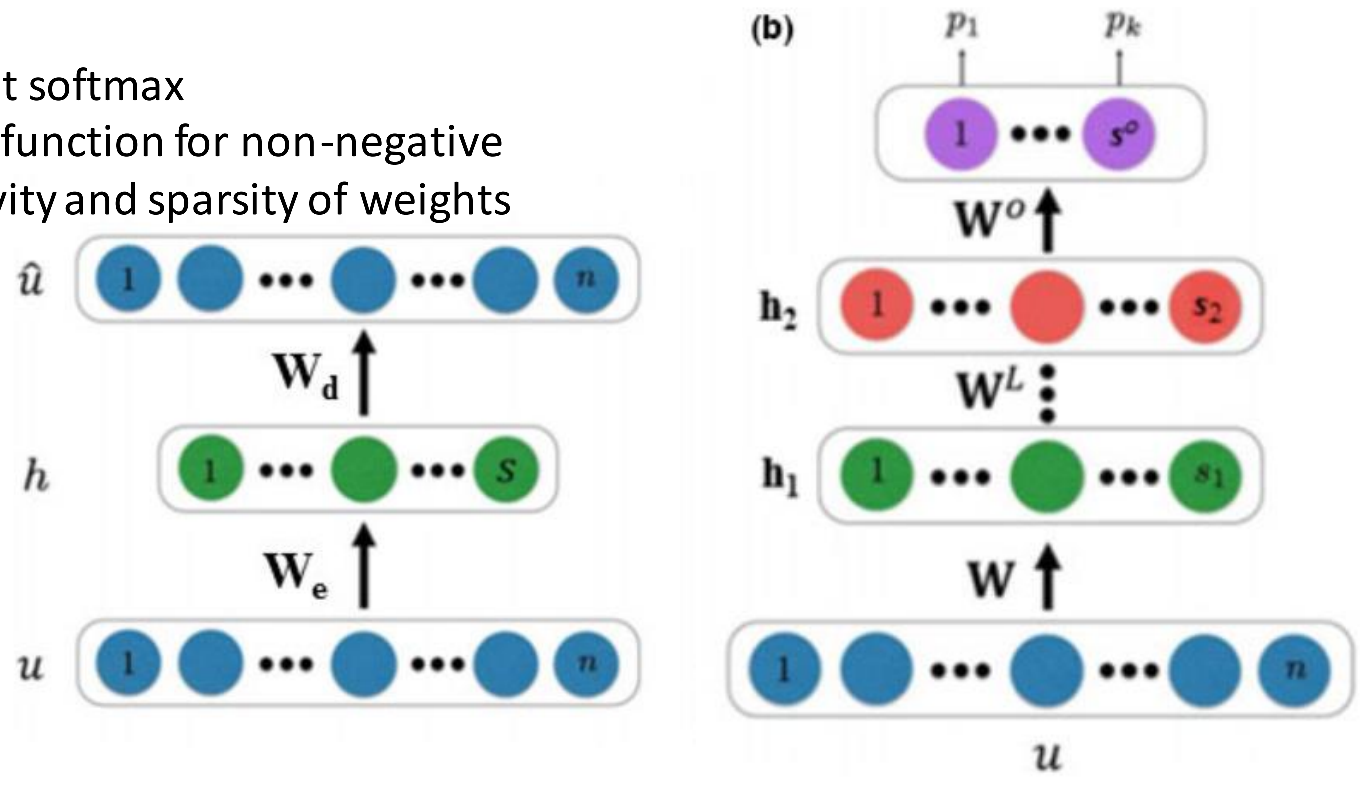
- For each subject, these three features of the extracted retina layers are described as a whole with a cumulative distribution function (CDF).
- 100 bins for Reflectivity and Thickness. 1000 bins for Curvature.
- This fixed size input is fed into multi-stage classifier with stack of non-negatively constrained autoencoder(SNCAE).

CAD for OCT images

Architecture

Autoencoder with encoding and decoding weights. Encoder converts the input vector into a latent representation and Decoder converts it back to the original input while learning the weights.

SNCAE with two NCAE layers and an output softmax
Layer. In this, there is a penalty in the cost function for non-negative numbers. This encourages the non-negativity and sparsity of weights



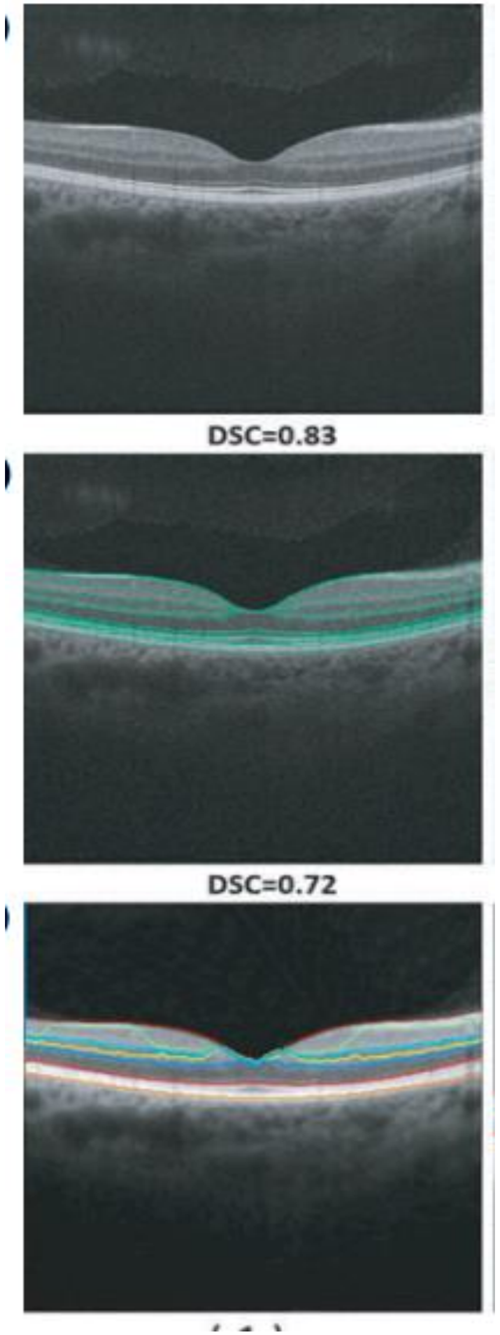
CAD for OCT images

Results

- This system was tested and validated on 52 subjects: 26 normal and 26 abnormal.
- The robustness and accuracy of the approach is measured using **Agreement Coefficient(AC)** and **Dice Similarity Coefficient(DSC)**. Higher DSC scores were obtained than previous segmentation approaches on both images with high and low SNR.
- The classifier decides whether the DR or no DR, based on the CDFs of the most discriminative features (the INL curvature, MZ reflectivity, and NFL thickness).
- 100 percent accuracy on leave-one-out-cross validation.

Classifier	Accuracy	Sensitivity	Specificity	AUC
DFCN (proposed)	**100%**	**100%**	**100%**	**0.98**
K-Star (K*)	95%	95%	95%	0.94
K-Nearest-Neighbor (K-NN)	90%	90%	90%	0.90
Random forest	85%	85%	85%	0.87
Random tree	72%	72%	72%	0.71

Classifier	Training accuracy (fourfold cross-validation)	Testing		
		Accuracy	Sensitivity	Specificity
DFCN (proposed)	**95%**	**92%**	**83%**	**100%**
K-Star (K*)	93%	89%	89%	89%
K-Nearest-Neighbor (K-NN)	91%	84%	84%	83%
Random forest	85%	82%	82%	82%
Random tree	83%	81%	81%	81%



GoogleNet and AlexNet

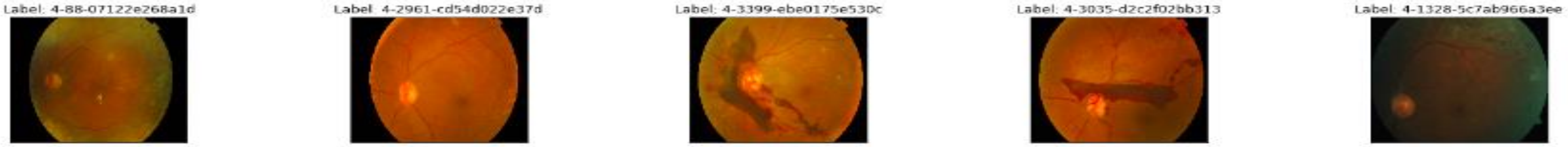
- This paper explores the use of Convolutional Neural Network(CNNs) to detect Diabetic Retinopathy particularly focusing on **Early** stage retinopathy.

- This paper used **transfer learning** on models GoogleNet and AlexNet which were trained on ImageNet dataset to do a 2-nary, 3-nary and 4-nary classification and measured the metrics.

GoogleNet and AlexNet

Data and Pre-processing

Kaggle dataset is used which has 35000 images with 5 classes



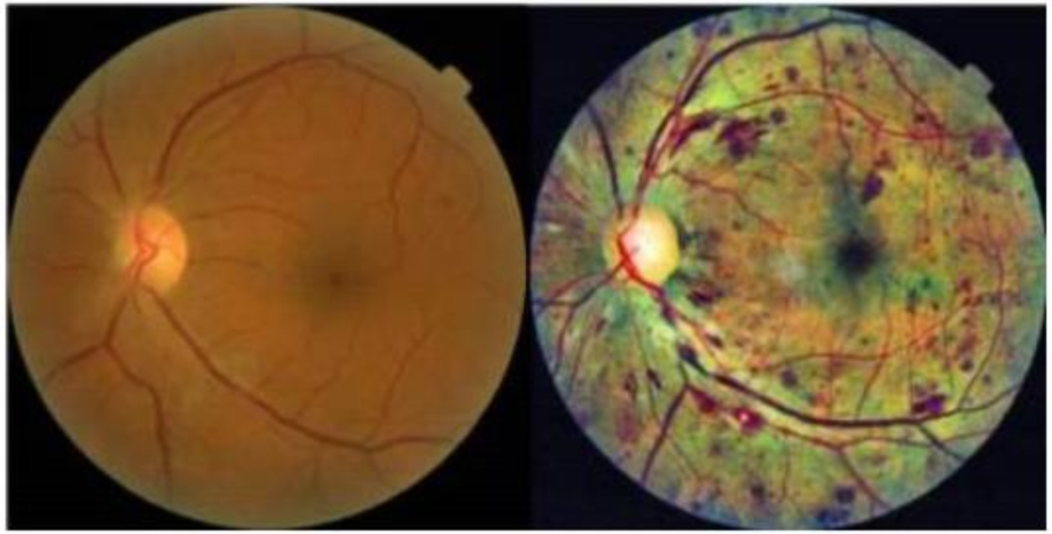
Messidor-1 dataset is used which consists of physician verified 1200 images. This is a set of higher fidelity images used to make a good quality model

GoogleNet and AlexNet

Data and Pre-processing

Since there are many images with **too much black area** which is **uninformative** and hinders the learning of the convolutional network, these are removed using **Otsu's** method.

Images are then normalized and Contrast Limited Adaptive Histogram Equalization (**CLAHE**) is applied to get better discernible features.

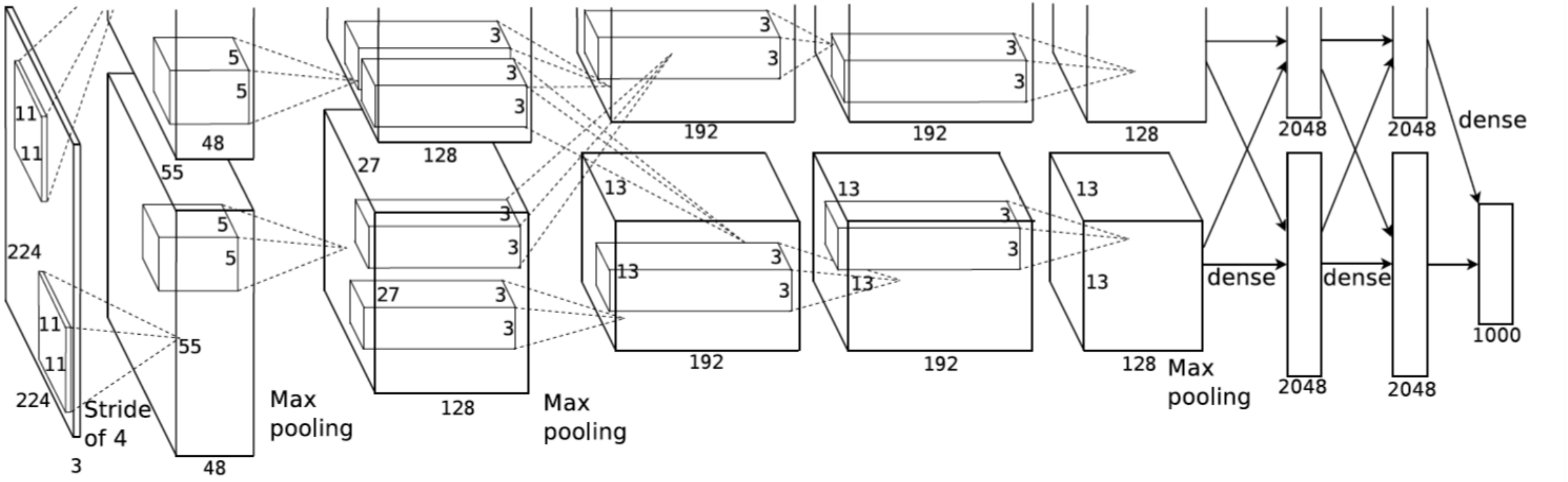


Before and After preprocessing

GoogleNet and AlexNet

AlexNet

It is better version of LeNet(1998). AlexNet is deeper, with more filters per layer, and with stacked convolutional layers. It consists of **11x11, 5x5,3x3**, convolutions, max pooling, **dropout**, data augmentation, **ReLU** activations, SGD with **momentum**. It attached ReLU activations after every convolutional and fully-connected layer.

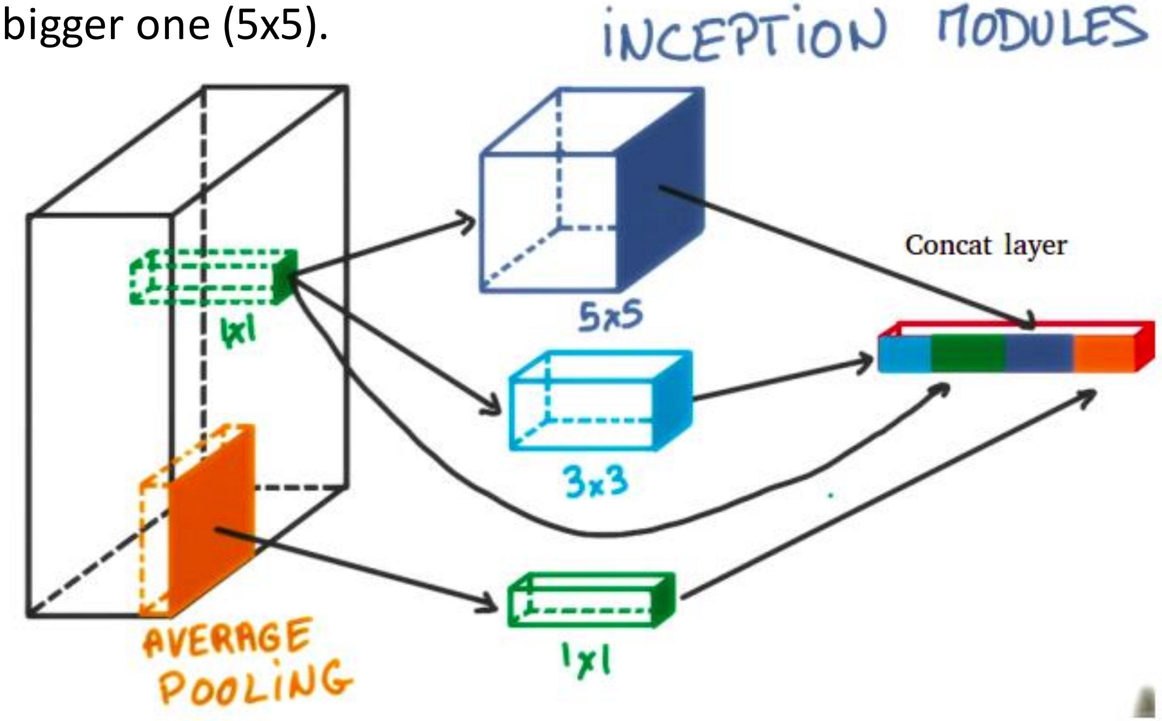


GoogleNet and AlexNet

GoogleNet

The network uses a CNN inspired by LeNet but implements a novel element which is called an **inception module**. Also the network uses batch normalization, image distortions and RMSprop.

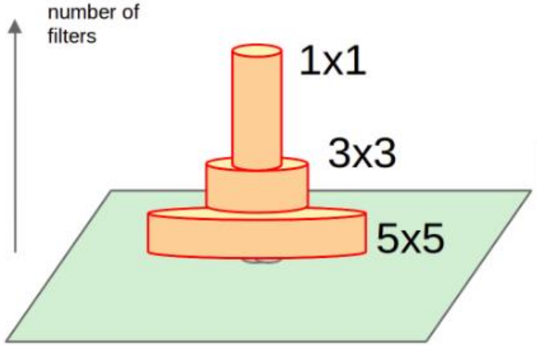
The idea of the inception layer is to **cover a bigger area, but also keep a fine resolution** for small information on the images. So the idea is to convolve in parallel different sizes from the most accurate detailing (1x1) to a bigger one (5x5).

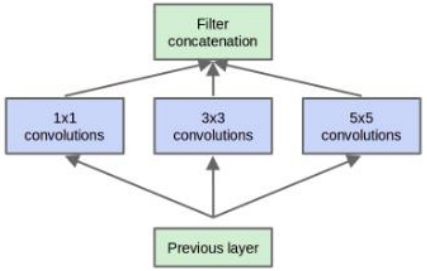


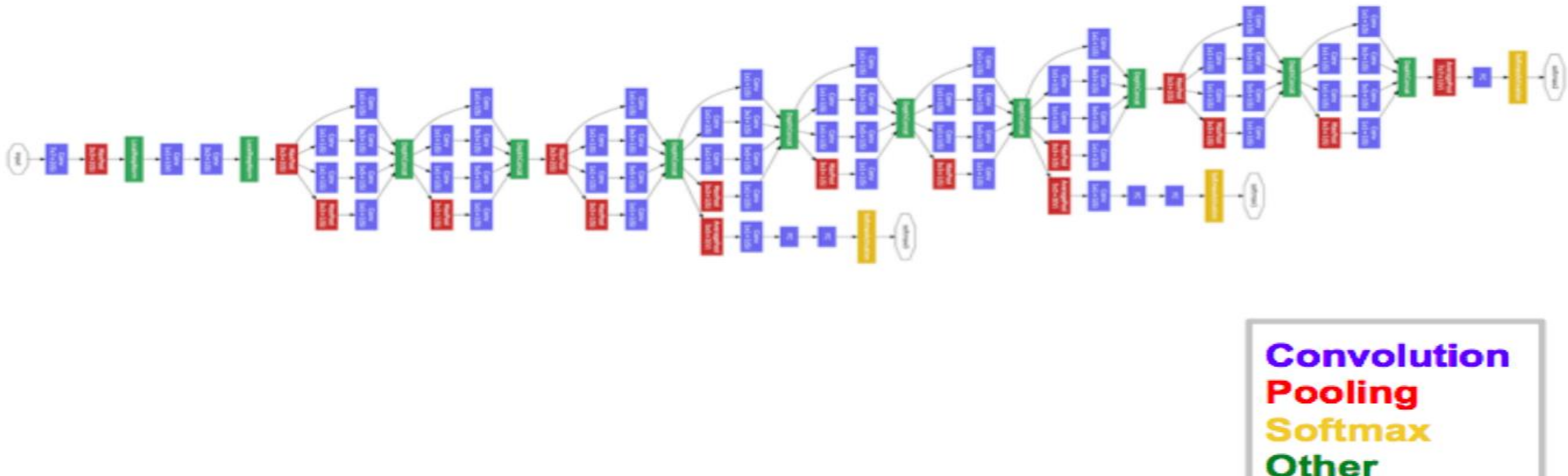
GoogleNet and AlexNet

GoogleNet

GoogleNet has 22 layer, and almost 12x less parameters (So faster and less then Alexnet and much more accurate). It also uses bottleneck approaches (dim. Reduction) to avoid parameter from increasing too much.







GoogleNet and AlexNet

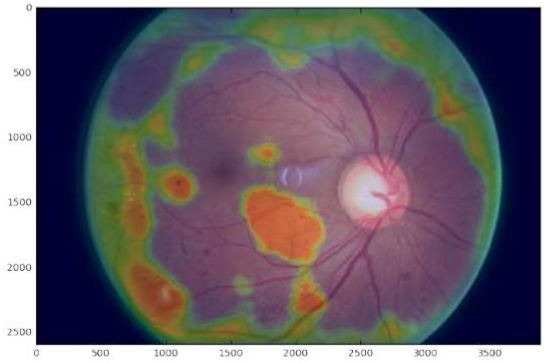
Experiment Results

- 3-nary classification sensitivity increased from 0 to 29.4%
- GoogLeNet model achieved the highest sensitivity of 95% and specificity of 96% with data augmentation and preprocessing techniques.
- For 3-nary and 4-nary classification, authors were unable to achieve significant sensitivity levels for **mild** class.
- This is attributed to CNN's inability to detect subtle features
- Authors hypothesize this can be due to camera artifact errors and due to low fidelity labelling (misleading or incorrect)

GoogleNet and AlexNet

Experiment Results

- Using heatmap for features detected by CNN, large features were the ones responsible for abnormality detection, while small features were not detected.
- 4-ary classifier encounters a problem of simply not having enough images to effectively train a deep CNN such as GoogLeNet. (Underfitting)



GoogleNet and AlexNet

Conclusion

- Authors were able to achieve state-of-the-art performance with CNNs using binary classifiers, the model performance degrades with increasing number of classes.
- Medical images are fraught with subtle features that can be crucial for diagnosis. Fortunately, the most often deployed architectures have been optimized to recognize macroscopic features such as those present in the ImageNet dataset.
- Therefore we require a new paradigm for diagnosing diseases via CNN models.

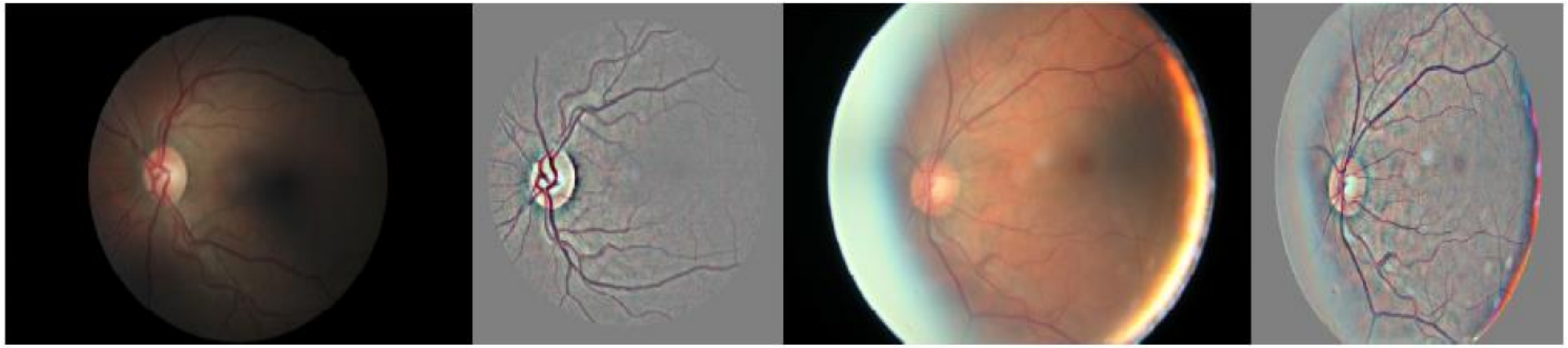
Deep Learning for prescreening in eye-pacs

Steps:

- Preprocessing
- CNN structure + Algorithm
- Results

Preprocessing

The images were rescaled to have the exact radius of 300px. After that the colored images were converted to grayscale. Following these steps, the images were clipped to 90% size to remove boundary effects and the final image resized to 224 x 224px.



CNN Architecture

Resnet – 50 is used.

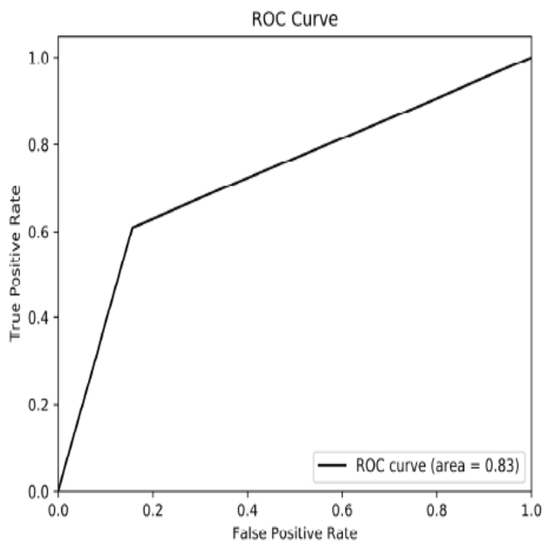
- ReLU has been used as the activation function.
- Batch normalization has been used after each convolutional layer.
- A dropout layer has also been added to the convolutional block to avoid overfitting.Inputs are dropped at a rate of 50%

CNN Architecture

- The weights trained for ImageNet ILSVRC algorithm were used as the initial weights for our model (transfer learning). The model was then trained with stochastic gradient descent with Adam optimization.
- learning rate : 0.0001
- beta-1: 0.9
- beta-2: 0.99
- Epochs of training: 100
- loss function : categorical cross-entropy function

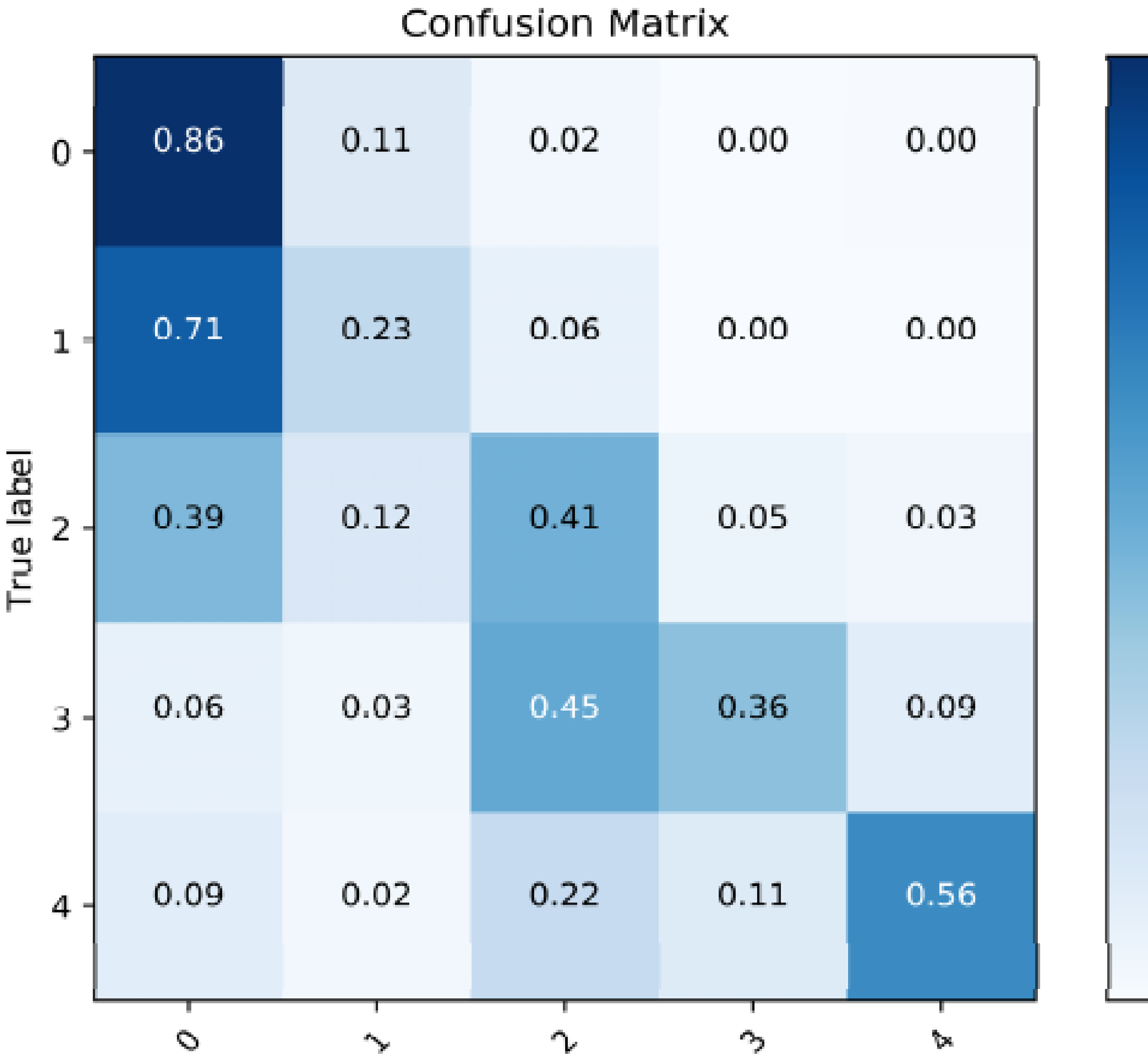
Results

- AuROC (area under receiver operating curve) of 0.83 was achieved.Accuracy is measured by the area under the ROC curve.
- Also specificity of 84%(true negative rate) and 61% sensitivity(true positive rate) was achieved across the full dataset.



Confusion Matrix

- The confusion matrix indicates problem in detecting levels 1, 2 and 3 and thus unable to identify subtle differences between adjacent classes.
- This could also be the effect of not ensuring equal test images for each label (in our case more images for 0 label).
- This is also possible due to low resolution of the images leading to the inability of the model to identify fine details such as exudates and hemorrhages.



Conclusion

- We analysed 2 types of data 2D fundus and OCT scans. DIfferent methods were analysed for both datasets.
- Most of the CNN based methods accurately differentiated between Proliferative DR and No DR. Also CNN based methods identified features which were observable but could not identify subtle features.
- Also localisation of features was achieved using CNN with RAM. Early stage DR can be identified using OCT scans by segmenting the Retinal layers and identifying discriminating features which gave satisfactory performance.