CADx PROJECT (Skin Lesion Classification)

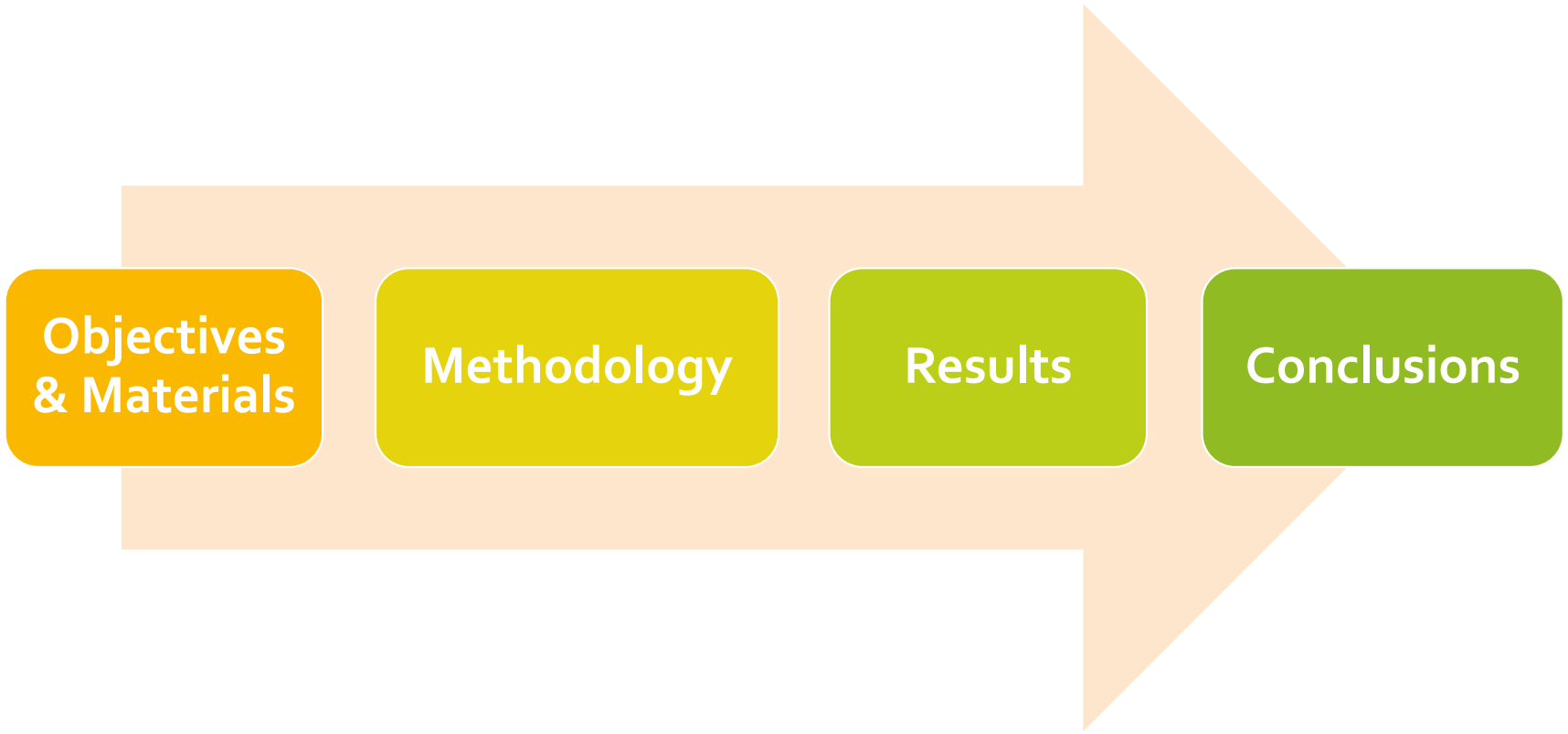
Supervised By

- Dr. Xavier Llado
- Dr. Arnau Oliver
- Kaisar Kushibar

Presented By

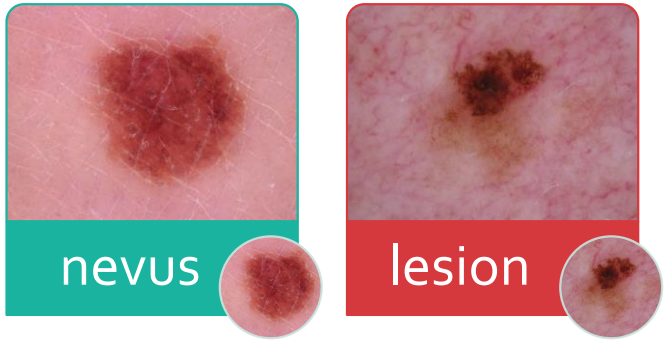
- Md. Kamrul Hasan
- Basel Alyafi
- Fakrul Islam Tushar

Outline

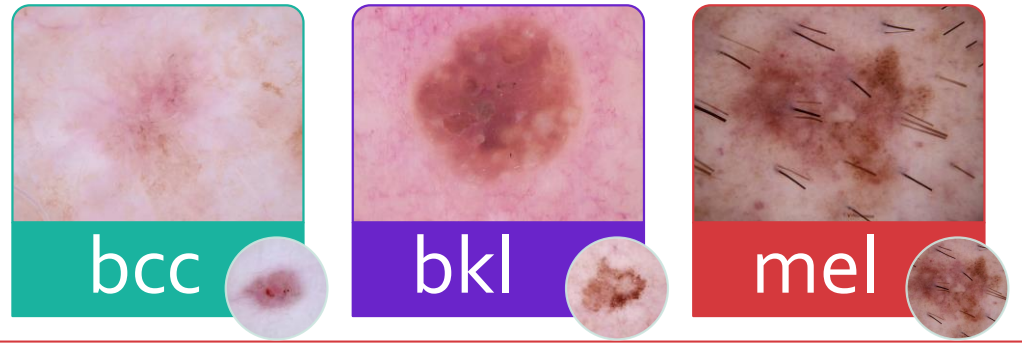


Objectives

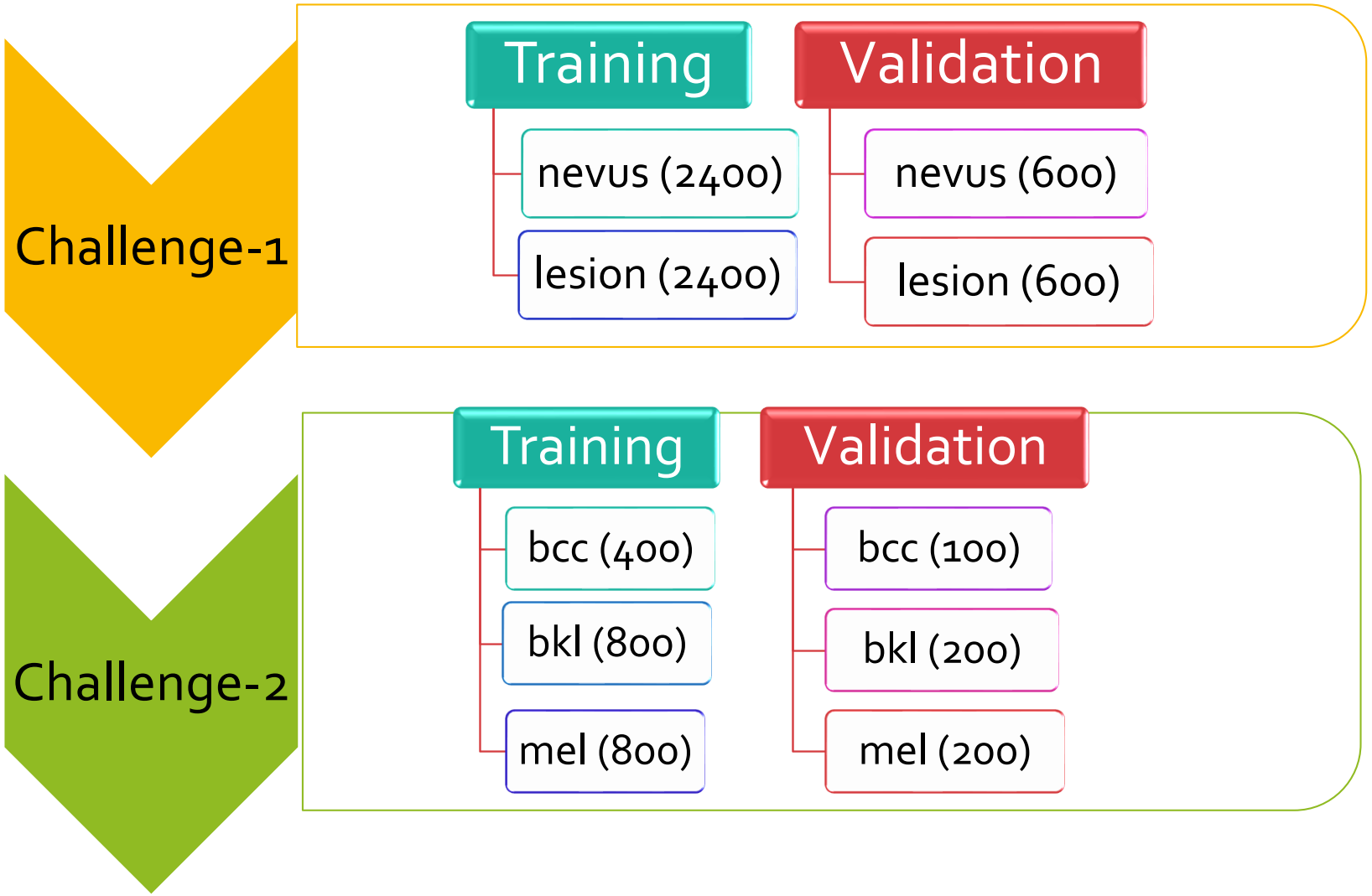
Challenge-1



Challenge-2

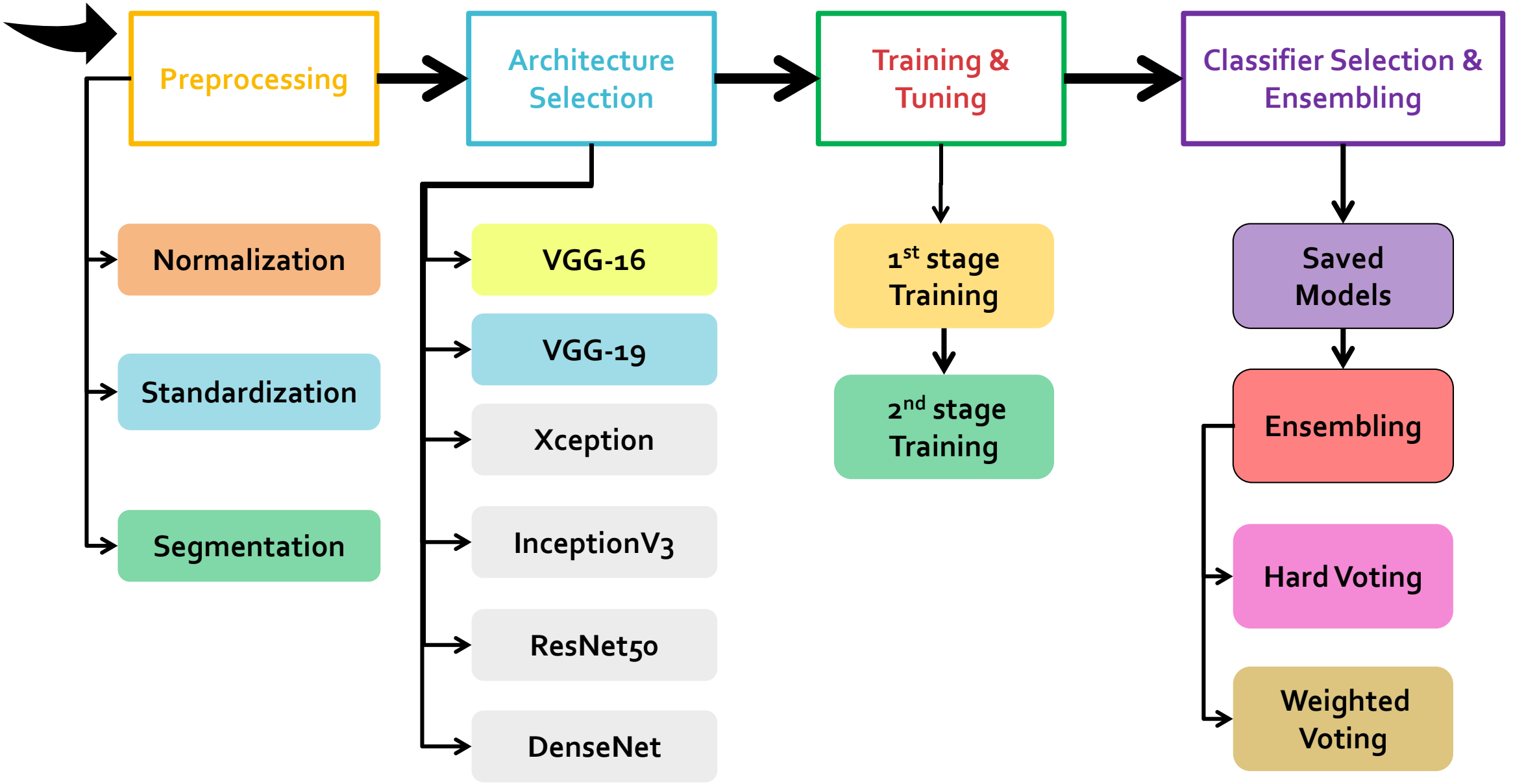


Materials (Image data)

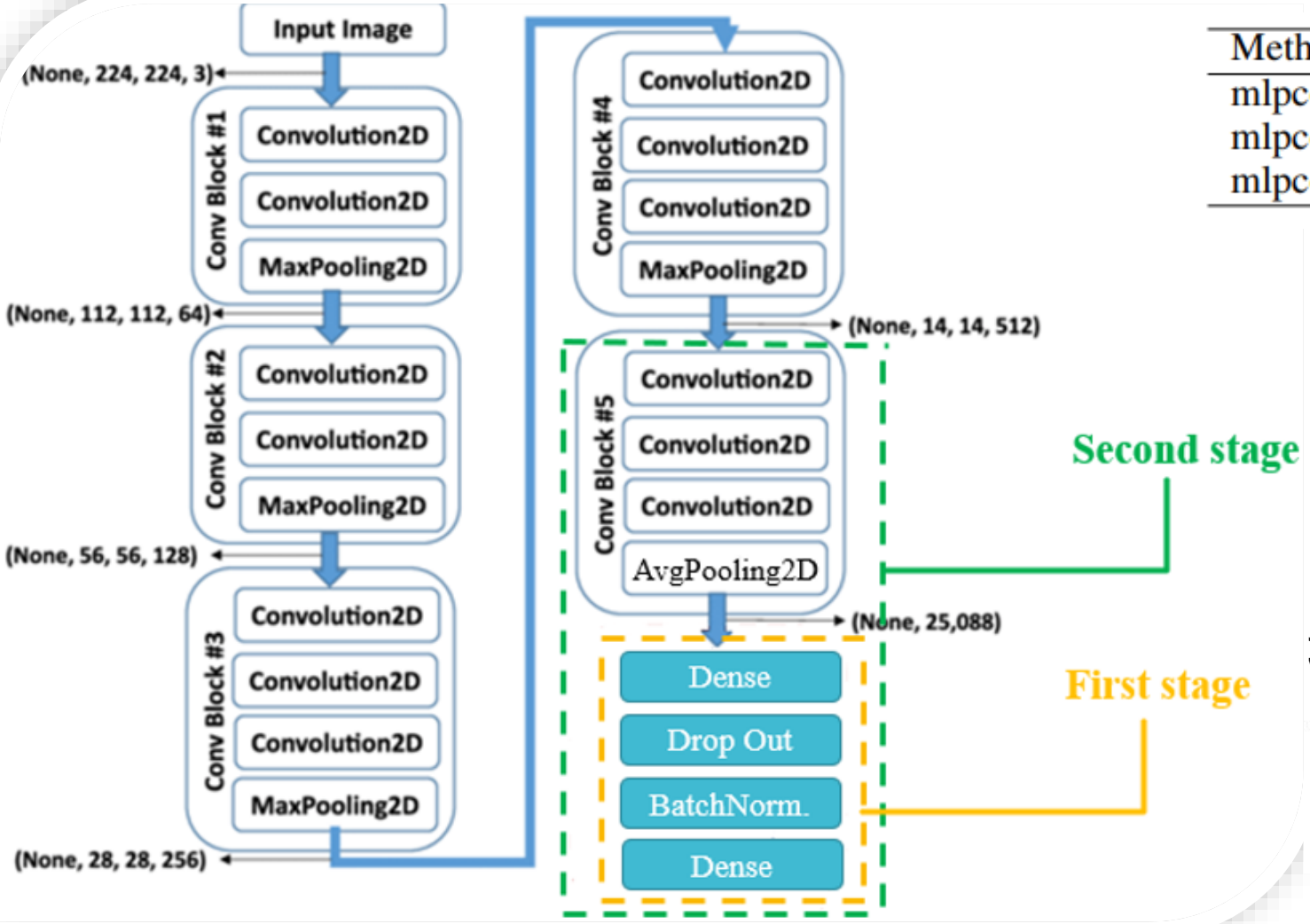


Methodology

- Overview of the work.

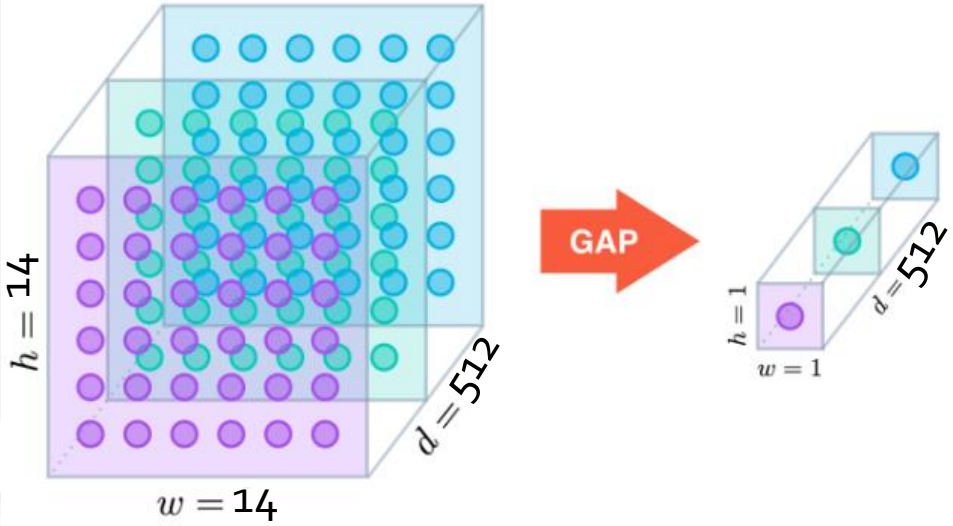

Preprocessing
Architecture Selection
Training & Tuning
Classifier Selection & Ensembling
Normalization
Standardization
Segmentation
VGG-16
VGG-19
Xception
InceptionV3
ResNet50
DenseNet
1st stage Training
2nd stage Training
Saved Models
Ensembling
Hard Voting
Weighted Voting

Training & Tuning



Method	Testing Error
mlpconv + Fully Connected	11.59%
mlpconv + Fully Connected + Dropout	10.88%
mlpconv + Global Average Pooling	10.41%

Global average pooling compared to FCN [1].



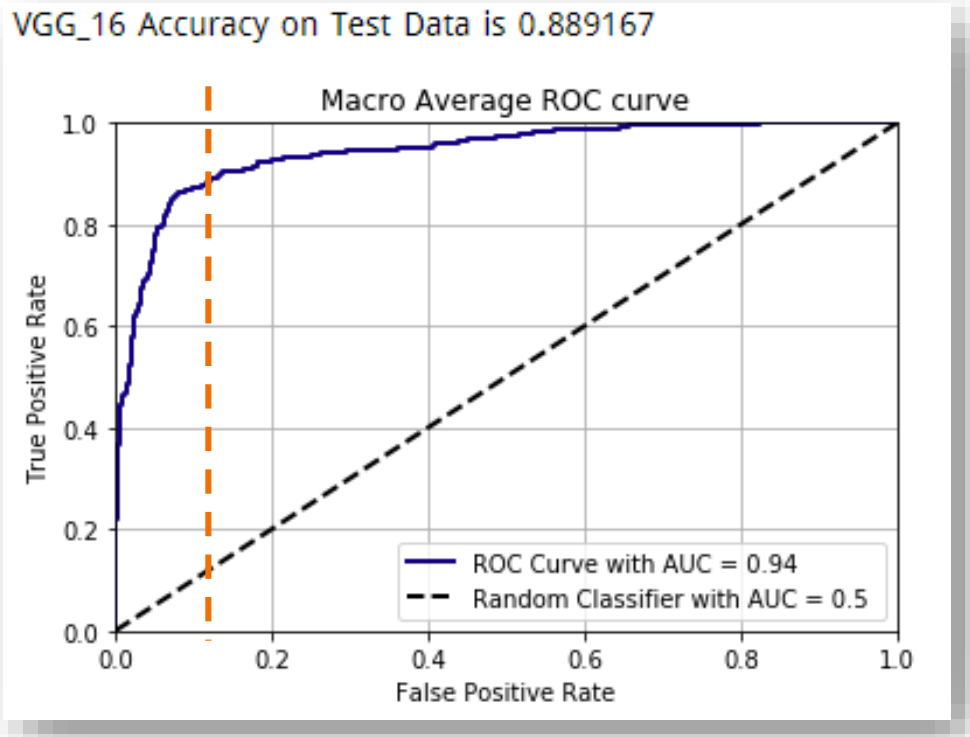
Global average pooling [2]

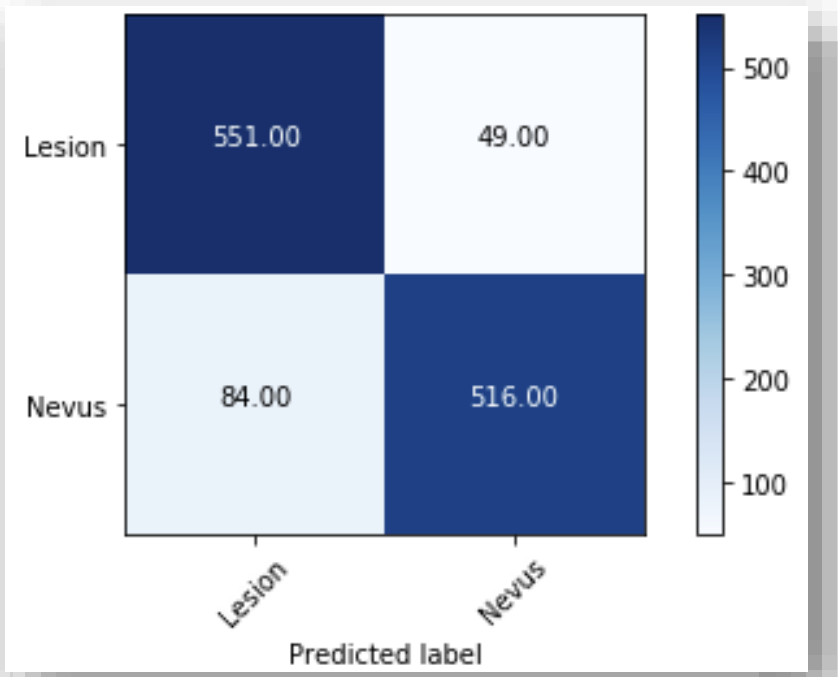
Challenge-1

Stage	Optimizer	Learning rate	Momentum	Nestrov	Metric	Epochs	Loss
1	Rmsprop	1e-3	-	-	Accuracy	30	**binary_ crossentropy**
2	SGD	1e-4	0.9	True	Accuracy	30	

Experiments-1 on VGG-16

Results Experiment-1





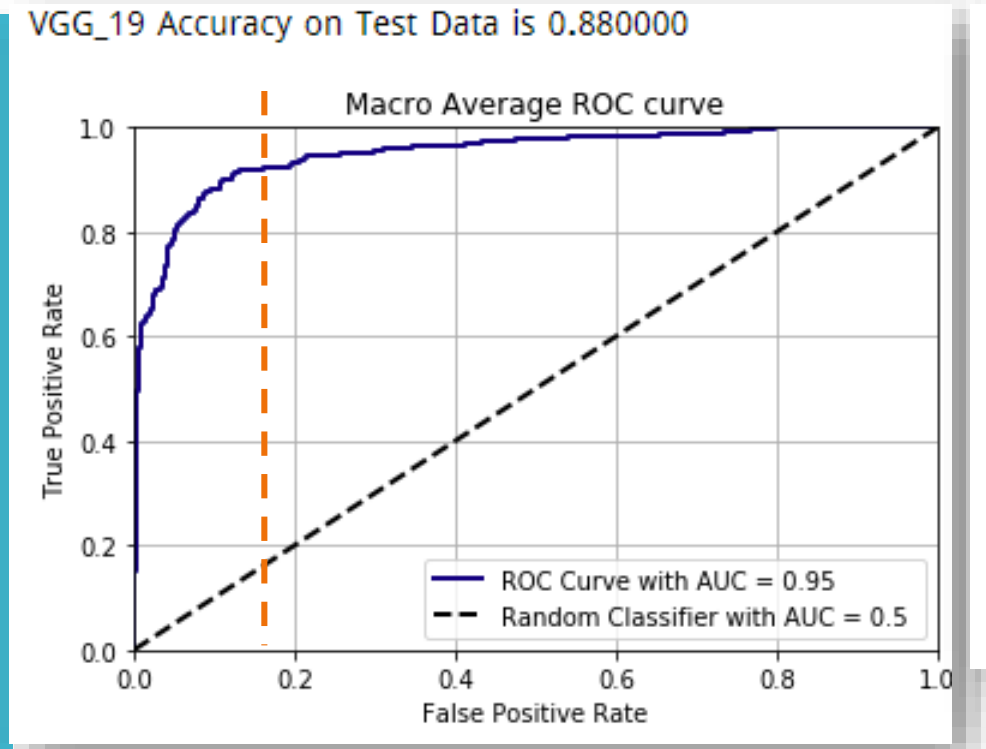
	precision	recall	f1-score	support
Lesion	0.87	0.92	0.89	600
Nevus	0.91	0.86	0.89	600

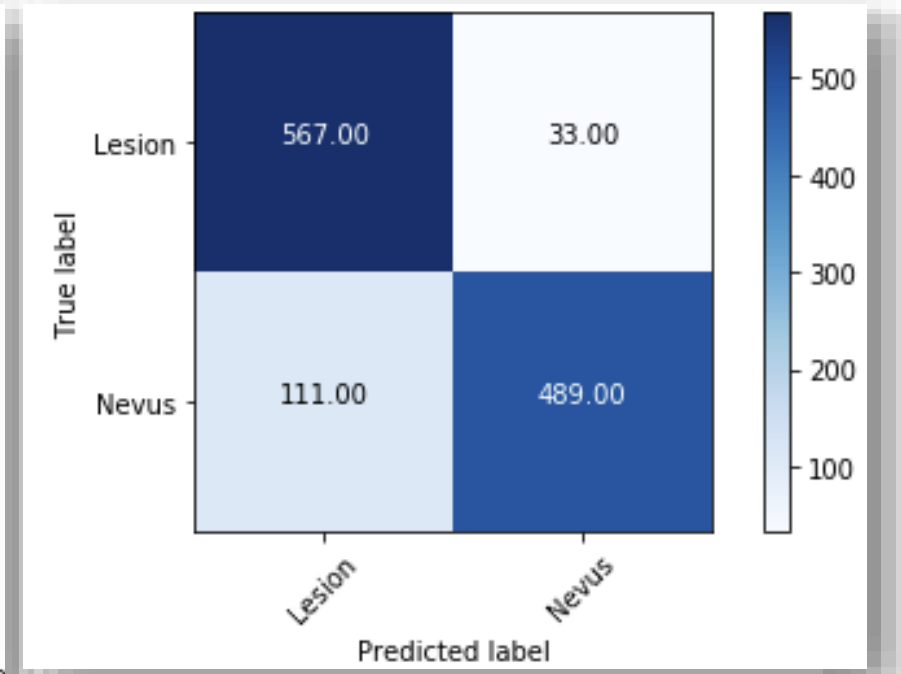
Stage	Optimizer	Learning rate	Momentum	Nestrov	Metric	Epochs	Loss
1	SGD	1e-4	0.0	False	Accuracy	30	**binary_ crossentropy**
2	SGD	1e-4	0.9	False	Accuracy	30	

Experiments-2 on VGG-16

Results Experiment-2

VGG_19 Accuracy on Test Data is 0.880000



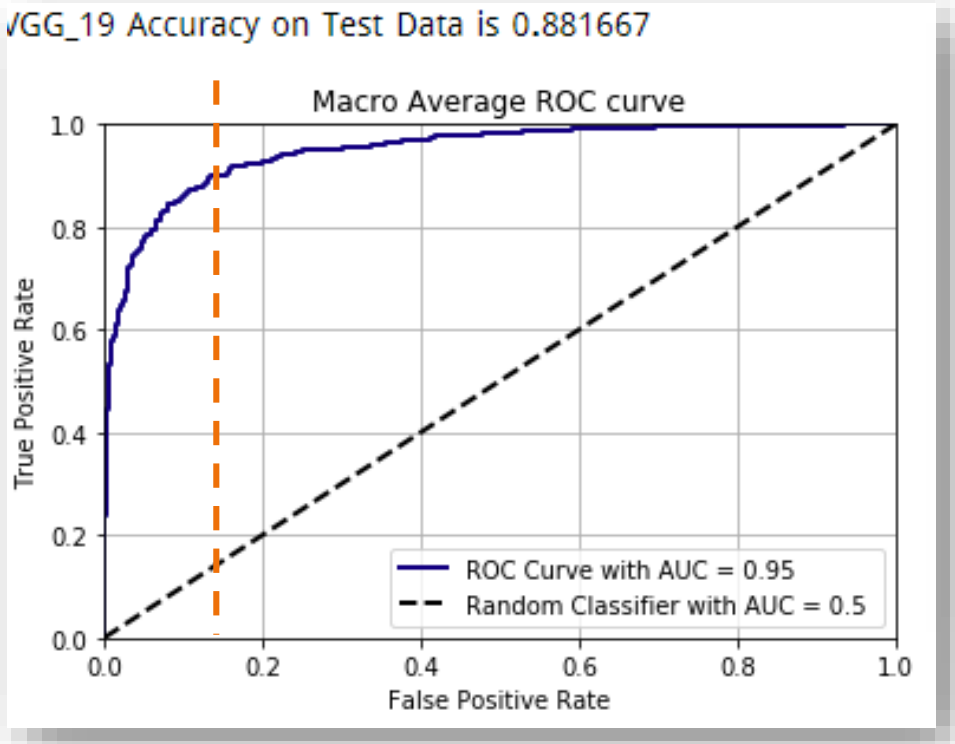


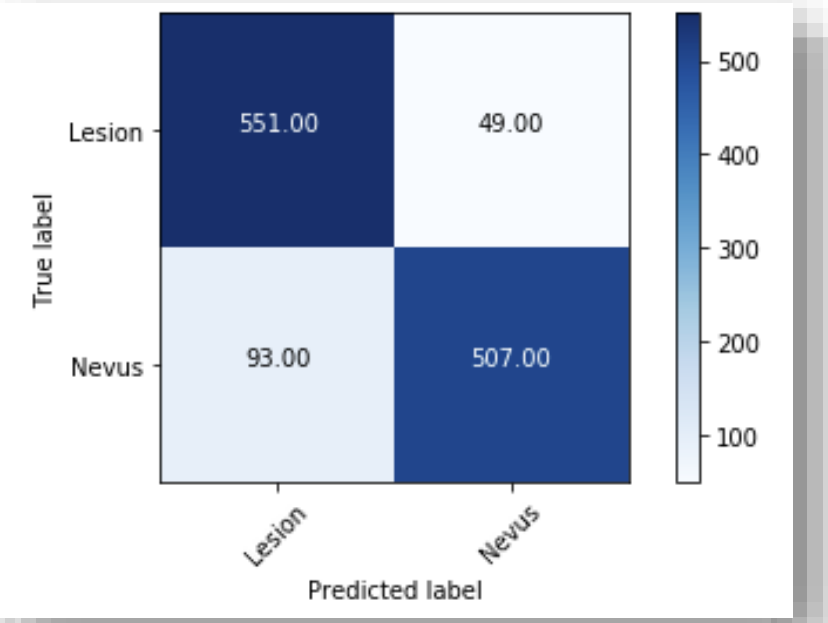
	precision	recall	f1-score	support
Lesion	0.84	0.94	0.89	600
Nevus	0.94	0.81	0.87	600

Stage	Optimizer	Learning rate	Momentum	Nestrov	Metric	Epochs	Loss
1	Rmsprop	1e-3	-	-	Accuracy	30	**binary_ crossentropy**
2	SGD	1e-4	0.9	True	Accuracy	30	

Experiments-3 on VGG-19

Results for Experiment-3





	precision	recall	f1-score	support
Lesion	0.86	0.92	0.89	600
Nevus	0.91	0.84	0.88	600

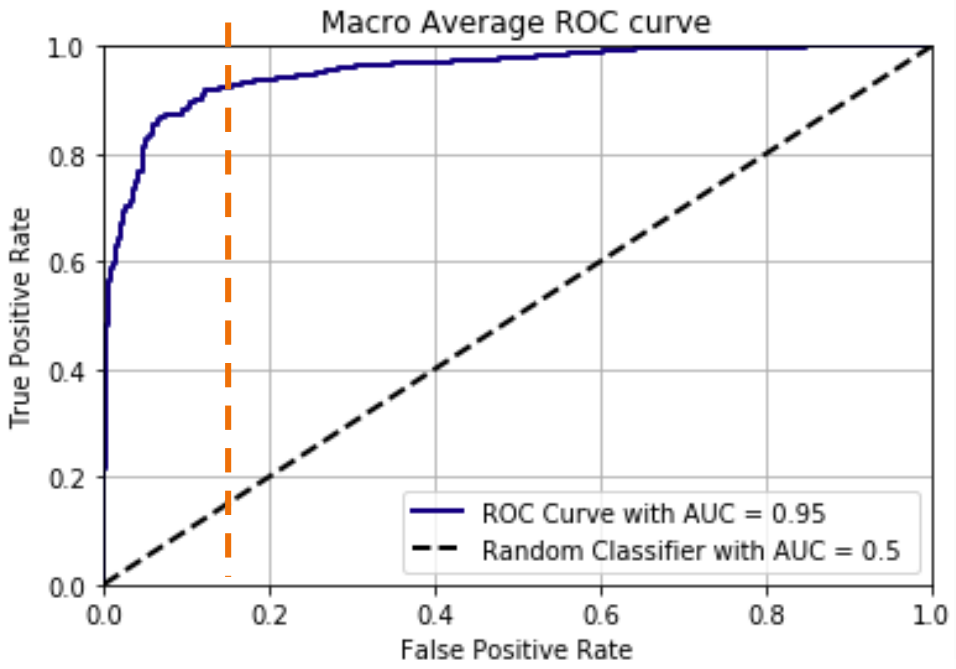
VIVA Democracia

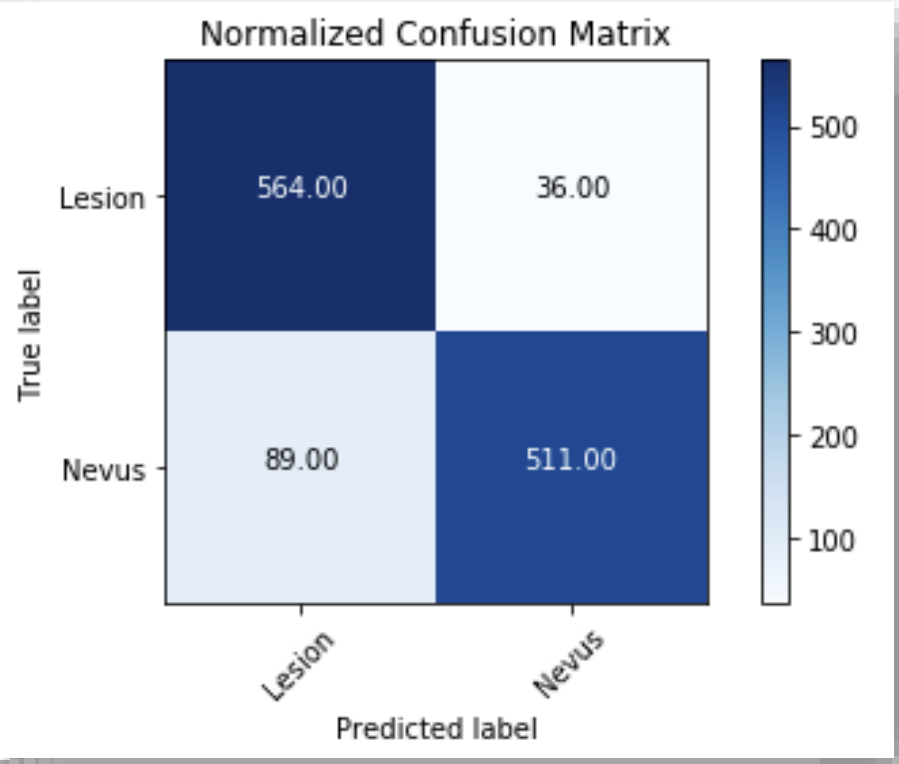
Hard voting:
Count classifiers votes

Weighted voting:
Weigh probabilities by classifiers AUC.

Ensembled CNN Hard Voting

Ensemble Accuracy on Test Data is 0.895833





	precision	recall	f1-score	support
Lesion	0.86	0.94	0.90	600
Nevus	0.93	0.85	0.89	600

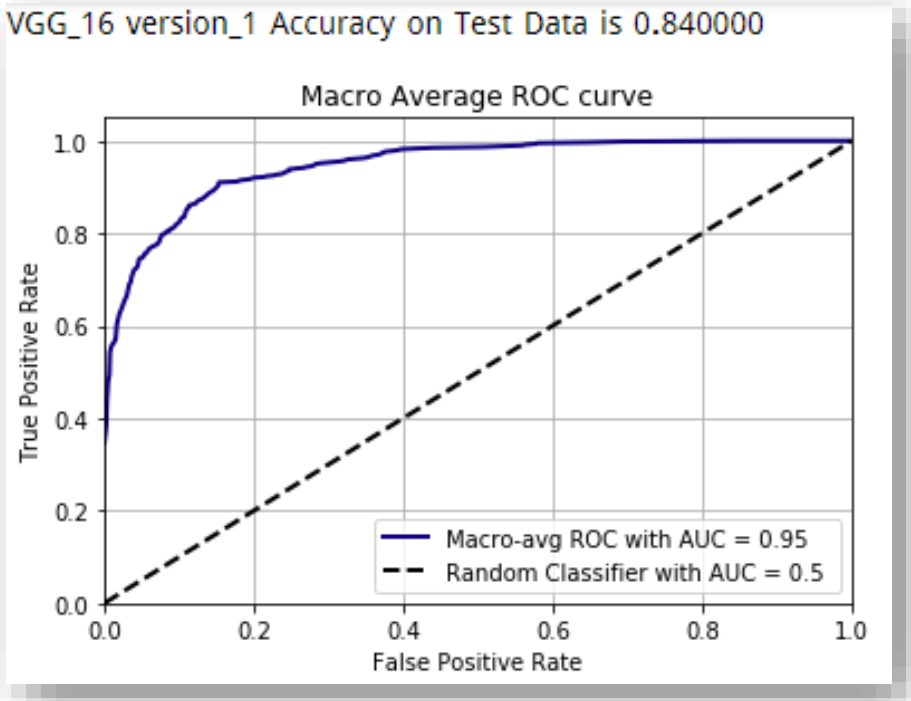
Challenge-2

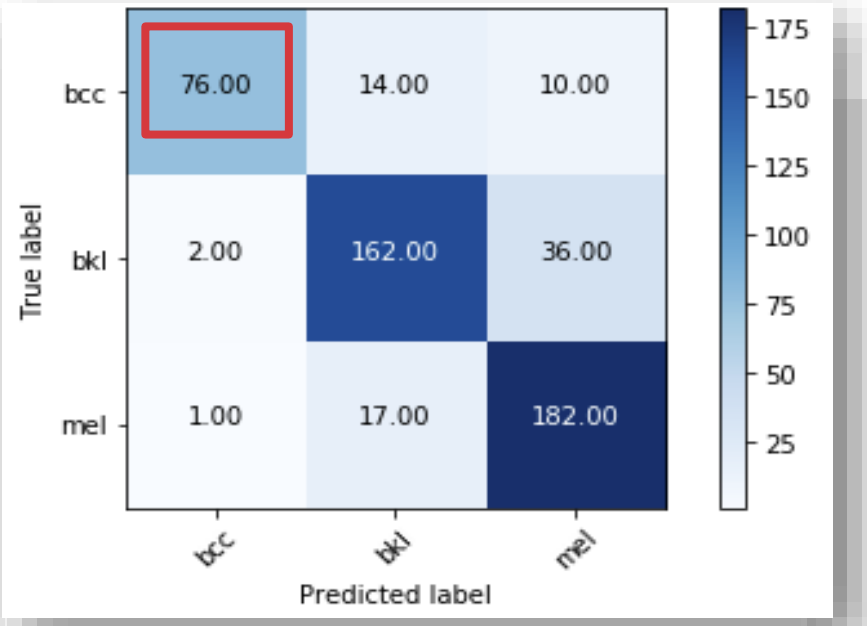
Stage	Optimizer	Learning rate	Momentum	Nestrov	Metric	Epochs	Loss
1	Rmsprop	1e-3	-	-	Accuracy	30	**categorical_ crossentropy**
2	SGD	1e-4	0.9	True	Accuracy	30	

Experiments-1 on VGG-16

2

Results Experiment-1





	precision	recall	f1-score	support
class: bcc	0.96	0.76	0.85	100
class: bkl	0.84	0.81	0.82	200
class: mel	0.80	0.91	0.85	200
micro avg	0.84	0.84	0.84	500
macro avg	0.87	0.83	0.84	500
weighted avg	0.85	0.84	0.84	500

No of errors = 80/500

Experiments-2 & 3 on VGG-16

Exp-2: **size-weighted** loss function
class_weight={'bcc': 2, 'bkl': 1, 'mel': 1}

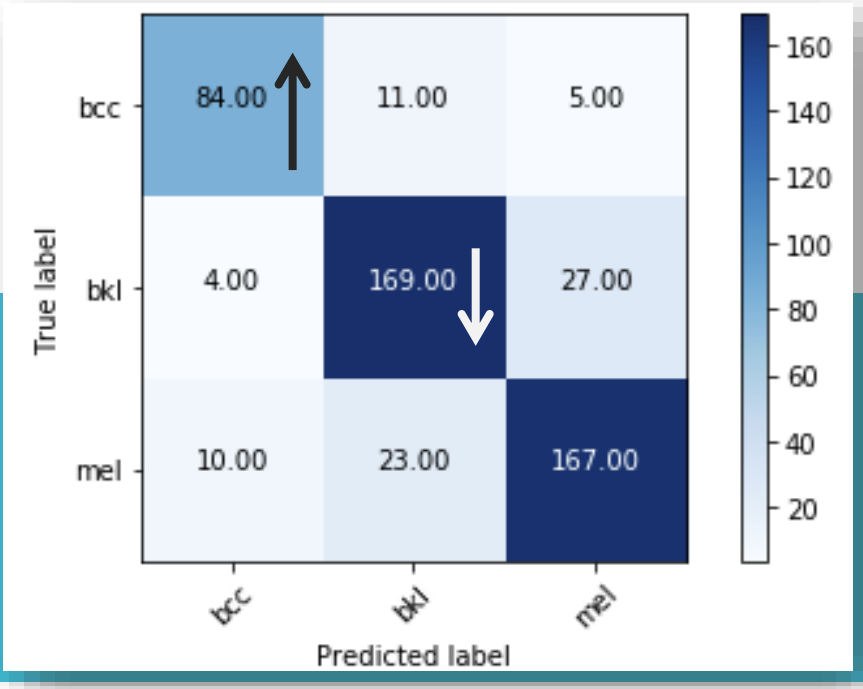
Exp-3: standardization instead of normalization

	precision	recall	f1-score	support
class: bcc	0.86	0.84	0.85	100
class: bkl	0.83	0.84	0.84	200
class: mel	0.84	0.83	0.84	200
micro avg	0.84	0.84	0.84	500
macro avg	0.84	0.84	0.84	500
weighted avg	0.84	0.84	0.84	500

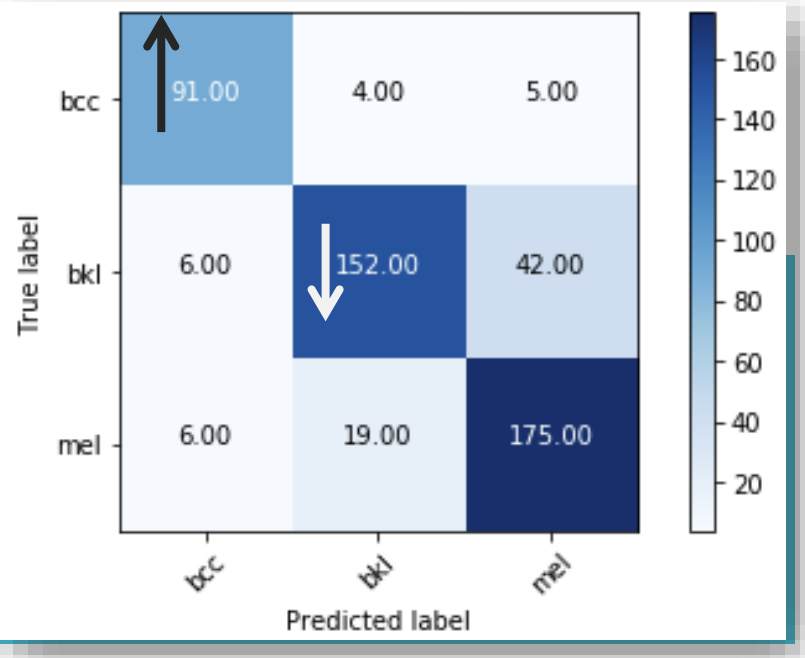
No of errors = 80/500

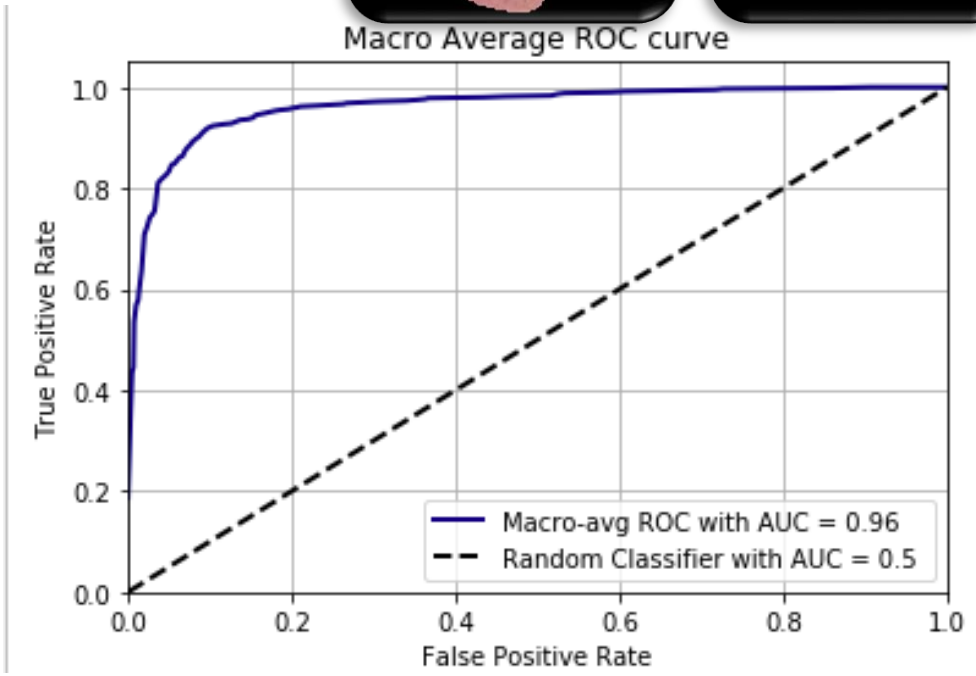
	precision	recall	f1-score	support
class: bcc	0.88	0.91	0.90	100
class: bkl	0.87	0.76	0.81	200
class: mel	0.79	0.88	0.83	200
micro avg	0.84	0.84	0.84	500
macro avg	0.85	0.85	0.85	500
weighted avg	0.84	0.84	0.84	500

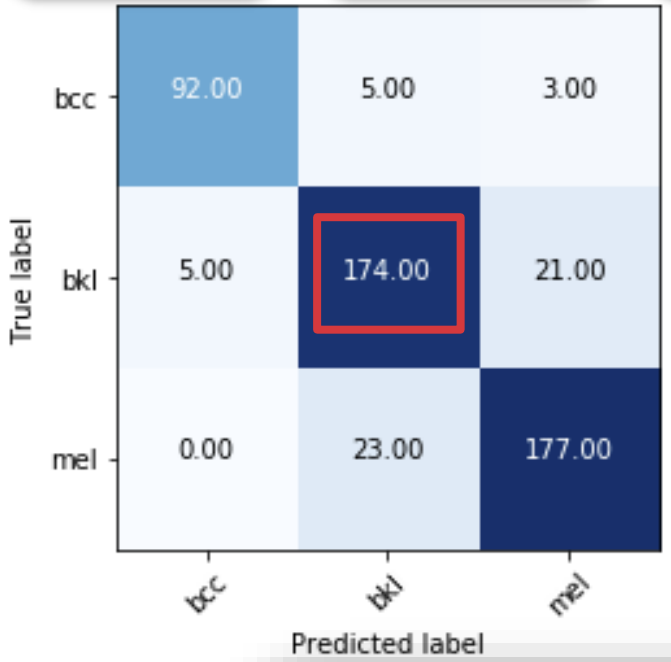
No of errors = 82/500

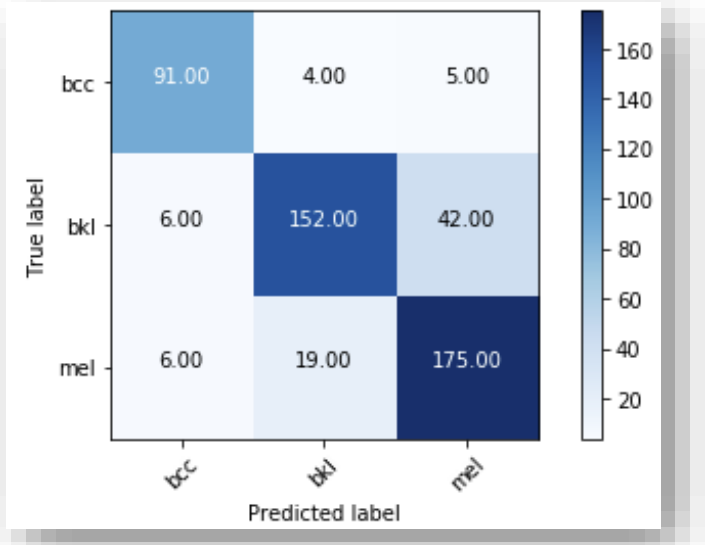


Results Experiment-2 & 3









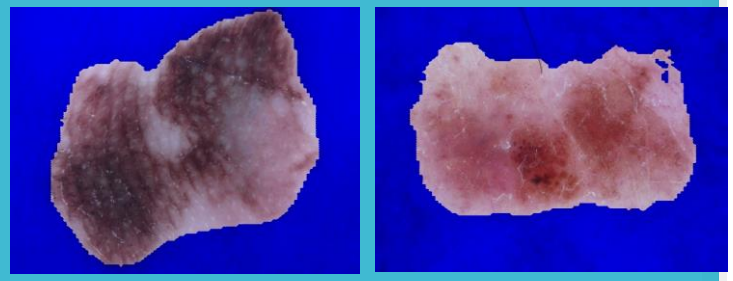
3rd Experiment

Results Experiment-4

	precision	recall	f1-score	support
class: bcc	0.95	0.92	0.93	100
class: bkl	0.86	0.87	0.87	200
class: mel	0.88	0.89	0.88	200
micro avg	0.89	0.89	0.89	500
macro avg	0.90	0.89	0.89	500
weighted avg	0.89	0.89	0.89	500

No of errors = 57/500

Results Experiment-5 & 6 & 7



Synthetic data; (RG) segmented images merged with the original B channel.

Exp. 5
Standard. Preprocess
No segmentation

	precision	recall	f1-score	support
class: bcc	0.95	0.78	0.86	100
class: bkl	0.82	0.86	0.84	200
class: mel	0.85	0.89	0.87	200
micro avg	0.86	0.86	0.86	500
macro avg	0.88	0.84	0.86	500
weighted avg	0.86	0.86	0.86	500

No of errors = 72/500

Exp. 6
2000X2
Synthetic input:
Original blue channel only

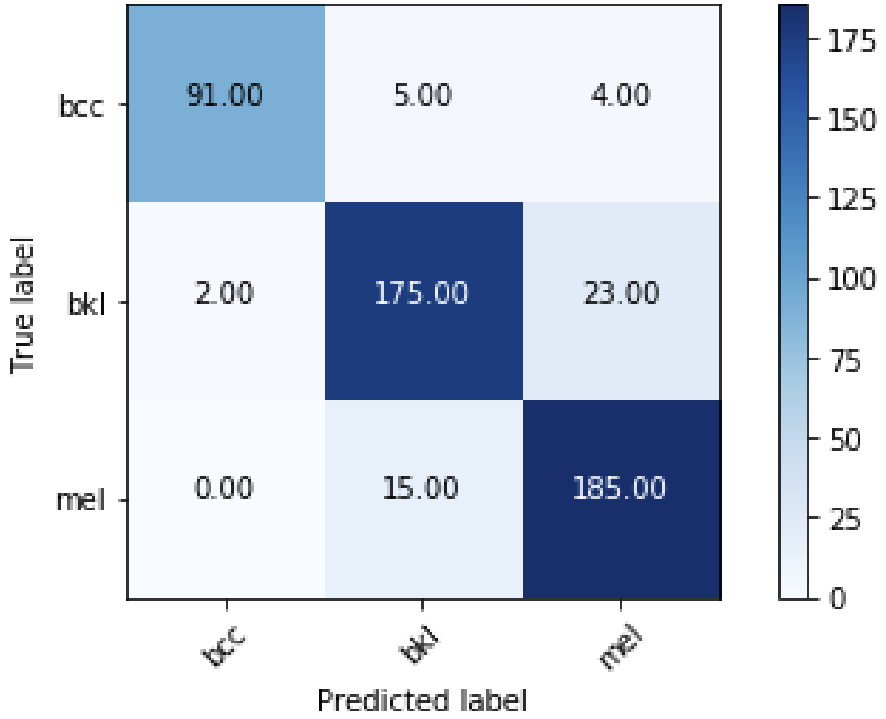
	precision	recall	f1-score	support
class: bcc	0.94	0.80	0.86	100
class: bkl	0.86	0.87	0.86	200
class: mel	0.85	0.91	0.88	200
micro avg	0.87	0.87	0.87	500
macro avg	0.88	0.86	0.87	500
weighted avg	0.87	0.87		

No of errors = 65/500

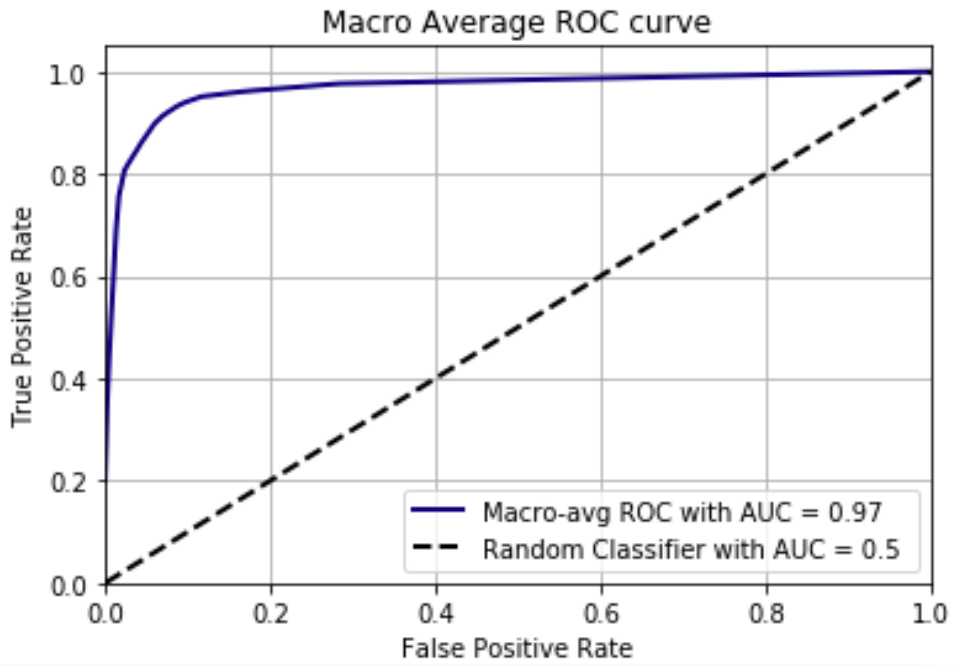
Exp. 7
2000X2
Images + seg.

	precision	recall	f1-score	support
class: bcc	0.91	0.91	0.91	100
class: bkl	0.86	0.85	0.86	200
class: mel	0.87	0.89	0.88	200
micro avg	0.88	0.88	0.88	500
macro avg	0.88	0.88	0.88	500
weighted avg	0.88	0.88	0.88	500

No of errors = 62/500



Ensemble Accuracy on Test Data is 0.902000



Results
Ensembled CNN
Hard Voting

	precision	recall	f1-score	support
class: bcc	0.98	0.91	0.94	100
class: bkl	0.90	0.88	0.89	200
class: mel	0.87	0.93	0.90	200
micro avg	0.90	0.90	0.90	500
macro avg	0.92	0.90	0.91	500
weighted avg	0.90	0.90	0.90	500

No of errors = 49/500

Conclusions

- VGG worked well with the provided data.
- GAP is a good regularizer.
- In transfer learning, use original preprocessing.
- Two-stage training was useful.
- Good segmentation helps.
- Ensembling is better than single classifiers.
- Democracy always wins.