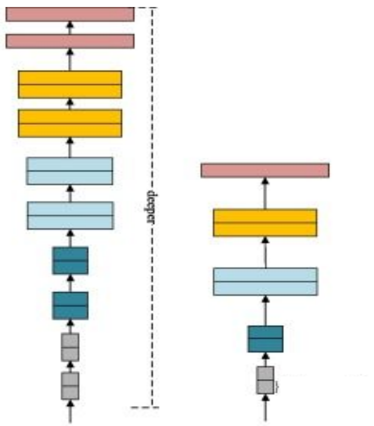
CAD: Skin Lesion Classification
Going Deeper

Manasi Kattel
Vladyslav Zalevskyi

Content

Literature Review



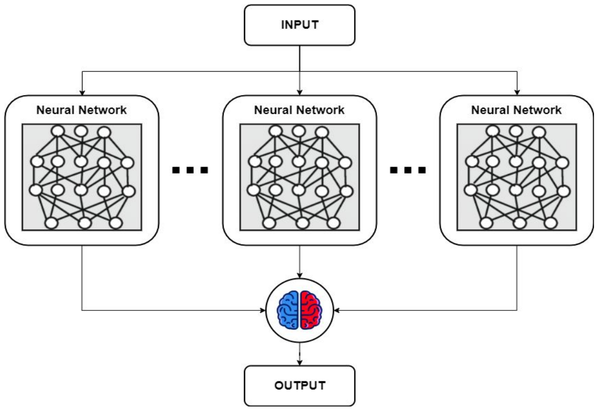
Size matters

ISIC 2019 winning olution [1]:

ensemble of Multi-Res EfficientNets +

SEN154 2

SIIM-ISIC Melanoma Classification winning solution [2]: ensembles of EfficientNet B3-B7, se_resnext101, resnest101



Ensembling matters

Instead of following monstrous ensembles and models we focused on:

- Single model architectures of different styles (convolutional and transformer)
- Tuning the models and the data
- Focus on losses, augmentations and ensembling
- Pretext learning

Literature Review

BACC OF DIFFERENT DCNN MODELS ON THE ISIC 2018 SKIN LESION CLASSIFICATION CHALLENGE TEST SET.

model	BACC	model	BACC	model	BACC
VGG-11	0.769	**DenseNet-169**	0.836	**RegNetX-3.2G**	0.842
VGG-13	0.771	**DenseNet-201**	0.829	**RegNetX-4.0G**	0.834
VGG-16	0.745	**DenseNet-161**	0.837	**RegNetX-8.0G**	0.831
VGG-19	0.750	**EfficientNet-b0**	0.838	**RegNetX-16G**	0.835
ResNet-18	0.812	**EfficientNet-b1**	0.842	**RegNetX-32G**	0.832
ResNet-34	0.825	**EfficientNet-b2**	0.853	**RegNetY-400M**	0.839
ResNet-50	0.834	**EfficientNet-b3**	0.845	**RegNetY-800M**	0.846
ResNet-101	0.838	**EfficientNet-b4**	0.842	**RegNetY-1.6G**	0.850
ResNet-152	0.835	**EfficientNet-b5**	0.843	**RegNetY-3.2G**	**0.858**
SENet-50	0.832	**EfficientNet-b6**	0.848	**RegNetY-4.0G**	0.848
SENet-101	0.845	**EfficientNet-b7**	0.847	**RegNetY-8.0G**	0.846
SENet-152	0.835	**RegNetX-400M**	0.823	**RegNetY-16G**	0.849
SENet-154	0.838	**RegNetX-800M**	0.828	**RegNetY-32G**	0.851
DenseNet-121	0.832	**RegNetX-1.6G**	0.833		

dataset	**7-PT**	**ISIC 2017**	**ISIC 2019**
Best model	RegNet Y-800M	RegNet Y-1.6G	RegNetY -8.0G
Balanced accuracy	0.652	0.743	0.59

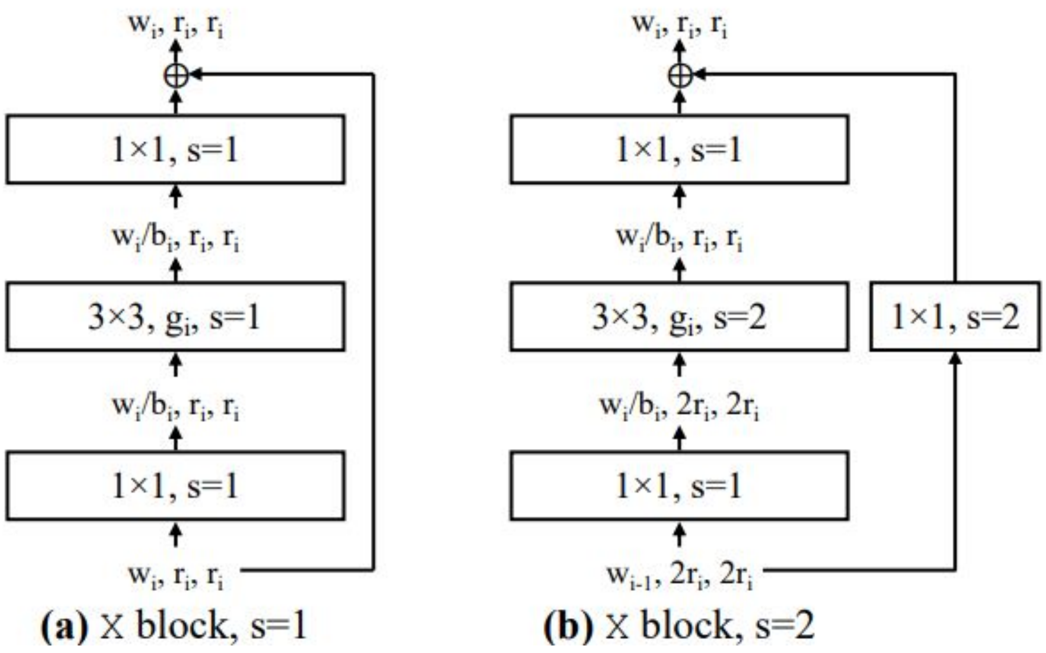
For the transformers we chose Swin architecture

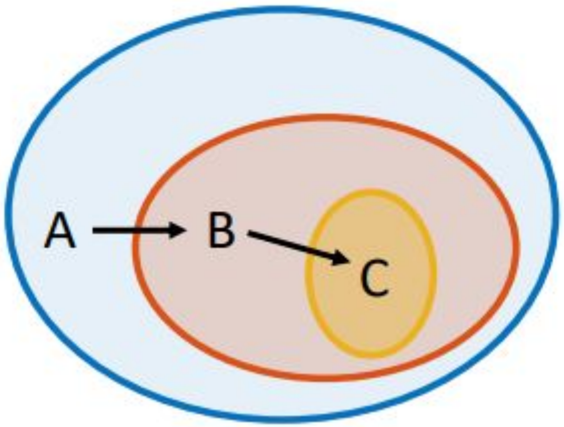
- still one of the best performing single-model architectures on ImageNet
- not very extensive research into transformers and skin lesion cad (not like for convnets)
- easily available with PyTorch

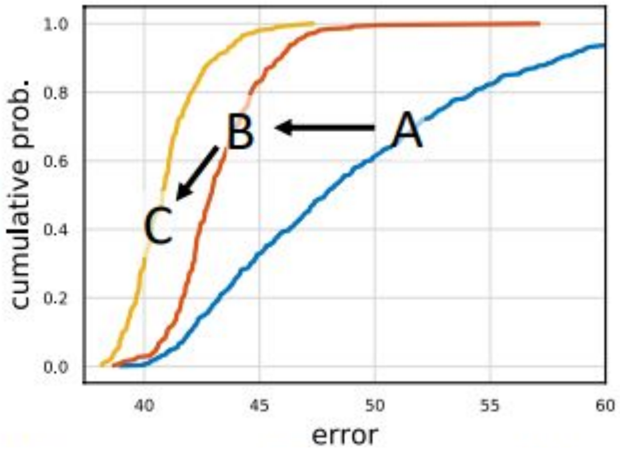
RegNetY

RegNet is a network design space made up of

- Model architectures
- **Different parameters that define a space of possible model architectures**
- Parameters can be the width, depth, groups, etc. of the network.





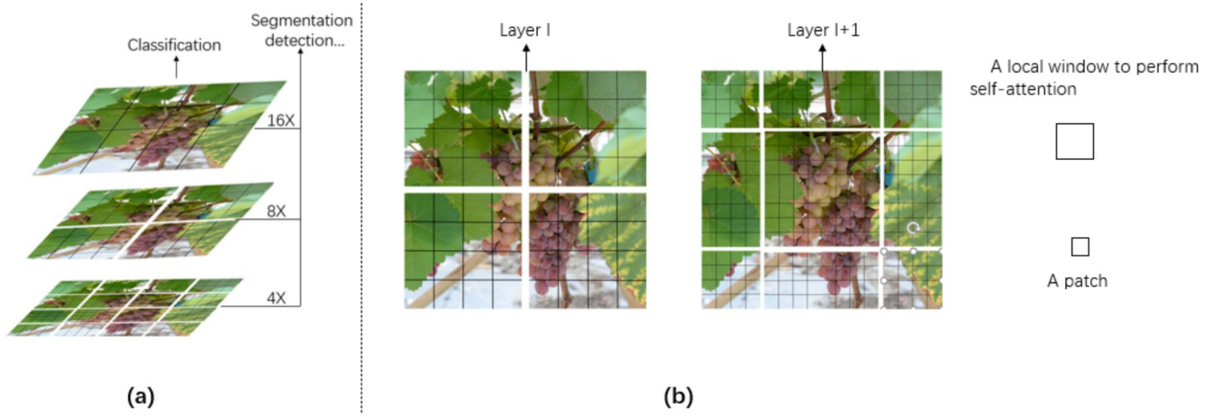


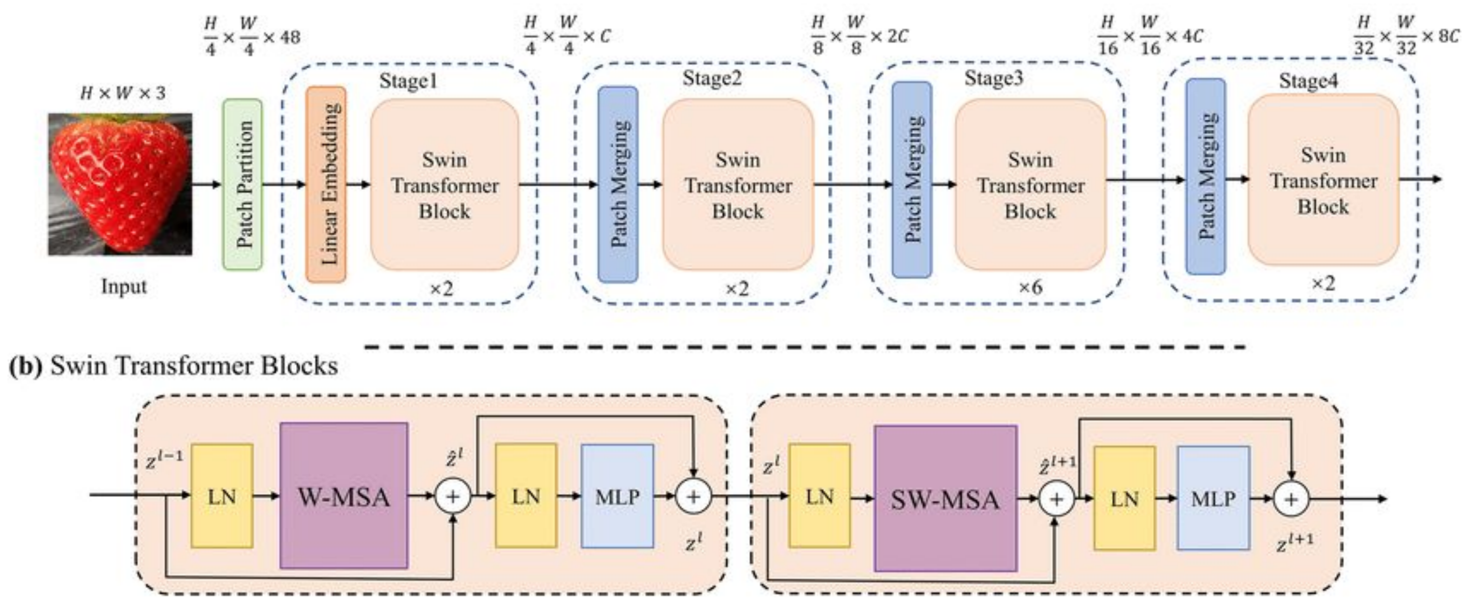
[4]

Swin Transformer

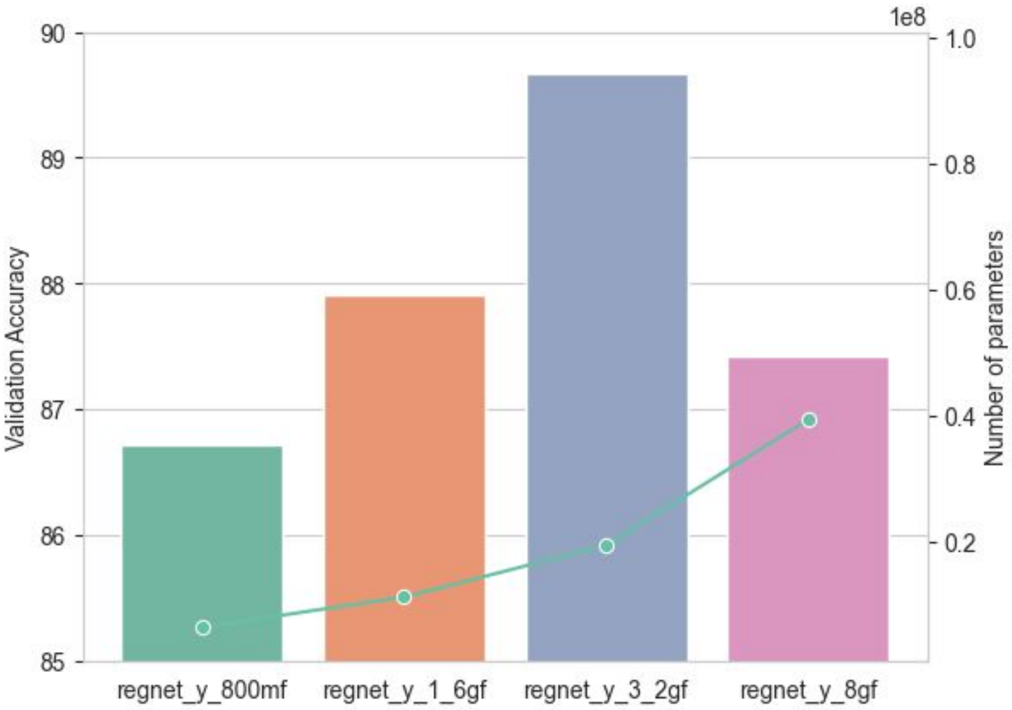
State-of-the-art performance in vision tasks; two key concepts

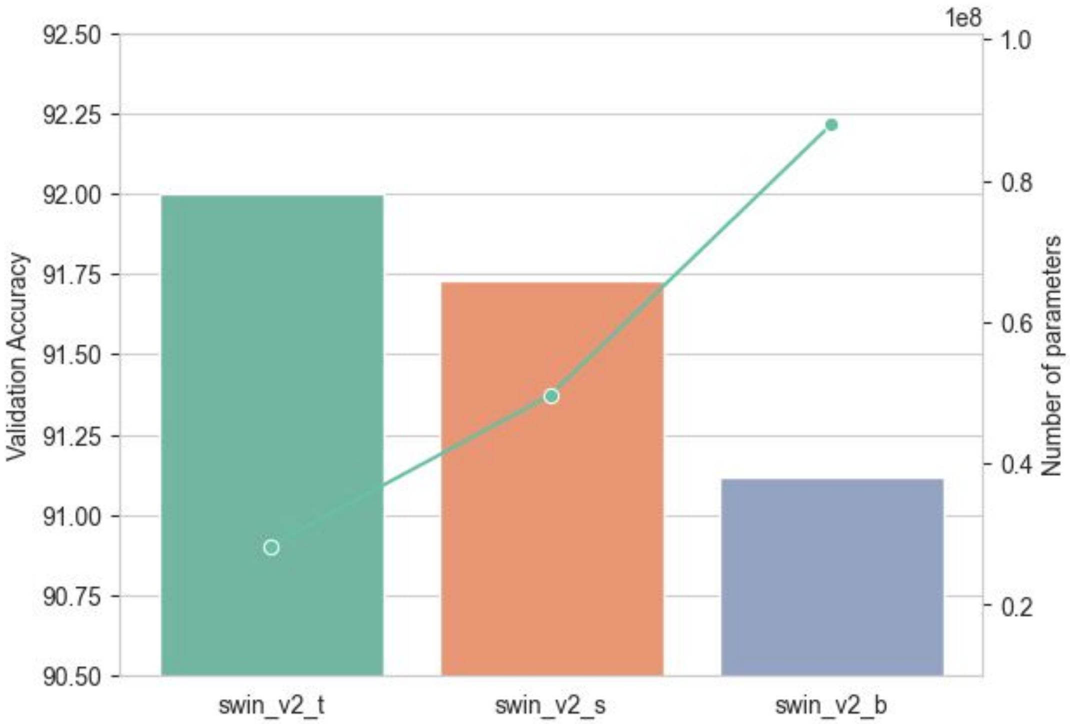
1. **hierarchical feature maps:** allows fine-grained prediction
2. **shifted window attention:** improves complexity





Model sizes experiment





Size greater than regnet_y_3_2gf, started overfitting, and smaller were underfitting!

Size greater than swin_tiny started overfitting!

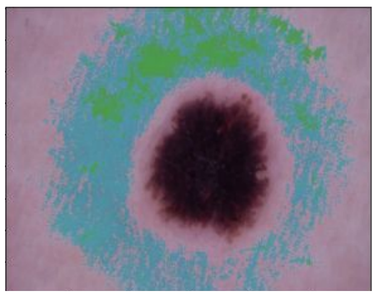
Augmentation

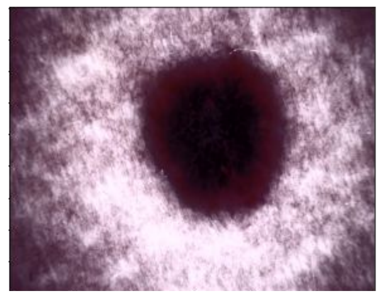
Modified randaugment [3]: 21 transformations(13 colour and 8 shape)

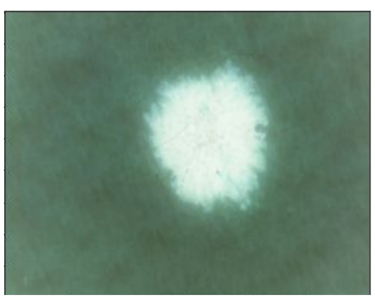
- Randomly select one transformation from {color} transformations, and then randomly select one transformation from {shape} transformations

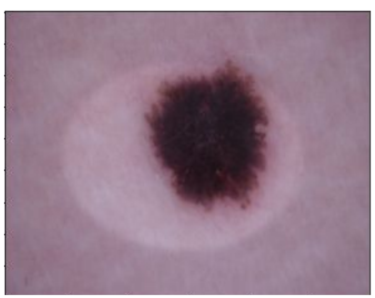
Color transformations

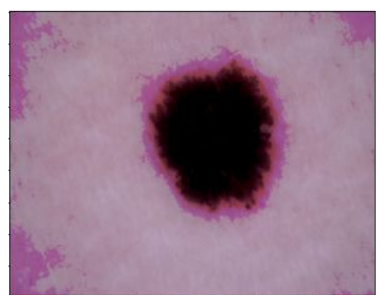

Auto-contrast


Polarize

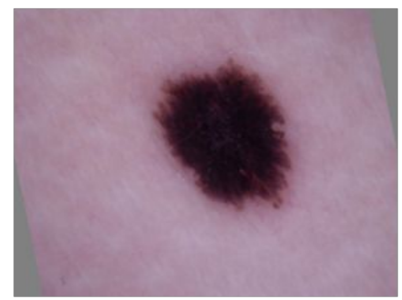

Equalize YUV

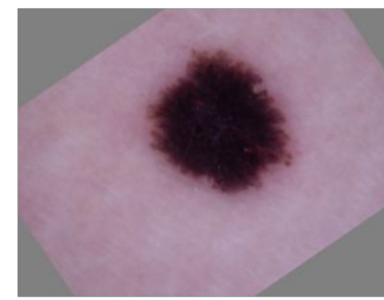

Invert

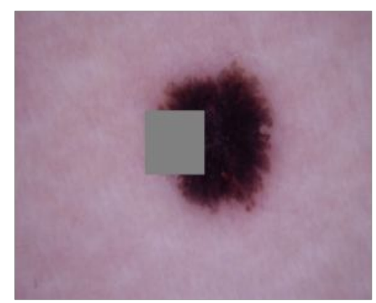

Mixup


Solarize-add

Shape transformations


Shear

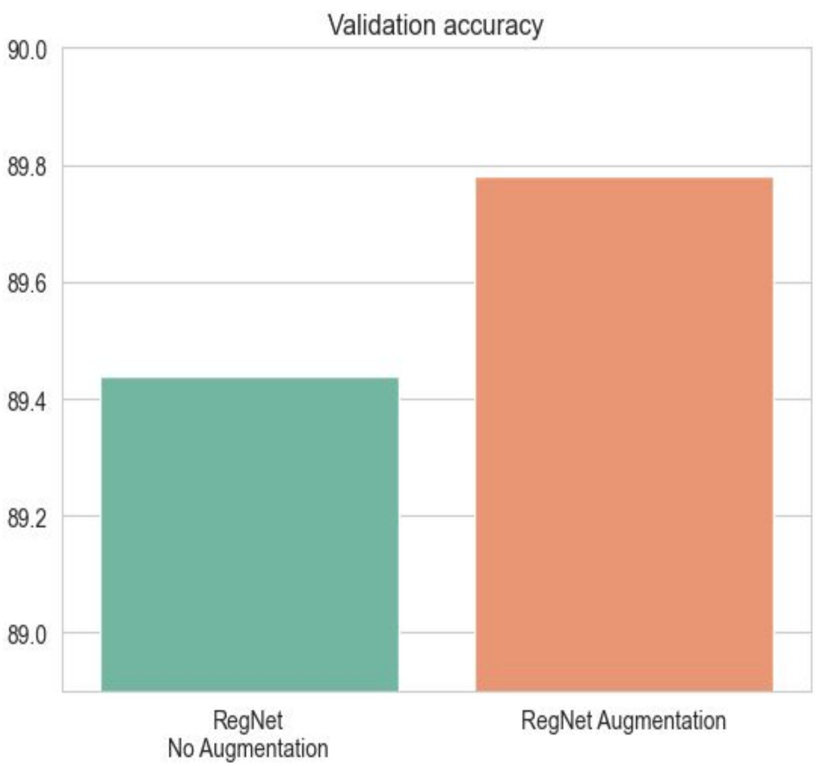

Rotate

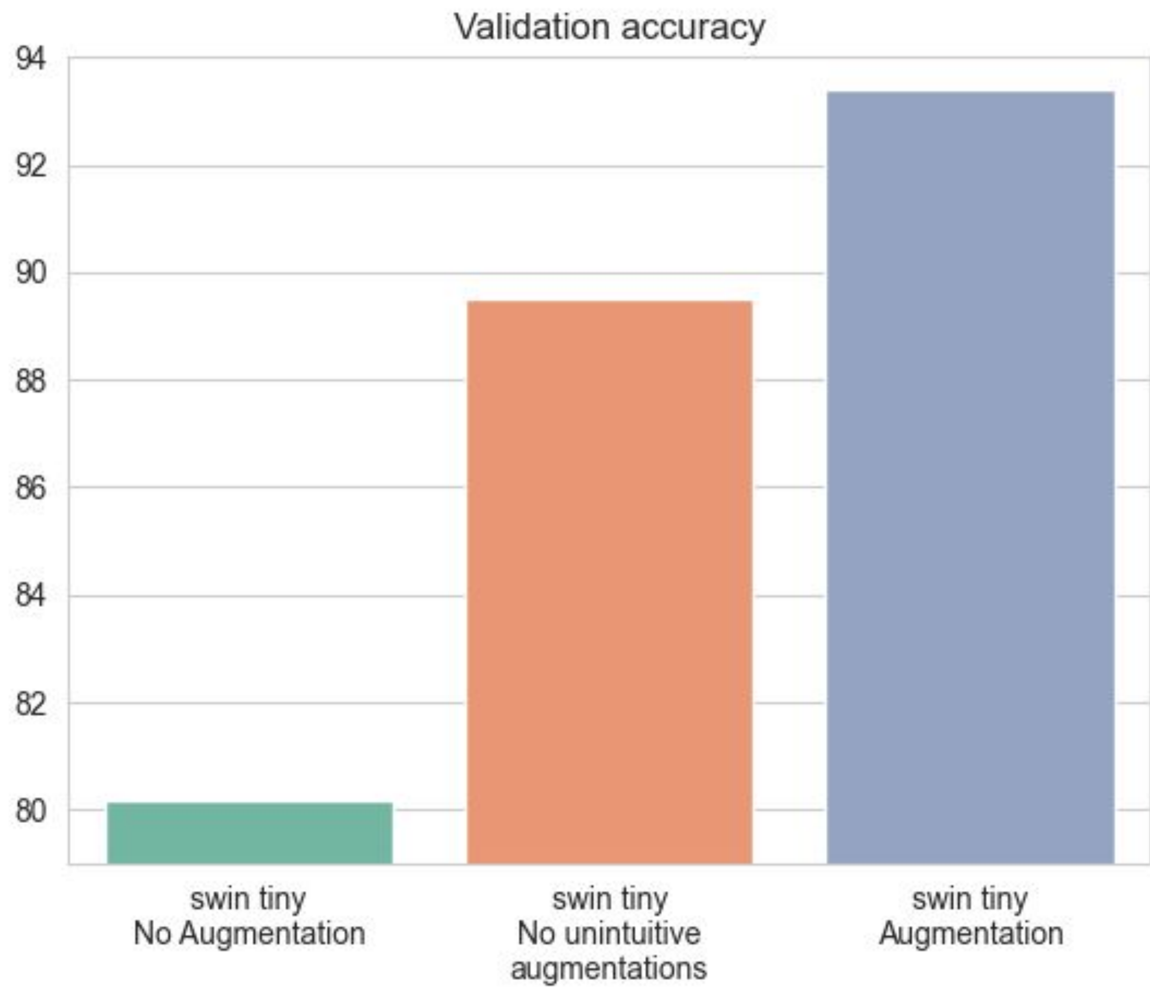

Cutout


Flip

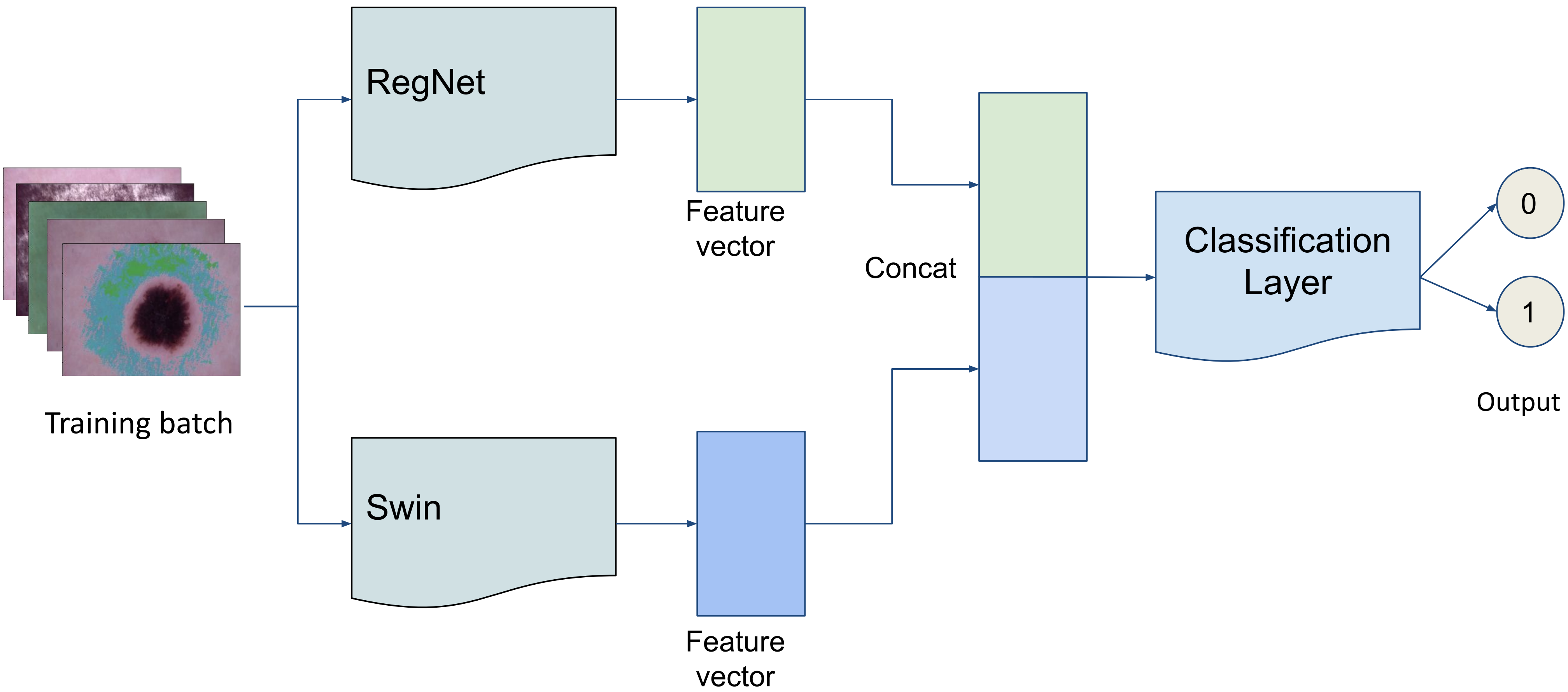
Challenge 1: Augmentation

- Experiments on challenge 1: binary problem





Challenge 1: Ensembling



Challenge 1: Ensembling

Freeze	Freeze the pretrained network and only train the linear layer
No Freeze	Do not freeze any layer on the ensemble model

Loss functions

Challenge 1: Cross-entropy loss.

Challenge 2: Losses that tackle class imbalance.

1. Focal loss

$$\mathrm{FL}(p_\mathrm{t}) = -(1 - p_\mathrm{t})^\gamma \log(p_\mathrm{t}).$$

- where $-\log(p_t)$ is the cross entropy loss
- $(1 - p_t)^\gamma$ is the modulating factor to down-weight easy examples and thus focus training on hard negative.
- The focusing tunable parameter γ smoothly adjusts the rate at which easy examples are down weighted.

Universitat de Girona

Loss functions

2. MWNL Loss [1]:

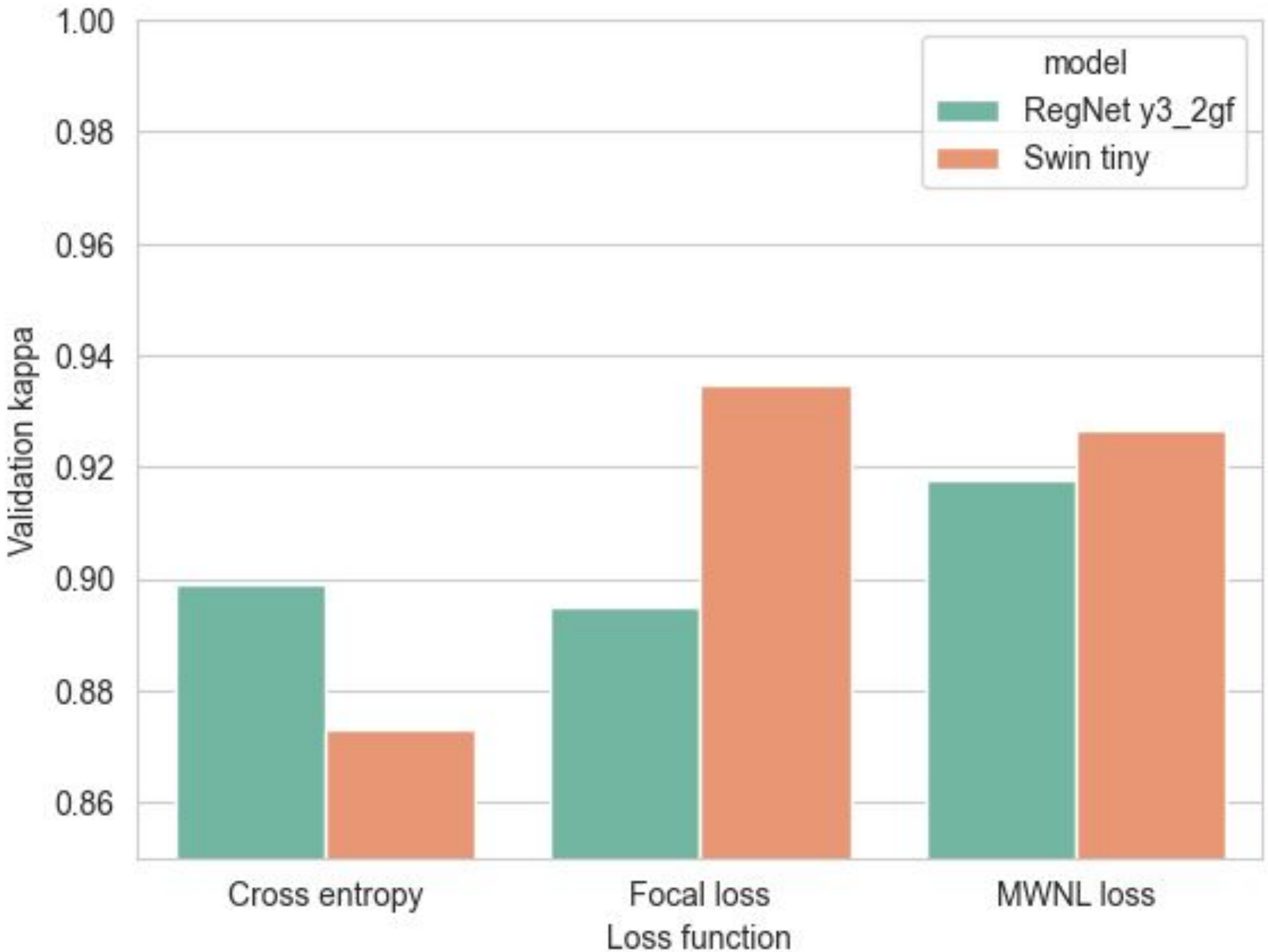
- Overcomes the class imbalance issue in sample number and classification difficulty
- Improves the accuracy of melanoma classification by adjusting the weight of the loss

$$\mathrm{MWNL}(z,y) = -C_y\left(\frac{1}{N_y}\right)^{\alpha}\sum_{i=1}^{C} Loss_i .$$

where

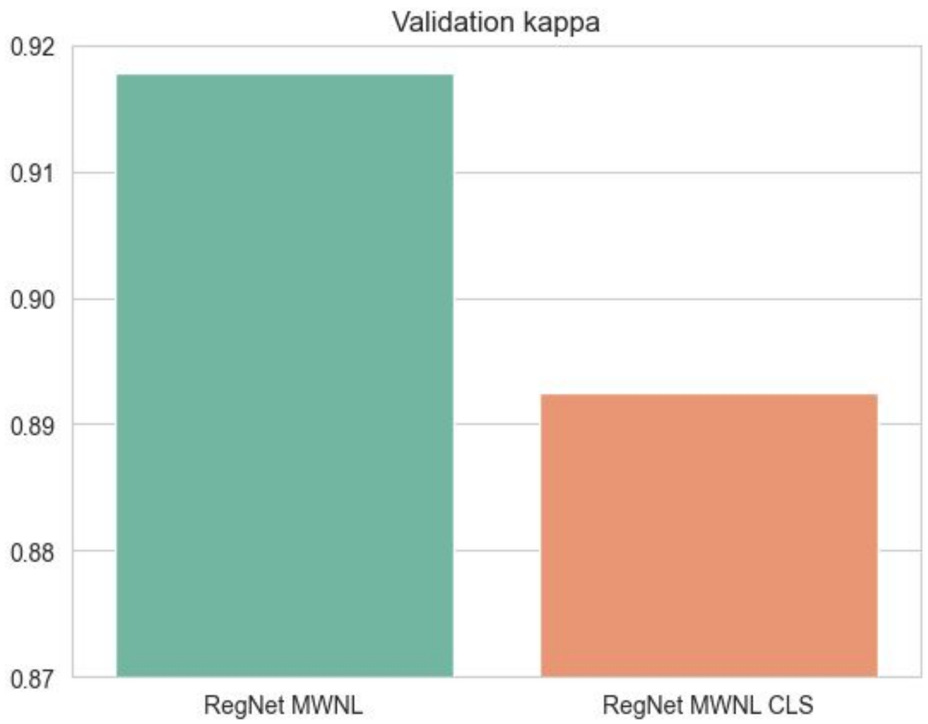
$$Loss_i = \begin{cases} (1-p_i^t)^r \log(p_i^t) & p_i^t > T \\ G^* & p_i^t \le T \end{cases}$$

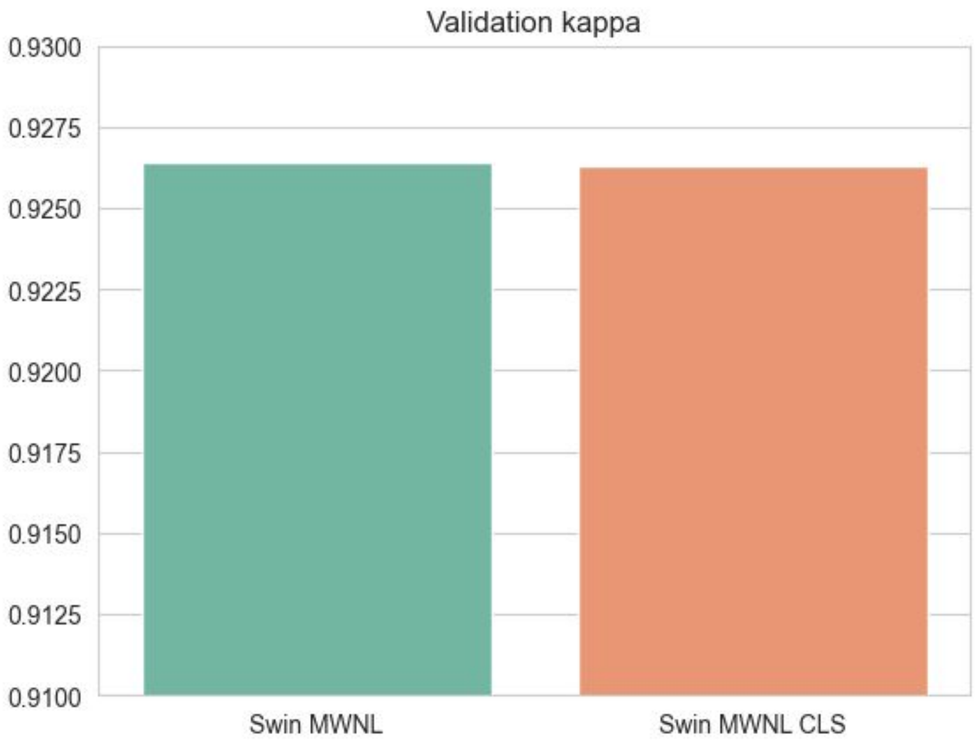
Challenge 2: Loss functions



Challenge 2: Cumulative Learning strategy

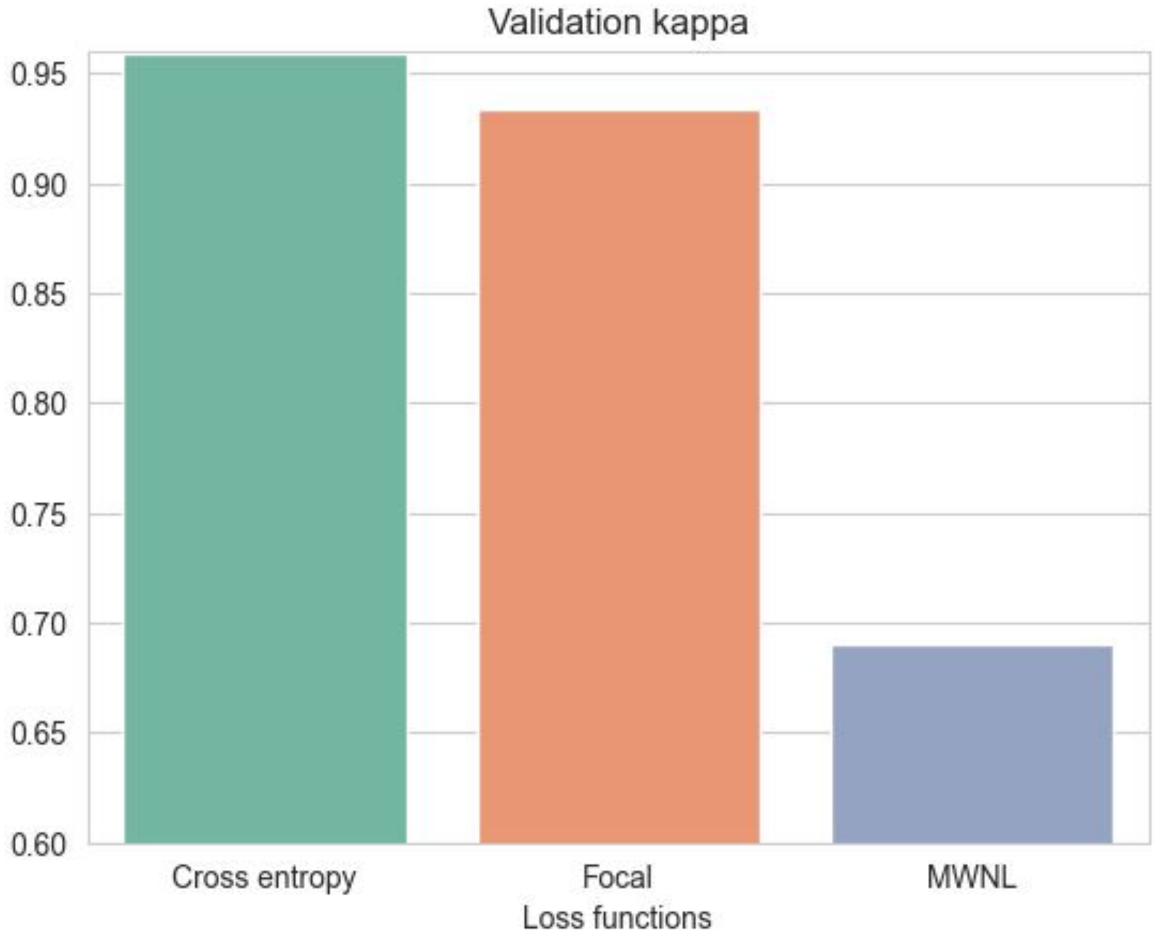
- First train the network on the originally imbalanced data.
- Then change the training gradually to a re-balancing mode.





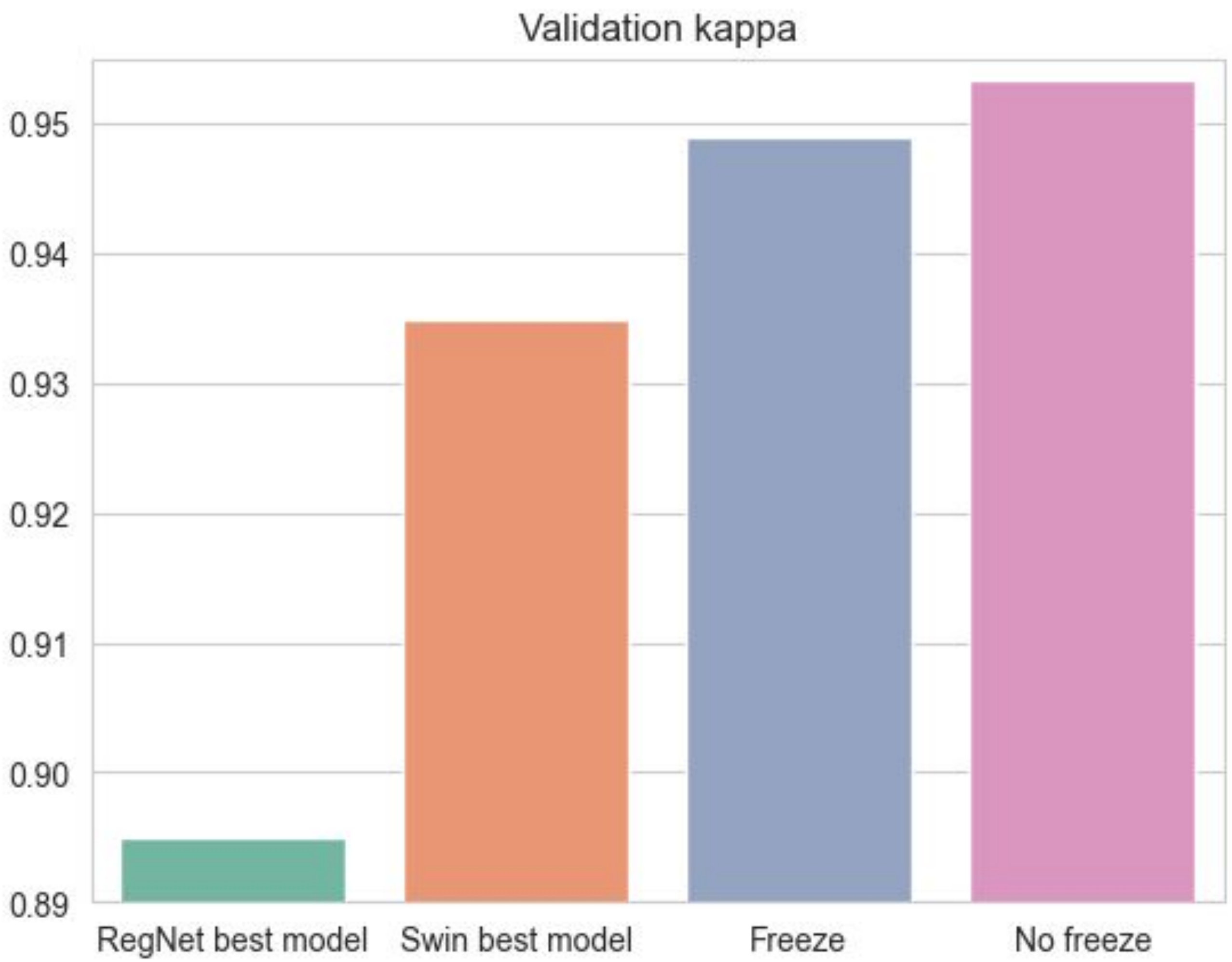
Balanced Sampling

- Weighted sampling of images to get balanced number of images in each batch (swin-tiny)



Challenge 2: Ensembling

Freeze	Freeze the pretrained network and only train the linear layer
No Freeze	Do not freeze any layer on the ensemble model



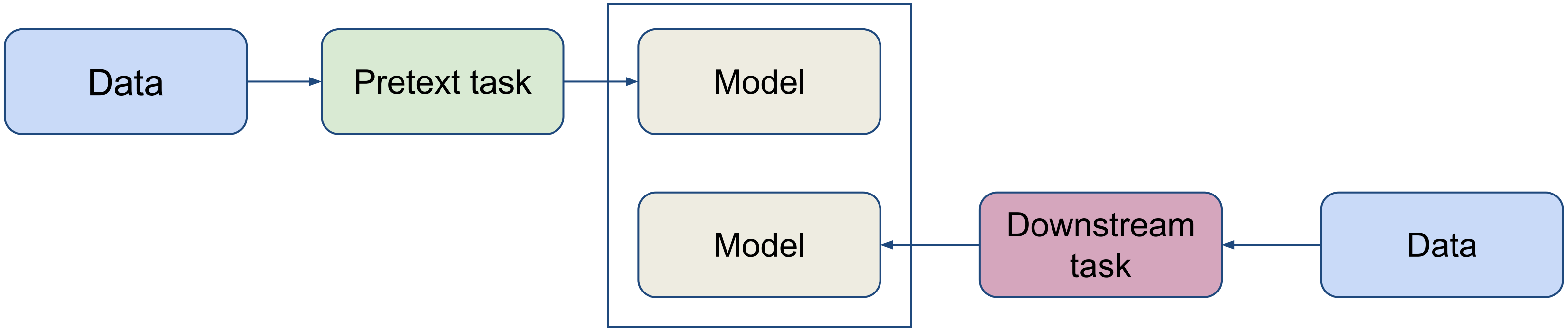
“Pretext learning”

Involves training a model for a task other than what it will actually be trained and used for. This Pretext Training is done prior to actual training of the model.

Needed to be performed with our tested models.

Pretext task to learn:

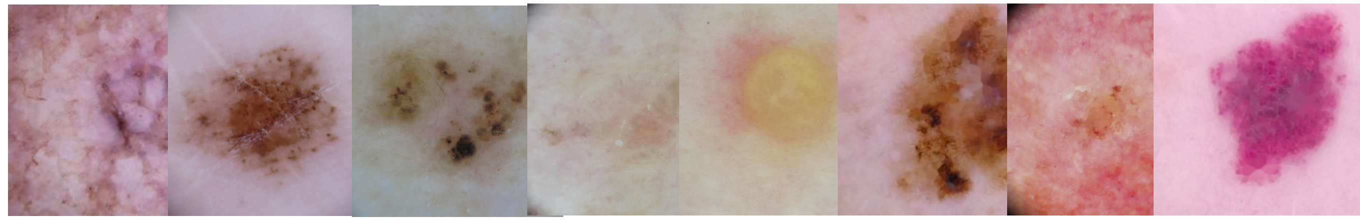
- lesion size
- lesion colors
- abcd scores
- other relevant patient medical data



Shared architecture/weights

“Pretext learning”

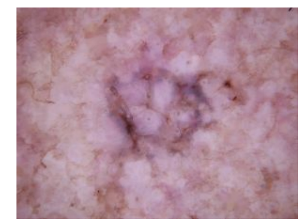
Pretext task
Eight class classification



ack nevus bcc bkl def mel scc vac

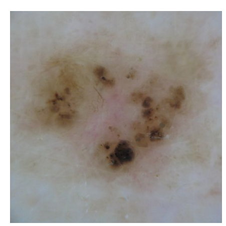
Challenge 1
Binary classification

nevus

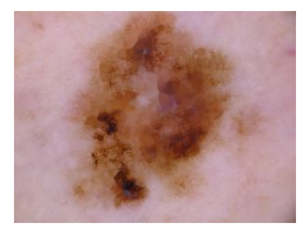


ack/bcc/bkl/def/mel/scc/vac

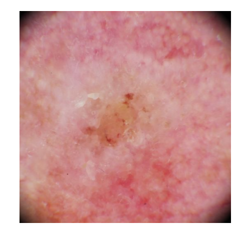
Challenge 2
Three-class classification



bcc

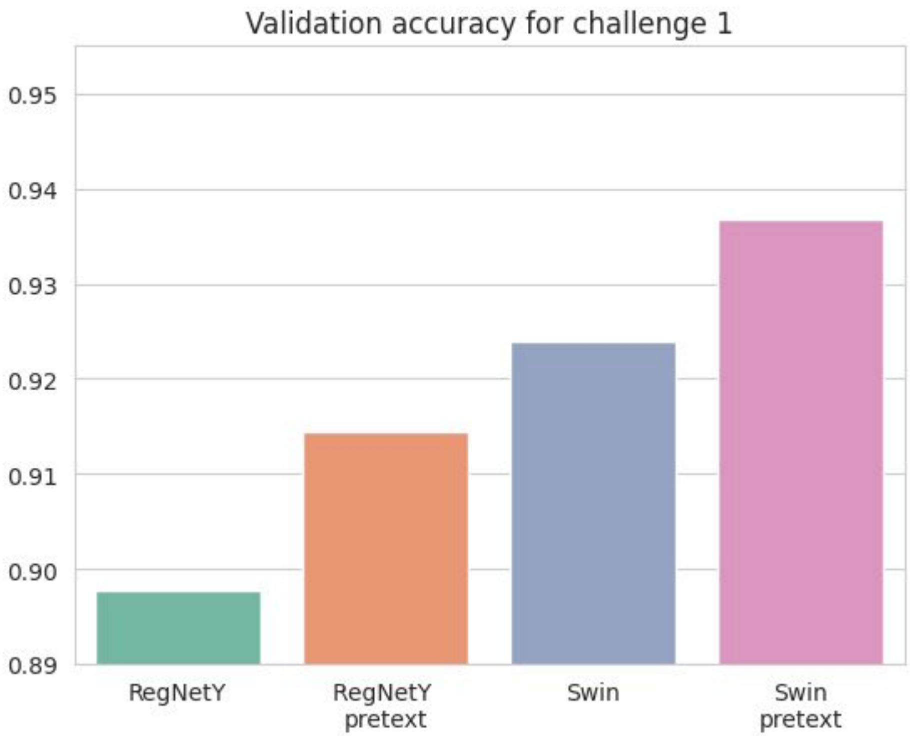


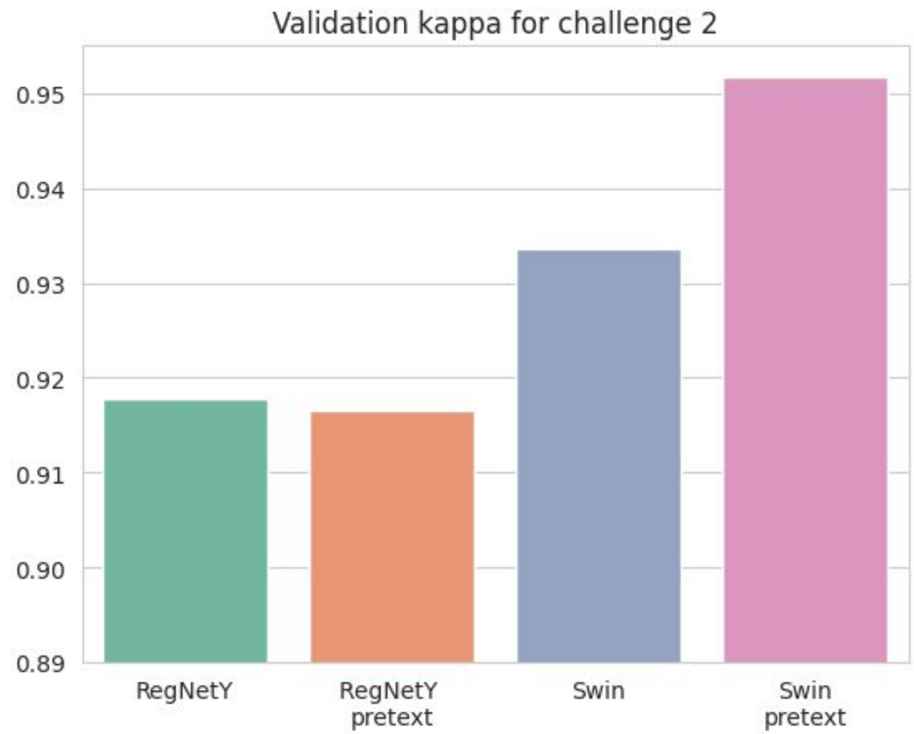
mel



scc

“Pretext learning” results





Both Swin and RegNetY improved performance with the pretext task for challenge 1.

Only Swin was able to maintain information learned during pretext training at challenge 2 training due it it’s bigger size and memory.

RegNetY - 0.818

Swin - 0.835

Universitat de Girona

Final models

Challenge 1

Ensemble (learnable feature fusion)

- RegNetY-3.2GF (with pretext initialization)
- Swin-v2-Tiny (with pretext initialization)

RandAugment

Cross entropy loss

Validation accuracy: 0.936

Challenge 2

Ensemble (learnable feature fusion)

- RegNetY-3.2GF (without pretext initialization challenge 1 transfer learning)
- Swin-v2-Tiny (with pretext initialization and challenge 1 transfer learning)

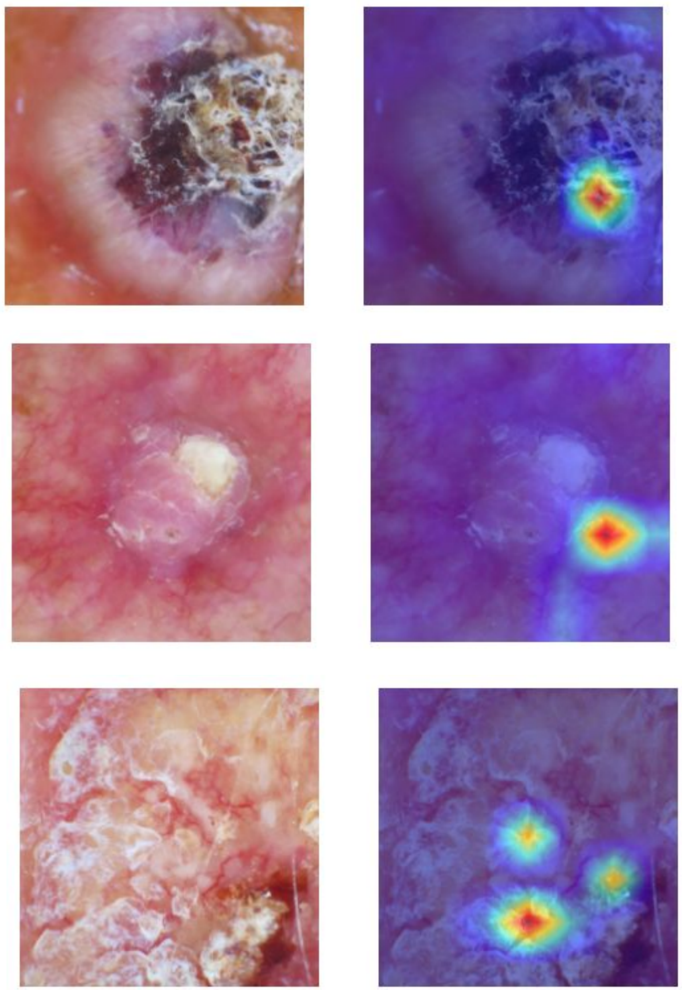
RandAugment

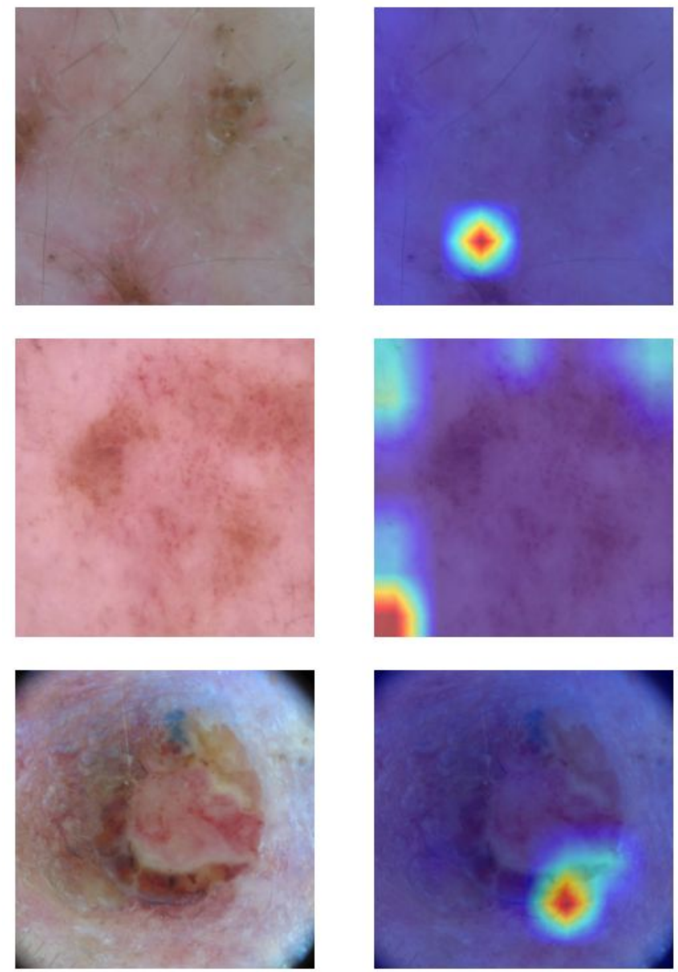
MWNL loss

Validation kappa: 0.9533

Grad-CAM

Grad-CAM of Correctly vs. incorrectly classified skin lesions





Conclusions

- Strong augmentations push models to learn a more robust set of features.
- Ensembling is a powerful tool that allowed us to combain and benefit from 2 different feature embeddings of convolutional and transformer models.
- Balanced sampling did help training the models and so did using sample-weight sensitive losses like focal or mwnl did.
- Bigger model sized are more prone to overfitting so the size needs to be fine-tuned depending on the problem and dataset.
- Pretext learning has great potential to improve the results, however the more training or fine tuning we perform over the model the more the initial weights change; only swin was able to benefit from it after challenge 1 and 2 fine tuning.

References

[1]https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/175412

[2] https://challenge.isic-archive.com/landing/2019/

[3] Yao, Peng & Shen, Shuwei & Xu, Mengjuan & Liu, Peng & Zhang, Fan & Xing, Jinyu & Shao, Pengfei & Kaffenberger, Benjamin & Xu, Ronald. (2021). Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification.

[4] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He and P. Dollár, "Designing Network Design Spaces," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10425-10433, doi: 10.1109/CVPR42600.2020.01044.

[5]Wang J, Zhang Z, Luo L, Zhu W, Chen J, Wang W. SwinGD: A Robust Grape Bunch Detection Model Based on Swin Transformer in Complex Vineyard Environment. Horticulturae. 2021; 7(11):492. https://doi.org/10.3390/horticulturae7110492

[6] Zheng, Hao & Wang, Guohui & Li, Xuchen. (2022). Swin-MLP: a strawberry appearance quality identification method by Swin Transformer and multi-layer perceptron. Journal of Food Measurement and Characterization. 16. 1-12. 10.1007/s11694-022-01396-0.