Deep Transfer Learning for Automated Diagnosis of Skin Lesions from Photographs

Doyoon Kim*[1], Emma Rocheteau*[2]

[1]Cleveland High School, California, US; doyoondeankim@gmail.com

[2]University of Cambridge, UK; ecr38@cam.ac.uk

*equal contribution

NEURAL INFORMATION PROCESSING SYSTEMS

Summary

- Melanoma is the deadliest form of skin cancer and requires expert knowledge by practitioners for diagnosis, which can be costly and inaccessible in certain parts of the world.
- With advancements in deep learning for improved diagnosis and increasing access to mobile technologies, this technology can be used for fast and efficient diagnosis.
- We compared the performance of several neural network architectures with and without transfer learning.

Model Architecture

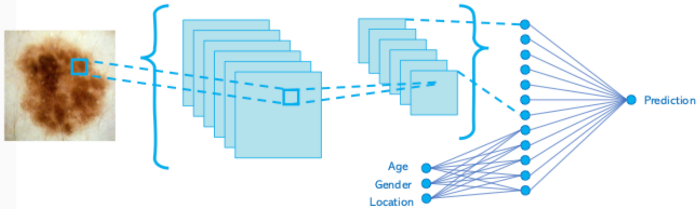

Figure 1: Model architecture. The CNN is different in each experiment.

- The CNN component (indicated in brackets) is different for each experiment. The static data is processed separately and concatenated to the CNN output before a final prediction is made.

Datasets

- We used the International Skin Imaging Collaboration (ISIC) 2020 dataset. It contains labelled photographs of skin lesions taken from various locations on the body.
- In total our data contained 37,648 skin lesion images.

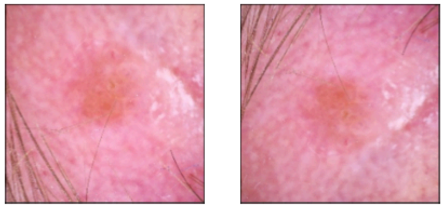

Figure 2: Example photographs in the training data. The original data is shown on the left and the augmented image is shown on the right.

- We performed data augmentation on the training data to introduce small variations in the form of random rotations, flipping, resizing, saturation shifts, etc. An example is shownabove.

Accuracy Metrics

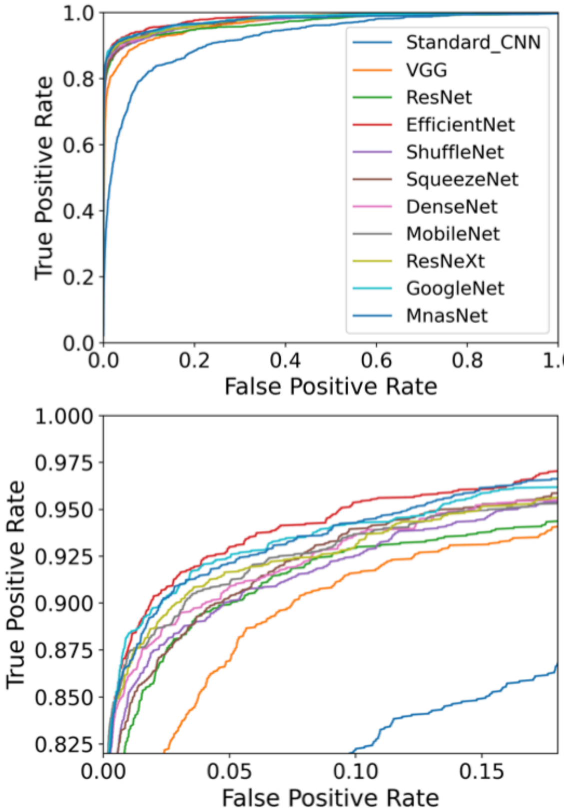

Figure 3: ROC curves of the TL models and Standard CNN.

Integrated Gradients

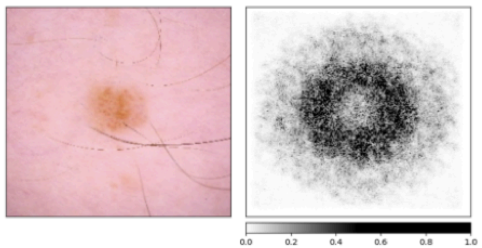

Figure 4: An image and its integrated gradient attributions (standard CNN).

Results

(a)

Model	Accuracy	AUROC	AUPRC	F1 Score
Standard CNN	0.914±0.004	0.759±0.030	0.484±0.026	0.633±0.035
General Practitioners†	-	0.83±0.03	-	-
VGG [30]	0.943±0.004	0.832±0.018	0.643±0.025	0.765±0.023
SqueezeNet [10]	0.949±0.003	0.860±0.014	0.687±0.011	0.801±0.008
ResNeXt [37]	0.952±0.009	0.878±0.022	0.712±0.035	0.818±0.023
DenseNet [9]	0.957±0.003	0.859±0.015	0.733±0.021	0.824±0.018
GoogleNet [33]	0.957±0.004	0.861±0.018	0.732±0.024	0.824±0.022
ResNet-50 [6]	0.959±0.003	0.869±0.016	0.744±0.018	0.835±0.016
MobileNet [29]	0.963±0.003	0.889±0.013	0.769±0.019	0.856±0.014
MnasNet [35]	0.963±0.008	0.900±0.010	0.771±0.039	0.859±0.023
ShuffleNet [18]	0.965±0.004	0.892±0.016	0.777±0.025	0.861±0.018
EfficientNet [17]	0.967±0.002	0.900±0.009	0.794±0.013	0.872±0.010
Dermatologists†	-	0.91±0.02	-	-

(b)

Model	Accuracy	AUROC	AUPRC	F1 Score
General Practitioners†	-	0.83±0.03	-	-
VGG [30]	0.959±0.003**	0.874±0.013**	0.740±0.016**	0.835±0.013**
ResNet-50 [6]	0.962±0.004	0.880±0.014	0.763±0.022	0.849±0.017
ShuffleNet [18]	0.963±0.006	0.896±0.024	0.769±0.040	0.857±0.028
SqueezeNet [10]	0.963±0.004**	0.902±0.015**	0.771±0.020**	0.861±0.015**
DenseNet [9]	0.966±0.003**	0.904±0.011**	0.786±0.018**	0.870±0.011**
Dermatologists†	-	0.91±0.02	-	-
MobileNet [29]	0.969±0.002**	0.916±0.007**	0.806±0.015**	0.884±0.009**
ResNeXt [37]	0.971±0.001**	0.918±0.006**	0.819±0.009**	0.891±0.005**
GoogleNet [33]	0.973±0.002**	0.921±0.006**	0.831±0.013**	0.898±0.008**
MnasNet [35]	0.974±0.002*	0.928±0.005**	0.832±0.013**	0.901±0.007**
EfficientNet [17]	0.975±0.002**	0.931±0.005**	0.840±0.010**	0.906±0.006**

Table 1: Performance of the models. (a) is without transfer learning, (b) is with.

- The error margins are 95% confidence intervals (CIs). We report the accuracy, area under the receiver operating characteristic curve (AUROC), area under the precision recall curve (AUPRC) and the F1 Score.
- Within each table, the results are ordered from least to best performance.
- In table (b), if the result is statistically better than the model without transfer learning in a one-tailed t-test ($p < 0.05$* and $p < 0.001$**), then it is indicated with stars.
- Results that significantly outperform general practitioners and dermatologists on AUROC (determined by a recent meta-analysis†[24]) are indicated in **green** and **blue** respectively ($p < 0.05$).

Conclusion

- We have demonstrated the benefit of transfer learning for melanoma diagnosis. EfficientNet and MnasNet were capable of outperforming dermatologists.
- In future work, we would like to extend our binary classification task to multiclass (other skin lesions such as benign keratosis, basal cell carcinoma, etc.
- GitHub: https://github.com/aimadeus/Transfer_learning_melanoma.