Skin Cancer Detection

Giovanni Budi, Andrew Chisolm, Noah Parker

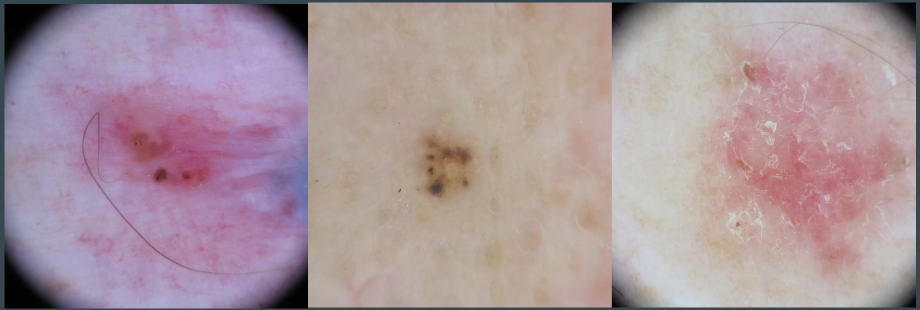
Motivation

- Early detection of skin cancer improves survival rates
- Extensive research has been done utilizing ML algorithms to classify images of skin lesions
- Most research uses *dermoscopic* images, which can only be taken from a healthcare provider
  - Limited access to specialized dermatological care in some regions
- To help detect skin cancer in underserved regions, we are using smartphone-quality images to give a simple malignant/benign classification

Two Datasets

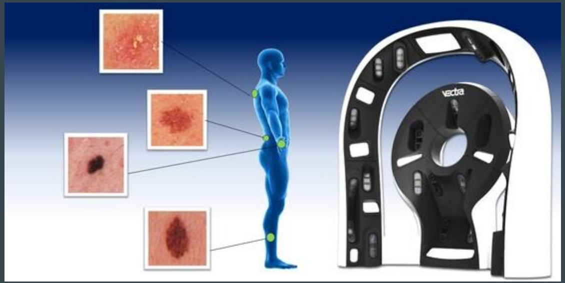
ISIC 2019

- Compilation of 25,331 dermoscopic images
- High quality
- Labelled into nine categories (some malignant, some benign)
- Sourced from Barcelona, Austria, and Australia



ISIC 2024

- Over 400,000 images cropped from 3D total body photographs
- Lower quality, closely resembles the quality of a cell phone camera
- Labelled into two categories (benign and malignant)
- Sourced from nine sites across Europe, Australia, and the United States



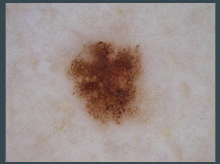
Approach

Three Models:

- ISIC 2019 Model
  - Replicating previous work
- ISIC 2024 Model
  - Using similar techniques from 2019 on the new data
- Fine-Tuned 2019/2024 Model
  - Use information gathered from the 2019 model to help inform the 2024 model

Key Questions:

- Is it possible to achieve the same positive results that the 2019, high quality data achieved on the 2024 data?
- What other insights can we gain on how these two datasets are different?



```
# Save best predictions
allData['bestPred'][cv] = predictions
allData['targets'][cv] = targets
# Write to File
with open(mdlParams['saveDirBase'] + '/CV.pkl', 'wb') as f:
    pickle.dump(allData, f, pickle.HIGHEST_PROTOCOL)
# Delte previously best model
if os.path.isfile(mdlParams['saveDir'] + '/checkpoint_best-' + str(oldBestInd) + '.pt'):
    os.remove(mdlParams['saveDir'] + '/checkpoint_best-' + str(oldBestInd) + '.pt')
# Save currently best model
state = {'epoch': step, 'valBest': mdlParams['valBest'], 'lastBestInd': mdlParams
['lastBestInd'], 'state_dict': modelVars['model'].state_dict(),'optimizer': modelVars
['optimizer'].state_dict()}
torch.save(state, mdlParams['saveDir'] + '/checkpoint_best-' + str(step) + '.pt')
```

Methodology

Image Preprocessing:

- Training augmentations: Resized crop, flip, rotation, color jitter, cutout, normalization.
- Validation: CenterCrop, fixed cropping, normalization.

Training Parameters:

- EfficientNet-B0 model with Batch size: 20, Learning rate: 0.000015, and 60 epochs.

Evaluation:

- Accuracy, Mean Accuracy, Sensitivity, Specificity, AUC.

Technical Challenges

- Class imbalance in datasets (e.g., 1:20 malignant-to-benign ratio).
- Variability in image quality between high-resolution (2019 dataset) and smartphone-quality images (2024 dataset).
    - Limited diagnostic accuracy with low-quality images.
- Run time

Results

Table 1.1: 2019 Model Results

Fold	Epoch	Accuracy	F1	Mean AUC	Weighted Accuracy	Specificity	Sensitivity
0	60	0.7735	0.8114	0.8826	0.7934	0.7330	0.8538
1	60	0.7777	0.8138	0.8812	0.7971	0.7365	0.8576
2	60	0.7811	0.8241	0.8699	0.7857	0.7717	0.7998
3	60	0.7833	0.8212	0.8835	0.7993	0.7503	0.8483
4	60	0.7799	0.8164	0.8780	0.7967	0.7435	0.8500
		0.7791	0.8174	0.8790	0.7944	0.7470	0.8419

Table 1.2: 2024 Model Results

Fold	Epoch	Accuracy	F1	Mean AUC	Weighted Accuracy	Specificity	Sensitivity
0	30	0.8752	0.9308	0.9045	0.8249	0.8805	0.7692
1	20	0.8940	0.9420	0.9334	0.8637	0.8966	0.8308
2	30	0.9104	0.9507	0.9411	0.8716	0.9153	0.8280
3	30	0.8715	0.9283	0.9055	0.8435	0.8745	0.8125
4	40	0.8624	0.9229	0.9316	0.8476	0.8640	0.8312
		0.8827	0.9349	0.9232	0.8502	0.8862	0.8143

Table 1.3: 2019/2024 Fine Tuned Model

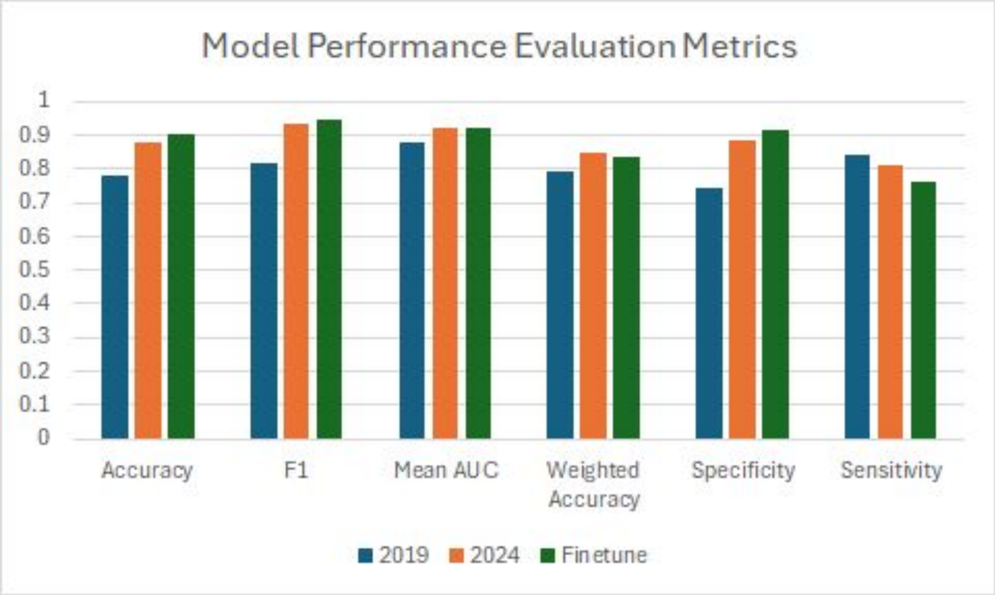
Fold	Epoch	Accuracy	F1	Mean AUC	Weighted Accuracy	Specificity	Sensitivity
0	40	0.8873	0.9379	0.9186	0.8312	0.8932	0.7692
1	60	0.9116	0.9523	0.9319	0.8212	0.9193	0.7231
2	50	0.9122	0.9518	0.9373	0.8625	0.9185	0.8065
3	30	0.9188	0.9560	0.9016	0.8446	0.9268	0.7625
4	60	0.9024	0.9468	0.9166	0.8253	0.9104	0.7403
		0.9065	0.9490	0.9212	0.8370	0.9136	0.7603

Results Continued ...

Table 1.4 : Model Performance Evaluation Metrics

Model	Accuracy	F1	Mean AUC	Weighted Accuracy	Specificity	Sensitivity
2019	0.7791	0.8174	0.8790	0.7944	0.7470	0.8419
2024	0.8827	0.9349	0.9232	0.8502	0.8862	0.8143
Finetune	0.9065	0.9490	0.9212	0.8370	0.9136	0.7603

- Improvements
  - Accuracy, F1, Specificity
- Consistent Metrics
  - Mean AUC, Weighted Accuracy
- Challenges
  - Sensitivity



Conclusion

- Key Takeaways
  - Transfer Learning Benefits: Fine-tuning significantly improved performance on lower-quality images, demonstrating the strength of transfer learning.
  - Specificity Gains: Fine-tuned model reduced false positives, making it more effective in distinguishing benign cases.
  - Challenges with Sensitivity
- Future Direction
  - Explore oversampling, class balancing, or alternative architectures to improve sensitivity.
  - Investigate data augmentation or super-resolution techniques to address image quality issues.
  - Further refine the model for better generalization across datasets with varying quality.

References

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging* (ISBI 2018), pp. 168–172. IEEE, 2018.

Combalia, M., Codella, N. C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Carrera, C., Barreiro, A., Halpern, A. C., Puig, S., et al. Bcn20000: Dermoscopic lesions in the wild. *arXiv preprint arXiv:1908.02288*, 2019.

Gessert, N., Nielsen, M., Shaikh, M., Werner, R., and Schlaefer, A. Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. *Methods X*, 7(100864), 2020.

Kurtansky, N. R., D'Alessandro, B. M., Gillis, M. C., Betz-Stablein, B., Cerminara, S. E., Garcia, R., Girundi, M. A., Goessinger, E. V., Gottfrois, P., Guitera, P., et al. The slice-3d dataset: 400,000 skin lesion image crops extracted from 3d tbp for skin cancer detection. *Scientific Data*, 11(1):884, 2024.

Tschandl, P., Rosendahl, C., and Kittler, H. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(180151), 2018.