AI ASSISTED MOLE DETECTION FOR ONLINE DERMATOLOGY TRIAGE IN TELEMEDICINE SETTINGS

Presented by Debarpan Das
Radiology and Diagnostic Imaging
University of Alberta
Date: 11 August, 2023

Supervisor: Professor Kumaradevan Punithakumar
Co-supervisor: Professor Derek Emery

UNIVERSITY OF ALBERTA

CONTENTS

- Goal and Motivation
- Relevant Background
- Datasets and Models
- Experimental Setup
- Results and analysis
- Conclusion
- Publication and references

Goal and Motivation

- Moles range from harmless, atypical to melanoma
- Without adequate knowledge, difficult to understand severity
- Teledermatology:
  1. Fast and easily accessible diagnosis
  2. Condition evaluation
  3. Triaging and streamlining clinical workflow
  4. Reduce wait time (92 days to 2 days)
- Moles are preferably treated in person
  1. In person dermoscopic images for diagnosis and assessment
  2. Minimize false negatives (High recall)
- Automatic system to detect moles from patient uploaded images
- If present, warn the patient and recommend to see dermatologist

Immersion oil dermatoscope. Source: [1]

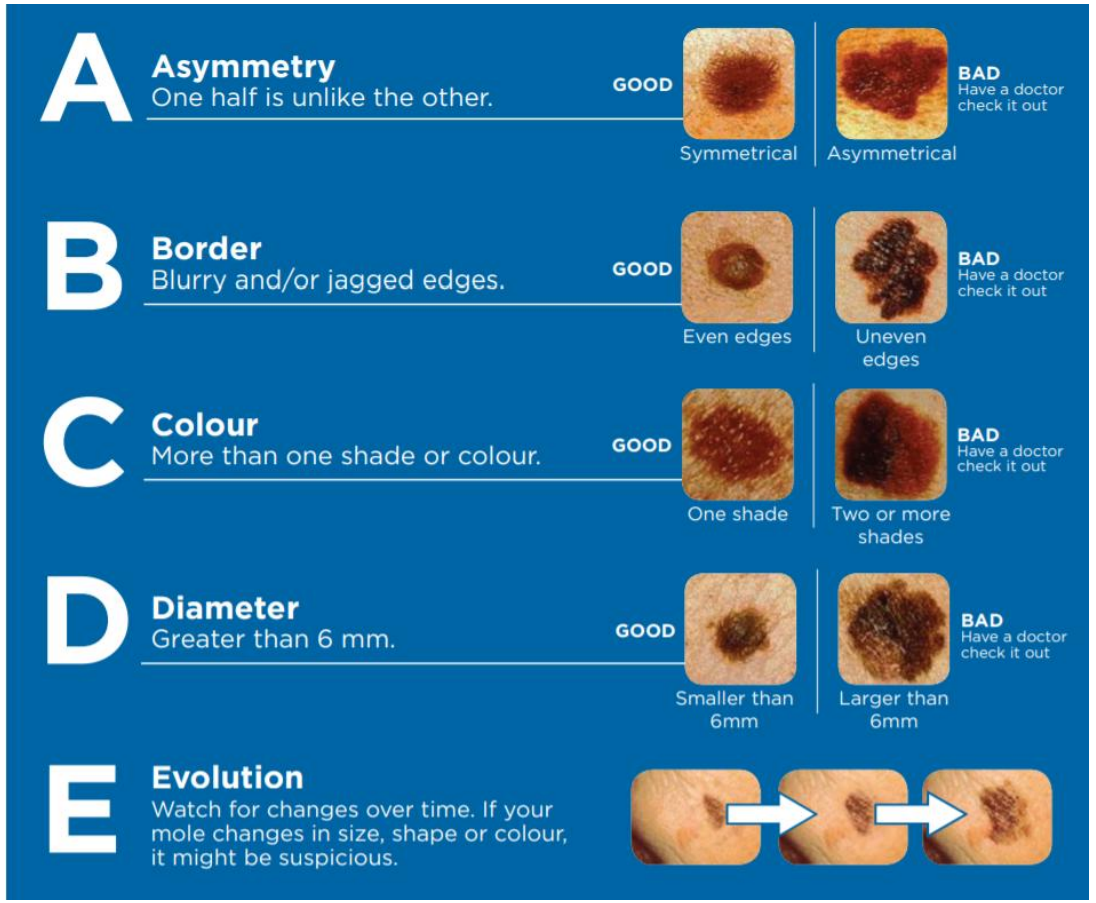
Digital dermatoscope with camera. Source: [2]

[1] https://en.wikipedia.org/wiki/Dermatoscopy

[2] https://www.pressebox.com/pressrelease/fotofinder-systems-gmbh/Lightweight-USB-camera-for-dermoscopy-and-trichoscopy/boxid/851020

Relevant Background

- Skin: largest organ in human body
- Pigmented cells/ melanocytes growing in clusters = mole (10-40/person avg)
- Individuals with small or several large moles: high risk of melanoma (deadly skin cancer)
- American Cancer Society, 2023: 97,610 new melanomas and 7,990 deaths
- Moles persisting into old age: high risk
- ABCDE method for self-exams and diagnosis of melanoma
- Most literature classifies types of moles (malignant/benign) or other skin conditions
- We detect the presence of any kind of mole and if found direct them to dermatologists, early diagnosis and risk mitigation

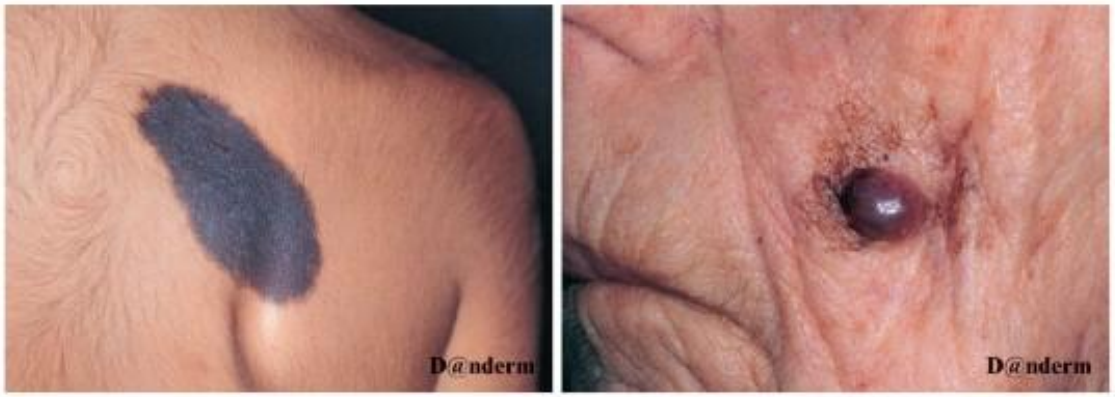


ABCDE method for melanoma detection [3]

[3] https://www.canadianskincancerfoundation.com/early-detection/the-abcdes-of-early-detection/

Thesis Objective

Developing an automated classification method to identify patient uploaded images with moles to assist dermatologists

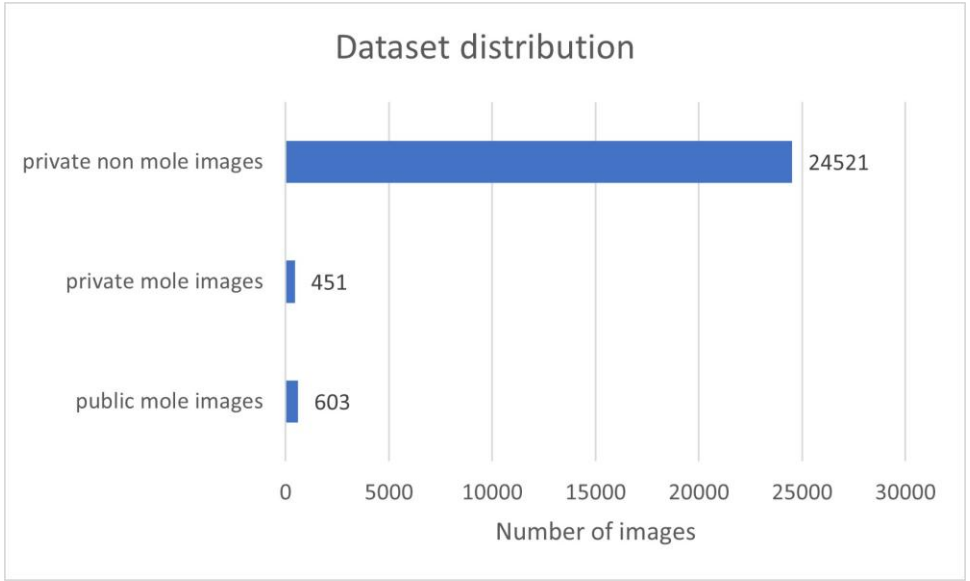


(a) Sample image with a nevus (b) Sample image of melanoma

Sample images of nevus and melanoma from Danderm dataset

Datasets and methods

- Private dataset from OROHealth: 40 different classes of skin diseases
- Labeled by 2 Canadian board-certified dermatologists
- Nevus, melanoma: mole, rest: non mole
- Huge dataset imbalance
- Use of public Danderm* dataset
- Moles: Nevus, melanoma and tumours from Danderm added to the previous one
- Minority class: 1.8% to 4.12%
- At max 600 images from each class while training

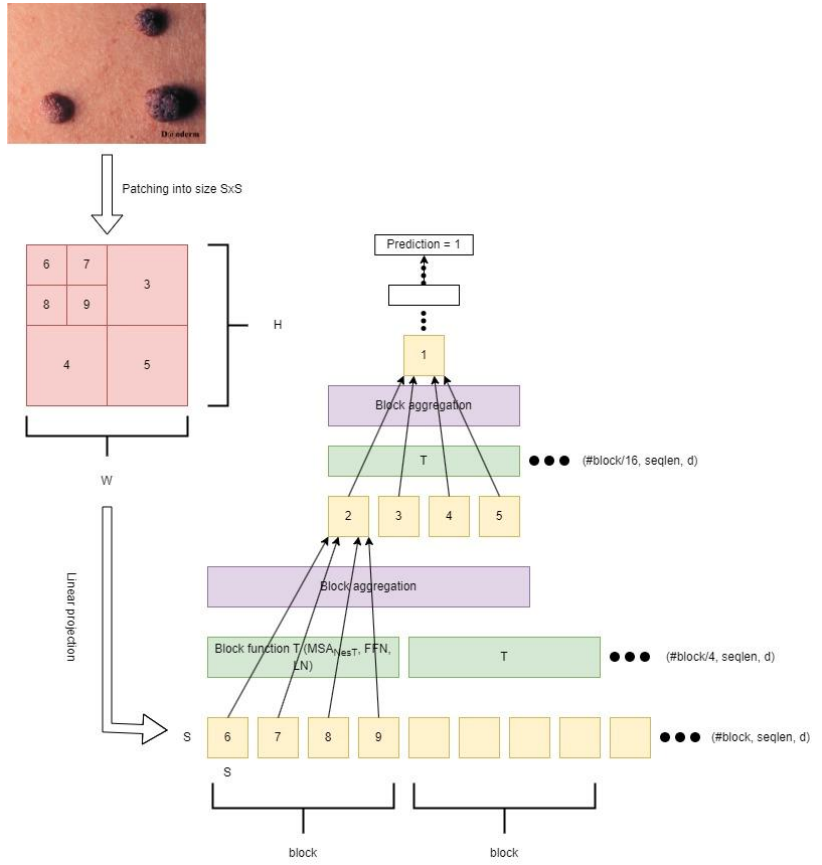


Bar plot representing dataset imbalance

* https://danderm-pdv.is.kkh.dk/

Datasets and methods (Contd.)

- NesT (Nested Hierarchical Transformer) trained on ImageNet dataset is used for classification
- Block aggregation to share info from spatially connected blocks
- Hierarchically nested local transformers
- Blocks at same hierarchy share same parameters =>applied parallelly => simplified arch + data efficiency
- Each block processes local info
- Global info from neighbouring blocks shared only while aggregation in image plane
- Hence arch utilizes both local and global info at every hierarchy level change
- Results compared against ViT (base, 32x32), BiT, Inception-v4



Hierarchical cross block aggregation in NesT

Experimental setup

- 5 models: NesT, ViT (base, 32x32 trained on ImageNet21k), BiT and Inception-v4 from timm (Hugging Face) library
- NVIDIA Tesla P4 GPUs from VM on GCP
- Data augmentation: image transforms such as random cropping, resizing, horizontal-vertical flip, random rotation, mix-up data augmentation
- Gradient accumulation instead of batch accumulation
- Cross entropy, Focal and Asymmetric single label loss
- Optimizers: SGD, Adam
- LR schedulers: One-cycle, cosine-annealing
- 22 different experimental permutations by varying optimizer, grad accumulation batch, loss function, learning rate

Gradient accumulation batch	Optimiser	Loss function	Learning rate
16	SGD	ASL	0.01
32	SGD	ASL	0.01
64	SGD	ASL	0.01
32	SGD	ASL	0.001
32	SGD	ASL	0.1
32	adam	ASL	0.01
32	adam	ASL	0.001
32	adam	ASL	0.1
32	adam	cross entropy	0.01
32	adam	cross entropy	0.001
32	adam	cross entropy	0.1
32	SGD	cross entropy	0.01
32	SGD	cross entropy	0.001
32	SGD	cross entropy	0.1
16	SGD	focal	0.01
32	SGD	focal	0.01
64	SGD	focal	0.01
32	SGD	focal	0.001
32	SGD	focal	0.1
32	adam	focal	0.01
32	adam	focal	0.001
32	adam	focal	0.1

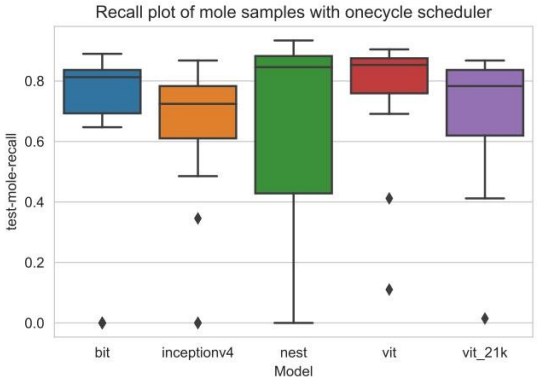
Set of 22 experiments performed on each of the 5 models

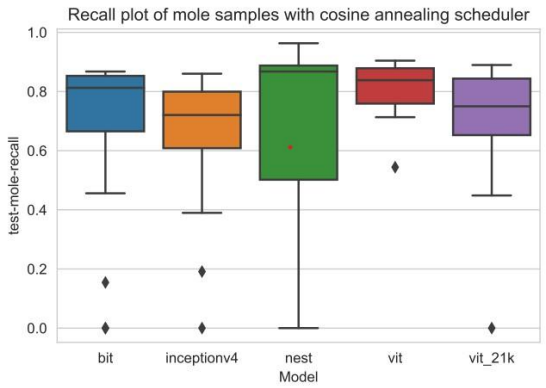
NesT	ViT	ViT (ImageNet21k pretrained)	BiT	InceptionV4
×1	×0.01	×0.01	×0.3	×0.1

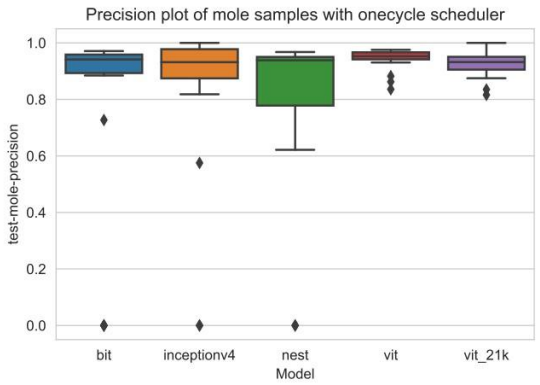
Multiplication indices of the max LR of different models

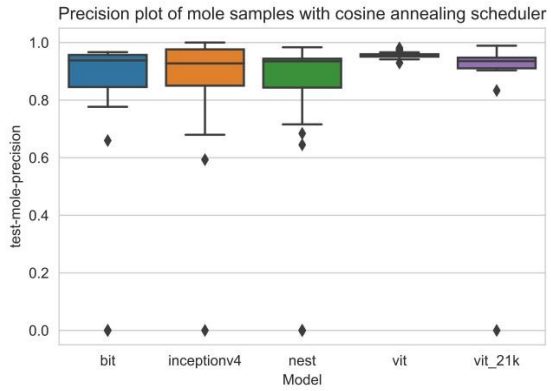
Results and analysis

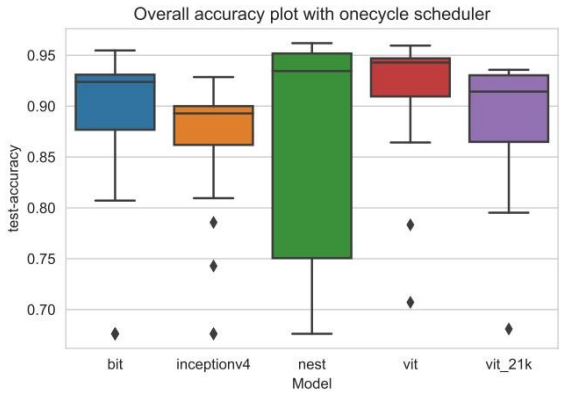
- Combined inference: NesT is the best performing model w.r.t. recall and overall accuracy
- Recall: mole detection, overlooked false negatives are more costly than false alarms
- Despite lower precision NesT: tradeoff
- Best hyper-param combination:
  - Batch size = 32
  - Optimizer = SGD
  - Loss = ASL
  - Max LR = 0.001

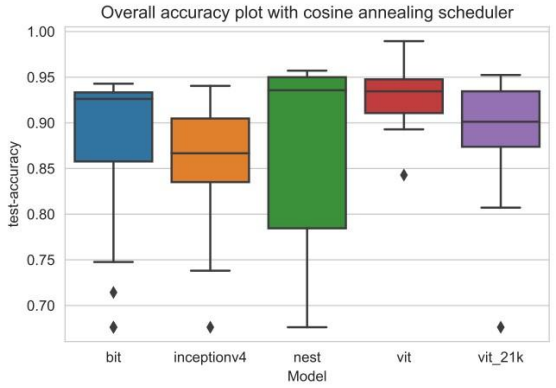




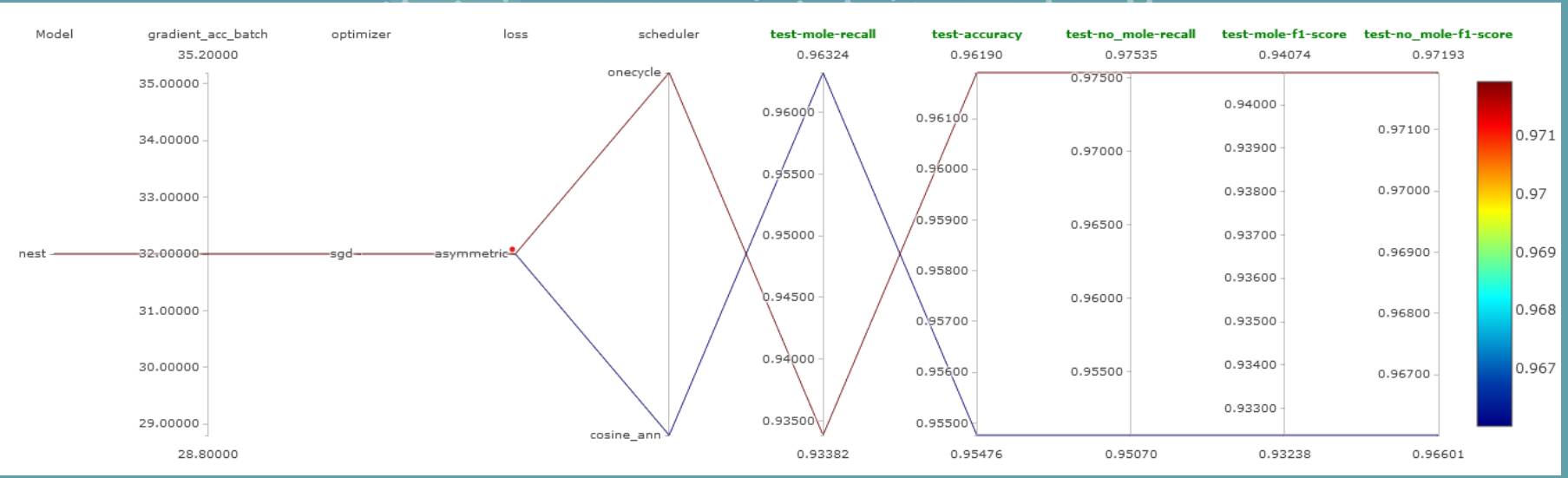








Results and analysis (Contd.)



Parallel plot depicting one-cycle scheduler vs cosine annealing scheduler

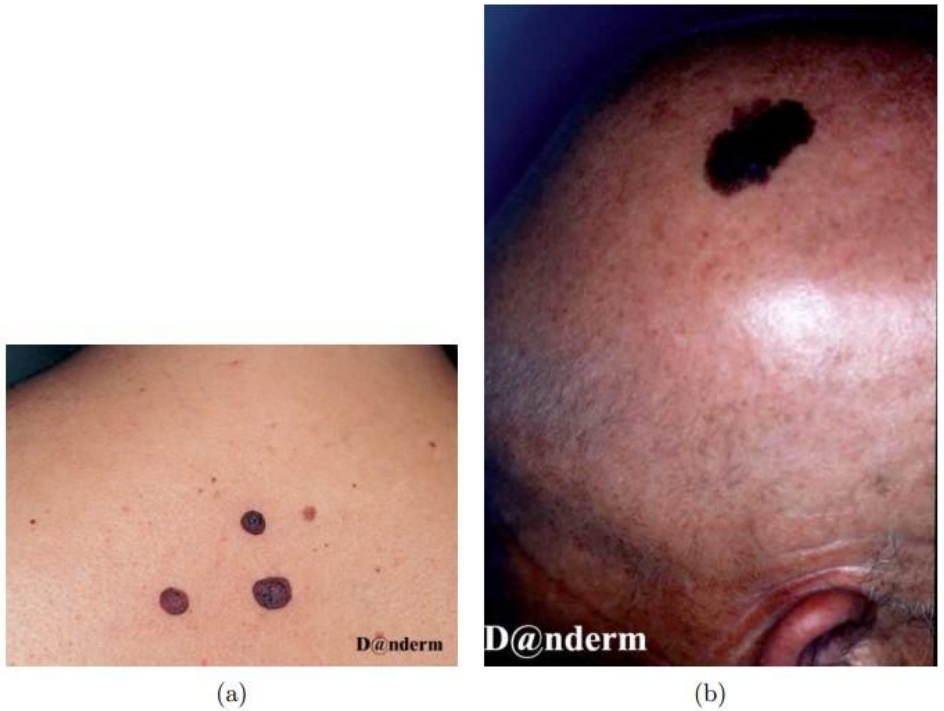
Model	Gradient accumulation batch	Optimiser	Loss function	Learning rate	Mole recall
BiT	32	sgd	asymmetric single label	0.03	0.8897
InceptionV4	32	sgd	asymmetric single label	0.01	0.8676
NesT	32	sgd	asymmetric single label	0.001	0.9338
ViT	32	sgd	cross entropy	0.0001	0.9044
ViT_21k	32	sgd	cross entropy	0.0001	0.8676

Mole recall values of best performing experiments of each model

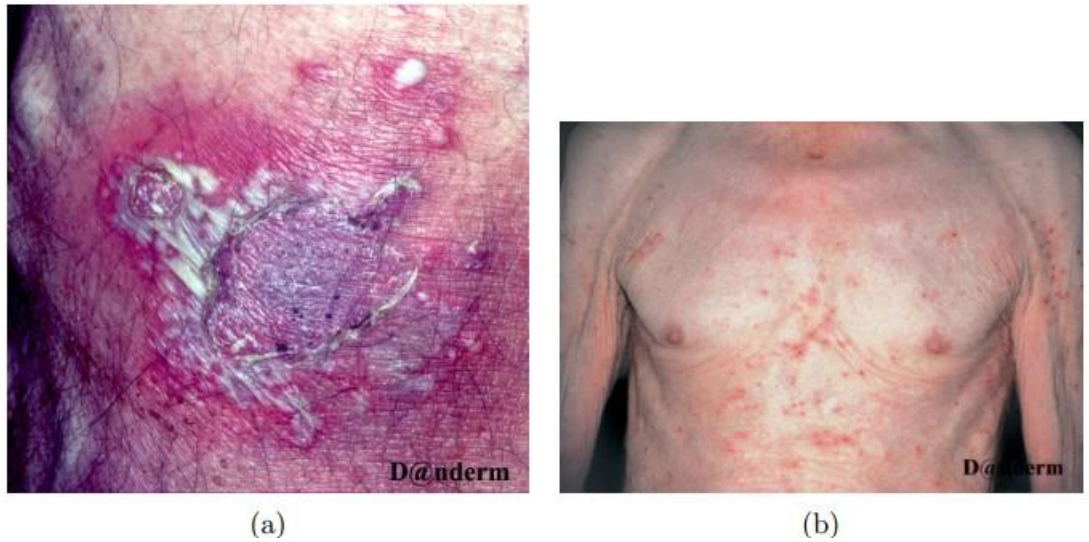
		Predicted label		
		Mole	Non-mole	Total
True label	Mole	127	9	136
	Non-mole	7	277	284
	Total	145	275	420

Confusion matrix of the best performing NesT model

Examples



Correctly classified mole samples (a) nevus image, (b) melanoma
Probability values: 94.9% and 50.5% respectively.



Correctly classified non mole samples (a) pustular psoriasis, (b) parasite infection
Probability values: 89.8% and 95.9% respectively.

* Lladó, X., Ganiler, O., Oliver, A. et al. Automated detection of multiple sclerosis lesions in serial brain MRI. Neuroradiology 54, 787–807 (2012). https://doi.org/10.1007/s00234-011-0992-6

Human Skin Detection

- Segmentation of human skin of different complexions
- Explicit thresholding techniques: not computationally expensive, easy to implement
- RGB, YCrCb, HSV intensity values are thresholded to fall in human skin color space
- CMYK color space [5] is defined as
  K = min(255-R, 255-G, 255-B)
  C = (255-R-K)/(255-K)
  M = (255-G-K)/(255-K)
  Y = (255-B-K)/(255-K)

$$RC = RULE\_RGB \cap RULE\_CMYK$$
$$RH = RULE\_RGB \cap RULE\_HSV$$
$$RHC = RULE\_RGB \cap RULE\_HSV \cap RULE\_CMYK$$
$$RHC\_Vote = min\_2vote(RULE\_RGB, RULE\_HSV, RULE\_CMYK)$$

- A hybrid color scheme is adopted to classify pixels according to their intensity values.
- The min_2vote() function returns True if at least of two of its arguments returns a True value

RGBA check:

$$rule\_1 = (R > 95)\&(G > 40)\&(B > 20)\&(R > G)\&(R > B)\&(abs(R - G) > 15)\&(A > 15)$$
$$rule\_2 = (R > 220)\&(G > 210)\&(B > 170)\&(G > B)\&(R > B)\&(abs(R - G) <= 15)$$
$$RULE\_RGB = rule\_1 \cup rule\_2$$

HSV check:

$$rule\_3 = H < 25$$
$$rule\_4 = H > 230$$
$$RULE\_HSV = rule\_3 \cup rule\_4$$

YCrCb check:

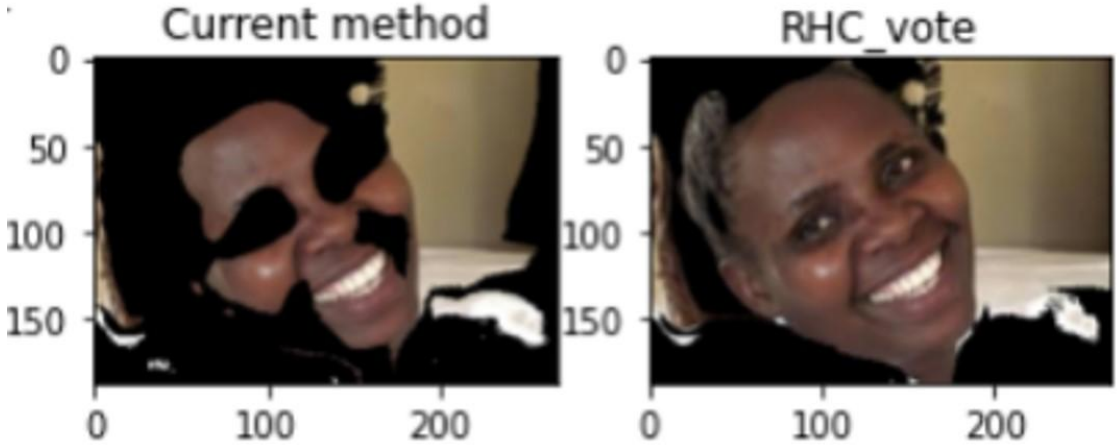
$$RULE\_YCRCB = (Cr > 135)\&(Cb > 85)\&(Y > 80)\&(Cr <= (1.5862 * Cb) + 20)\&(Cr >= (0.3448 * Cb) + 76.2069)\&(Cr >= (-4.5652 * Cb) + 234.5652)\&(Cr <= (-1.15 * Cb) + 301.75)\&(Cr <= (-2.2857 * Cb) + 432.85)$$

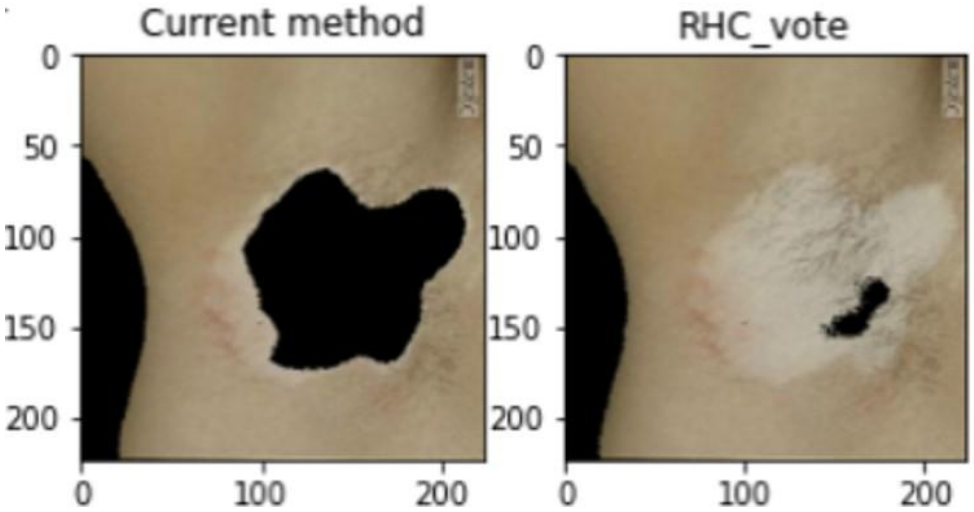
CMYK check:

$$rule\_5 = K < 205$$
$$rule\_6 = 0 <= C <= 0.05$$
$$rule\_7 = 0.0909 < Y < 0.945$$
$$rule\_8 = 0.1 <= Y/M < 4.67$$
$$RULE\_CMYK = rule\_5 \cap rule\_6 \cap rule\_7 \cap rule\_8$$

[5] A. Kumar and P. Shanmugavadivu, "Skin detection using hybrid colour space of rgb-h-cmyk," in *Progress in Advanced Computing and Intelligent Engineering*, B. Pati, C. R. Panigrahi, S. Misra, A. K. Pujari, and S. Bakshi, Eds., Singapore: Springer Singapore, 2019, pp. 75–83, isbn: 978-981-13-1708-8.

Examples





- Current method refers to the thresholds applied without CMYK space
- RHC_vote is better at detecting skin of darker shades. However, a lot of the background (False positive) spills into the foreground
- Current method [4] or the method without CMYK space is better at minimizing false positives but masks up a lot of the skin area in darker shades
- However ViT and NesT can detect skin features even without masking filters due to their superior learning capabilities

[4] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human skin detection using rgb, hsv and ycbcr color models," in *Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*, Atlantis Press, 2016/12, pp. 324–332, isbn: 978-94-6252-305-0. doi: 10. 2991/iccasp-16.2017.51. [Online]. Available: https://doi.org/10.2991/iccasp-16.2017.51.

Conclusion

- Promising first step in incorporating deep learning to successfully assess patients and triage them based on the nature of treatment required
- Our system achieved recall value = 93.4% on test images with moles and macro average recall value = 95.46% on the entire test dataset
- Scope of improving data: Most skin samples are light skinned, number of image samples with moles limited
- Future works should employ more diverse skin types and moles in the data

Publication reference

The results of this study were published in the journal ‘Informatics in Medicine Unlocked (2023)’ – Elsevier

Das D, Ergin E, Morel B, Noga M, Emery D, Punithakumar K. AI-assisted mole detection for online dermatology triage in telemedicine settings. Informatics in Medicine Unlocked. 2023 Jul 27:101311.

THANK YOU

QUESTIONS