Deep convolutional neural networks to identify skin cancer

Christoph Sadée, Pritam Mukherjee, Roberto Novoa, Olivier Gevaert
Biomedical Informatics Research, School of Medicine, Stanford University

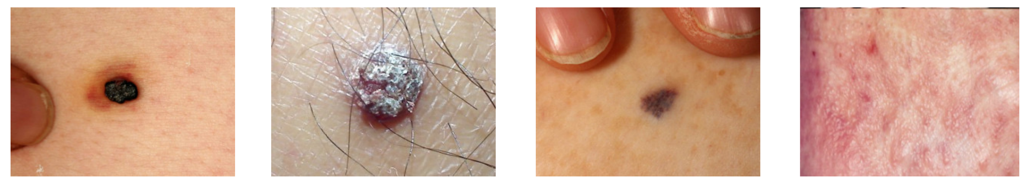
Abstract

Melanoma represents fewer than 5% of all skin cancers, but accounts for 75% of skin cancer related deaths [1]. Early detection profoundly increases survival rate. Diagnosis begins with visual examination that reveals a suspicious morphology, which then triggers biopsy and further work-up. Such exams are often prompted when patients notice unusual lesions or other healthcare providers spot worrisome features. Automated classification of skin lesions based on a single photograph poses numerous technical challenges [2], but would enable better access to initial screening of skin lesions. Esteva et al. [3] demonstrated that these challenges can be overcome, using a deep convolutional neural network (CNN) to differentiate keratinocyte carcinomas (malignant) from seborrheic keratosis (benign) and melanoma (malignant) from nevi (benign) and many further – representing first the most common skin cancer and second the most deadly. Their model performed on par with classifications provided by clinical experts and represents exciting initial results. However these initial findings were based on curated images that are part of online dermatology repositories. Here we extend this work to real world clinical images from the Stanford Dermatology department with 45k images. Extract diagnoses from reports, match them to our model classes and filter images to finetune the Esteva model. We observe a 67% accuracy for malignant vs benign vs inflammatory classification.

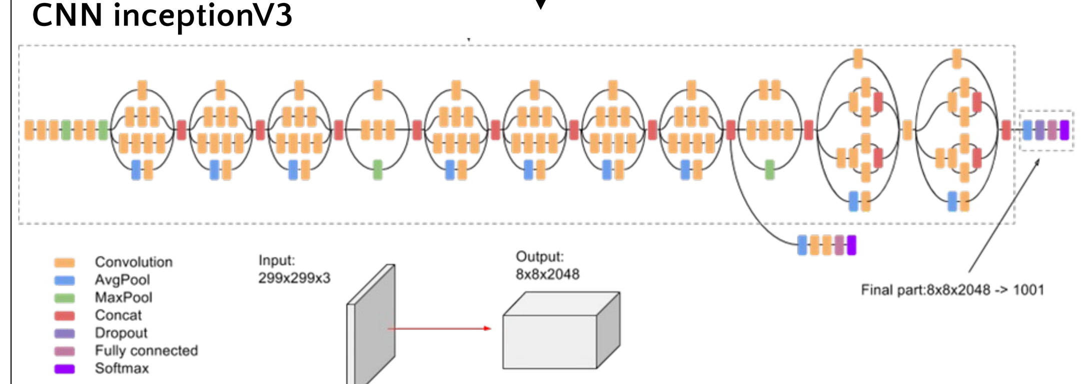
Introduction

Esteva et al 2017 have shown that CNNs can identify suspicious skin lesions by training the inceptionV3 CNN on scraped dermatology images from online repositories. Proposed applications range from early detection of malignant lesions to supporting clinical decision making.

Web scraped data



CNN inceptionV3



Output Classes

838 labels	9 labels	3 labels
0 1 2 . . . 837	0: benign dermal tumors cysts sinuses 1: benign dermal tumors cysts sinuses 2: epidermal tumors hamartomas milia and growths benign	0: benign
	3: epidermal tumors pre-malignant and malignant 4: genodermatoses and supernumerary growths 5: inflammatory	1: malignant
	6: malignant dermal tumor 7: pigmented lesions benign 8: pigmented lesions malignant	2: non-neoplastic

Esteva et al model 2017

Scraped images are procured photos of high quality. To deploy the model within a real world setting, we have to fine tune on clinical images.

Data

Available Stanford Clinical data

19'126 Reports

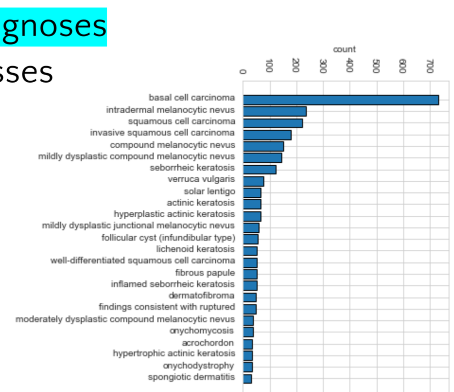
```
NARRATIVE: Accession No: XXXXXX SPECIMEN SUBMITTED: LEFT SOLE OF FOOT SUBMITTED ICD9 CODE: 238.2-238.9 DIAGNOSIS: SKIN, LEFT SOLE OF FOOT, BIOPSY -- ACRAL COMPOUND MELANOCYTIC NEVUS WITH SPECIAL SITE ATYPIA, NARROWLY EXCISED ON PLANE OF SECTION EXAMINED (SEE COMMENT) XXXXXXXX XXXXX COMMENT: This case has been reviewed in Dermatopathology consensus conference.
```

44'982 Images

Diagnosis extraction

- Key words in each report, identify diagnoses
- Diagnoses are matched to model classes
- ~60% of diagnoses are extracted

```
NARRATIVE: Accession No: XXXXXX SPECIMEN SUBMITTED: LEFT SOLE OF FOOT SUBMITTED ICD9 CODE: 238.2-238.9 DIAGNOSIS: SKIN, LEFT SOLE OF FOOT, BIOPSY -- ACRAL COMPOUND MELANOCYTIC NEVUS WITH SPECIAL SITE ATYPIA, NARROWLY EXCISED ON PLANE OF SECTION EXAMINED (SEE COMMENT) XXXXXXXX XXXXX COMMENT: This case has been reviewed in Dermatopathology consensus conference.
```



Foreign object detection

- Foreign object influences disease detection and requires cropping
- InceptionV3 can recognize rulers, band-aids and syringes
- ~20 % of data contains rulers

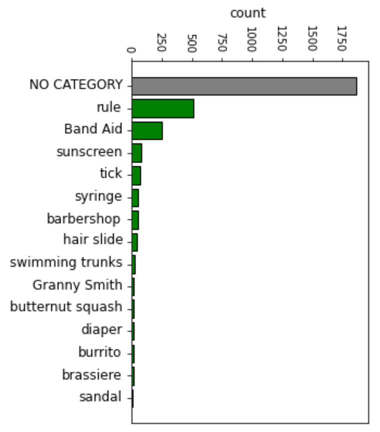
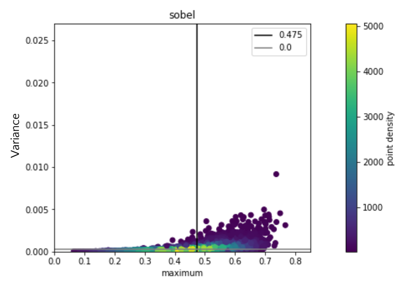


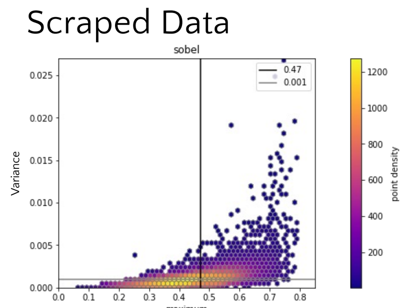
Image quality – Blur Detection

- Image sharpness influence disease detection
- Sobel filter highlights edges in figures
- Large variance and maximum in Sobel image = sharp image
- Small variance and maximum in Sobel image = blurry image

Clinical Data

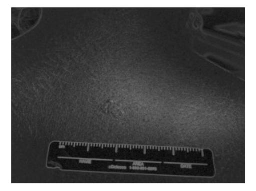


Scraped Data



Image

Sobel Image



Lesion Cropping

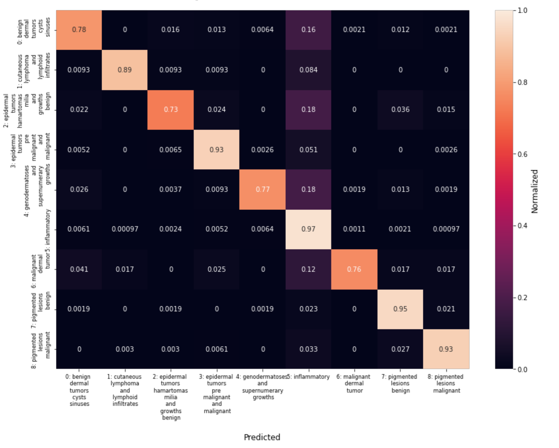
- Scraped data images are focused on lesion
- Clinical data often captures whole person images
- Identified rulers are aligned next to lesion
- Cropping image next to rulers to focus on lesions

Results

Pre-Clinical data performance

Esteva et al model was trained on >80k images. It classifies lesions into 9 model categories, with 93% accuracy. Inflammatory category has the most mis-classified lesions.

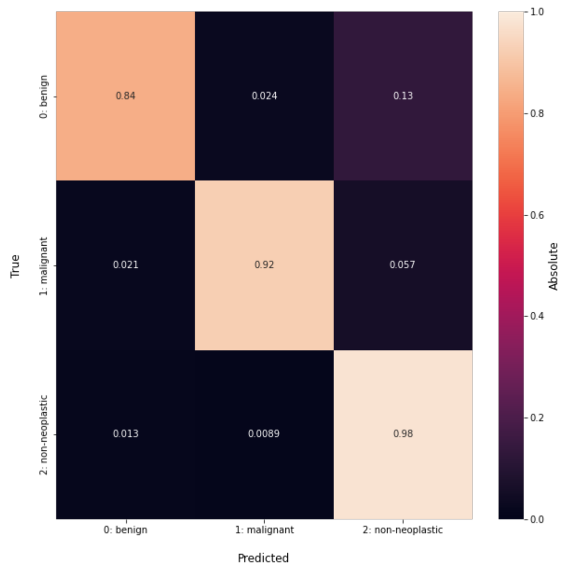
Scraped Test data



Esteva et al model classifies lesions into 3 model categories, benign, malignant and non-neoplastic with 95% accuracy

Labels	Accuracy
9	0.93
3	0.95

Scraped Test data



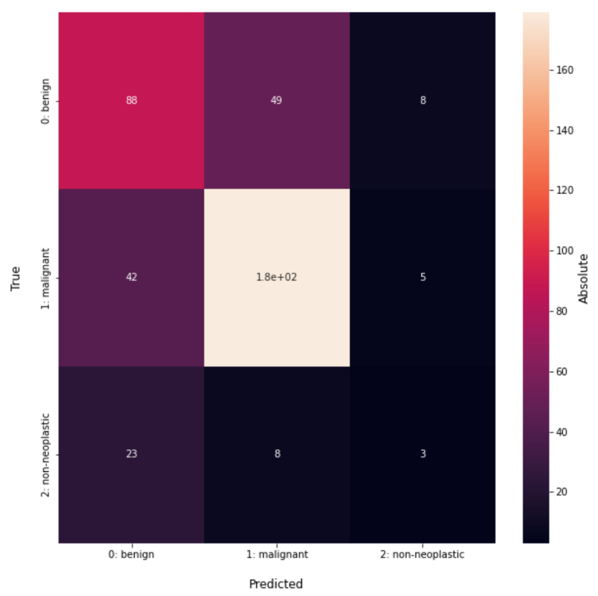
Clinical Data Performance

Model Fine tuning

- 1350 cropped clinical images
- 900 training images
- 450 test images
- Fine tuning all layers and lowering training rate to 10^{-4}

Finetuned model classifies lesions into 3 model categories, benign, malignant and non-neoplastic with 67% accuracy.

Clinical Test data



Labels	Accuracy
9	0.57
3	0.67

Next Steps & Conclusion

Next steps

- Further data extraction
- Cropping images if no ruler is present
- Removing blurry images
- Dermatologist annotate lesions with a purple pen, influencing classification.
  - Detect purple marks and filter
  - Remove or consider as confounder for classification
- Assess model performance on 9 level model classes

Conclusion

The inceptionV3 convolutional neural network can be used to accurately classify skin lesion in procured dermatology images with 95% accuracy when trained on >80k images. Translating the network to classify clinical dermatology images by finetuning, decreases accuracy to 67%. Although accuracy decreased, malignant and benign can still be identified with a fraction of the data available ~1000 images. We foresee further improvements with more data extraction and repeated finetuning.

References

References:
[1] Rigel, Darrell S. "Epidemiology of melanoma." Seminars in cutaneous medicine and surgery. Vol. 29. No. 4. WB Saunders, 2010.

[2] Chiem, Andy, Adel Al-Jumaily, and Rami N. Khushaba. "A novel hybrid system for skin lesion detection." 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information. IEEE, 2007.

[3] Esteva et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017.

General literature

Yu, Gang, et al. "A digital twin-based decision analysis framework for operation and maintenance of tunnels." Tunnelling and Underground Space Technology 116 (2021): 104125.

De Fauw, Jeffrey, et al. "Clinically applicable deep learning for diagnosis and referral in retinal disease." Nature medicine24.9 (2018): 1342-1350.

Litjens, Geert, et al. "A survey on deep learning in medical image analysis." Medical image analysis 42 (2017): 60-88.

Javed, Rabia, et al. "A comparative study of features selection for skin lesion detection from dermoscopic images." Network Modeling Analysis in Health Informatics and Bioinformatics 9.1 (2020): 1-13.

Albahli, Saleh, et al. "Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour." IEEE Access 8 (2020): 198403-198414.