Machine Learning for Automated Diagnosis of Skin Lesions

Student: Fábio Santos Nº80317
Supervisors: Filipe Silva, Pétia Georgieva
Seminar, Aveiro University, 11/12/2019

Context and Objectives

- Skin cancer is the **most common type of cancer** (1 in 5 persons)
- When detected on late stages survival rates are 23%
- **When detected early survival rates are 99%!**
- Skin lesions can be difficult to classify even for dermatologists
- **Support decision tool for automated diagnosis of skin lesions are needed**
- **eHealth/mHealth app which integrates well into the clinical workflow**

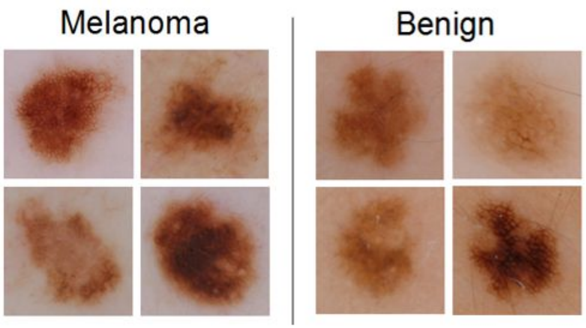


Fig 1 - Melanoma vs benign moles [1]

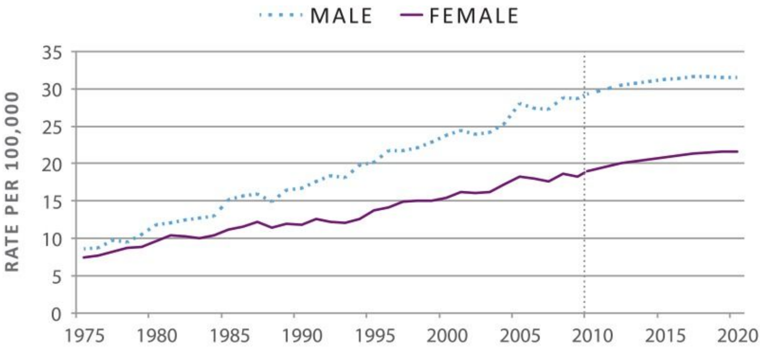


Fig 2 - Melanoma incidence rates [2]

Deep learning in medical imaging

- Advantages over other Ml methods:
  - No feature engineering
  - Transfer learning
- Rise of deep learning due to:
  - New methods to prevent overfitting
  - Rise of computing power along with GPU usage
  - High level modules to build neural nets
- Applied in contexts such as:
  - Diabetic retinopathy
  - Diagnosis of breast nodules
- Used as support tool in eHealth/mHealth apps
- Can beat human performance

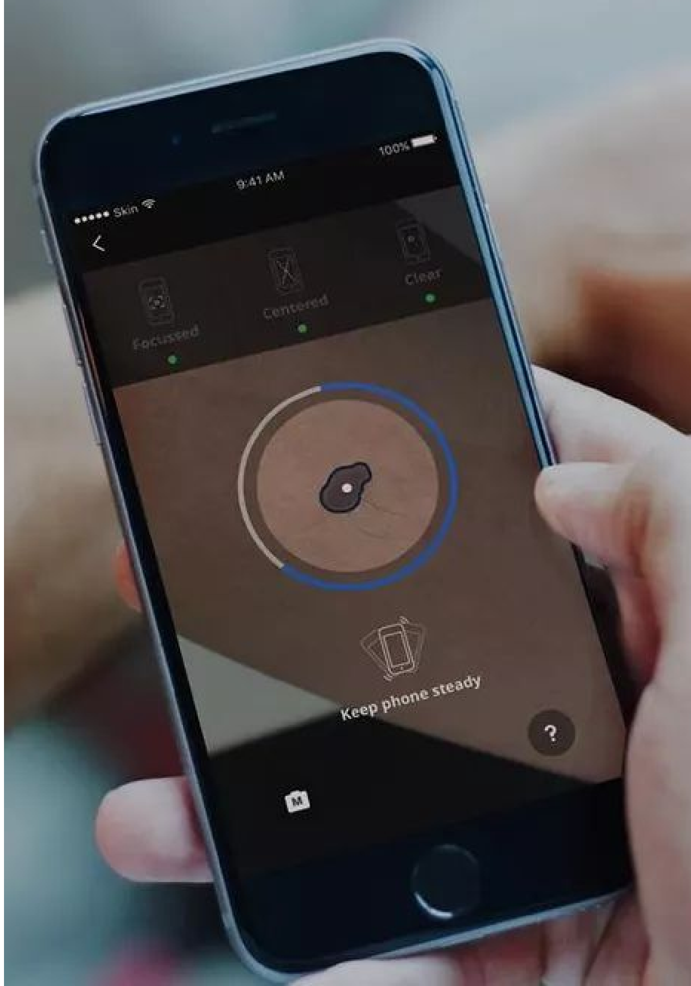


Fig 3 - SkinVision mHealth app [3]

Deep learning for skin lesion diagnosis (transfer learning)

- “Dermatologist-level classification of skin cancer with deep neural networks”, by Esteva et al. in 2017 [4]:
  - Most famous result in skin lesion classification using deep learning
  - Labelled skin images from multiple sources totalling **129450 samples!**
  - Used data augmentation
  - Transfer learning approach (InceptionV3 trained on ImageNet)
  - Evaluated against 21 board-certified dermatologists
  - **AUC: 0.96**
- “Man against machine”, by Haenssle et al, in 2018 [5]:
  - Similar approach to Esteva et al but with less data
  - Evaluated against **58 dermatologists**!
  - **AUC: 0.86**
- Most approaches use transfer learning

Fig 4 - Nature’s front page [4]

Deep learning for skin lesion diagnosis (from scratch)

- “New compact deep learning model for skin cancer recognition“, by Ly et al. in 2019 [6]:
  - Binary classification
  - Balances benign and malign cases
  - “PHDB” dataset comprised of multiple datasets (80,192 labeled images)
  - Multiple custom model architectures were created
  - Less parameters than most approaches
  - Deployed app for iOS and Android
  - 86% Accuracy

Actual	**Malignant**	**120**	16
	Benign	22	**114**
		Malignant	**Benign**
		Predicted	

Fig 5 - Confusion Matrix [6]

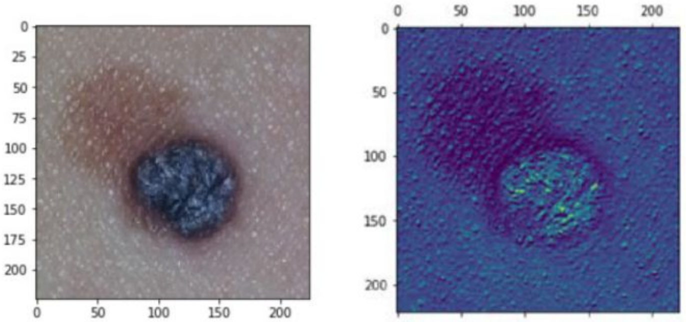


Fig 6 - Feature map example [6]

International Skin Image Collaboration

- Public dataset used as benchmark
- Places a yearly challenge
- ISIC 2018 part 3 top 3 winners:
  - Develop a classifier to distinguish between 7 classes
  - Top 3 submissions by Nozdryn-Plotnicki et al. (Metaoptima) [7] :
    - Data augmentation with preprocessing
    - Ensemble of multiple models trained on ImageNet (transfer learning)
    - AUC for best submission: ~98.3
- ISIC 2019 winner:
  - Develop a classifier to distinguish between 9 classes
  - Best submition by Gessert et al. [8]:
    - Transfer learning from EfficientNets model pre trained on ImageNet
    - Preprocessing and data augmentation techniques before training
    - AUC from optimal ensemble: 95.4 ± 0.5

Production grade applications

- Metaoptima's Dermengine web app:
  - Change detection tool
  - Image processing such as hair removal
  - Visual search tool
  - Uses ML to compare similar lesions

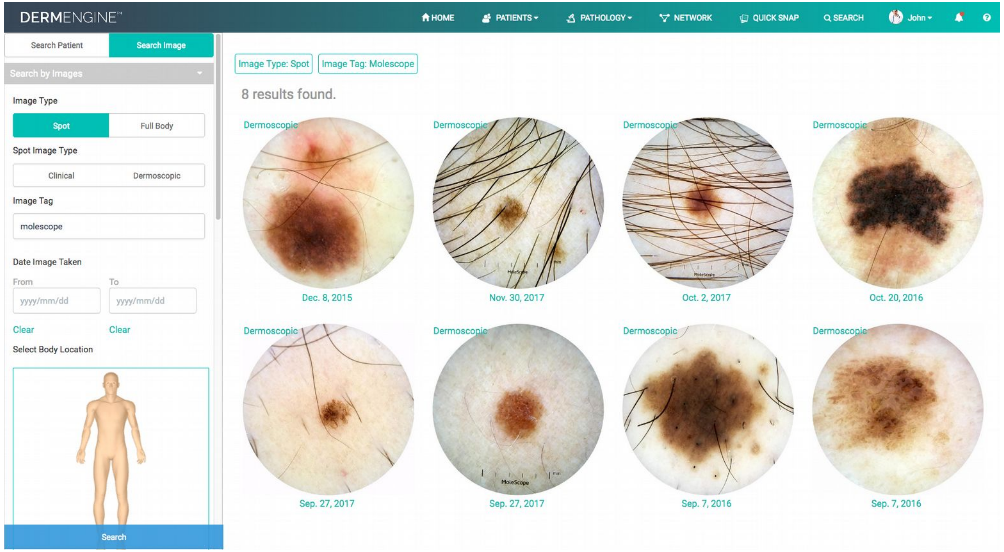


Fig 7 - Demengine web app [9]

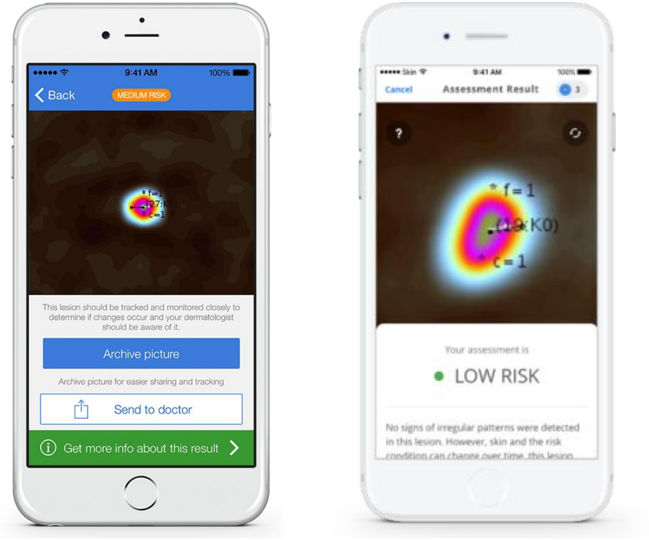


Fig 8 - SkinVision mobile app[3]

- Skin vision mobile app:
  - Patient grade mHealth app
  - ~30 second response time
  - Low/Medium/High risk response
  - 97% Sensitivity, 78% Specificity
  - Notifications to remind patients

Deep learning

- Implemented through artificial neural networks
- Require large amounts of data and computational power
- Multiple layers, each with multiple neurons
- Neurons acts like a function with parameters
- Parameters are learned in iterations based on a optimizer
- Different architectures for different purposes: RNN's, CNN's, LSTM

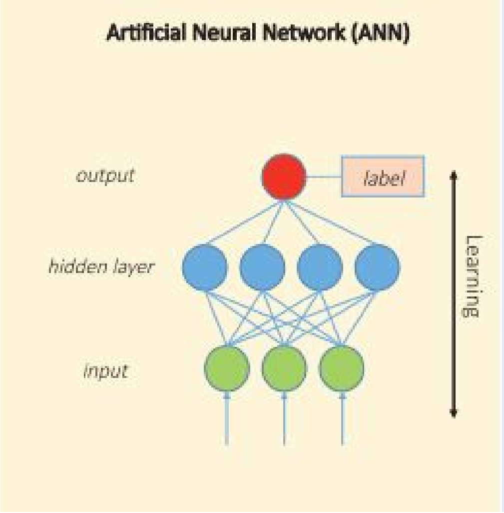


Fig 9 - Artificial neural network [10]

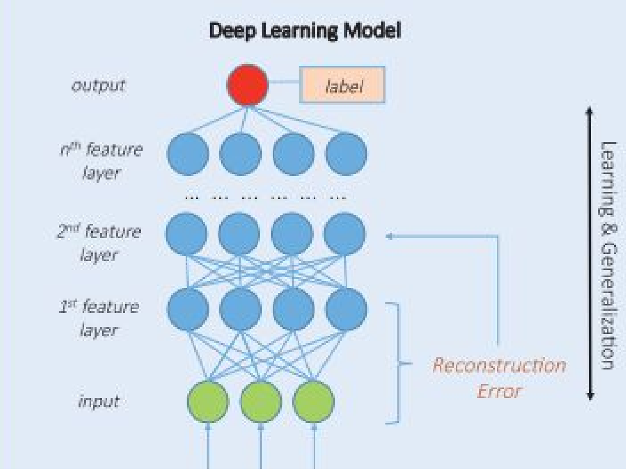


Fig 10 - Deep Neural Network [10]

Convolutional neural networks (CNN’s/DCNN’s)

- Inspired by the way that the visual cortex works
- Widely used in image recognition
- 3 main differences from ANN’s:
  - Local receptive fields
  - Shared weights (makes it translation invariant)
  - Pooling layers to remove noise and reduce parameters needed
- Can have multiple sets of convolutional layer + pooling layer

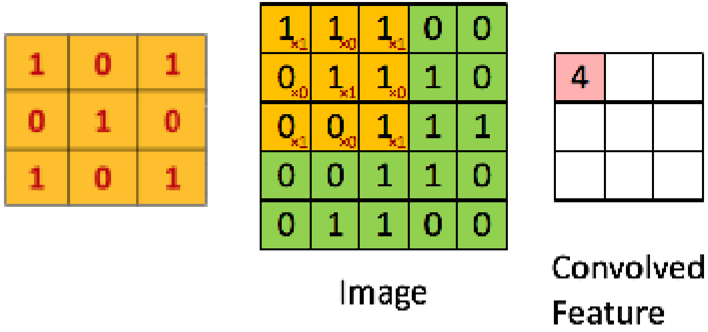


Fig 11 -Convolution Operation

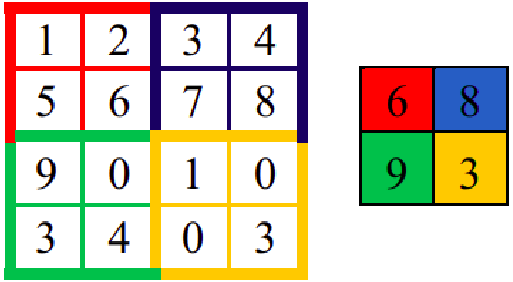


Fig 12 - Pooling operation

CNN’s Architectures

- ImageNet dataset:
    - Visual dataset designed for use in visual recognition software research.
    - Contains more than 14 Million hand labelled images with over 20000 classes
    - Imagenet large scale visual recognition challenge (ILSVRC)
- VGGNet:
    - 2014’s ILSVRC winner
    - Most common variation has 16 weight layers
    - Increased depth to build abstract features
- ResNet (Residual neural network):
    - 2015’s ILSVRC winner
    - Beats human performance on ImageNet
    - Less parameters than VGGNet (more lightweight)
    - 152 layers!

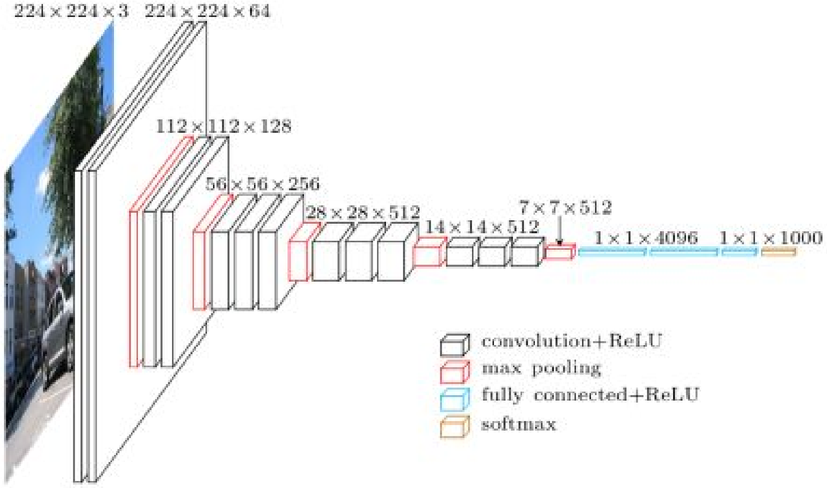


Fig 13- VGGNet Architecture [11]

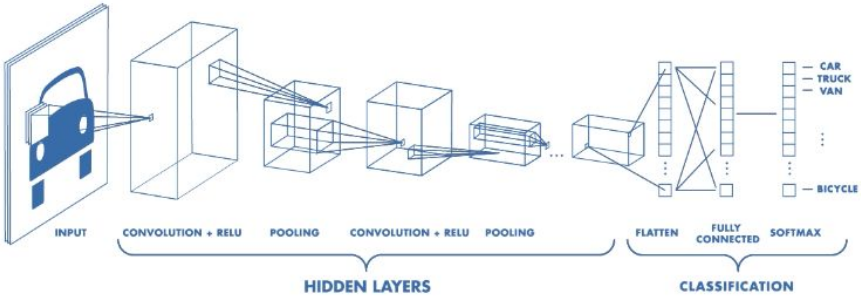
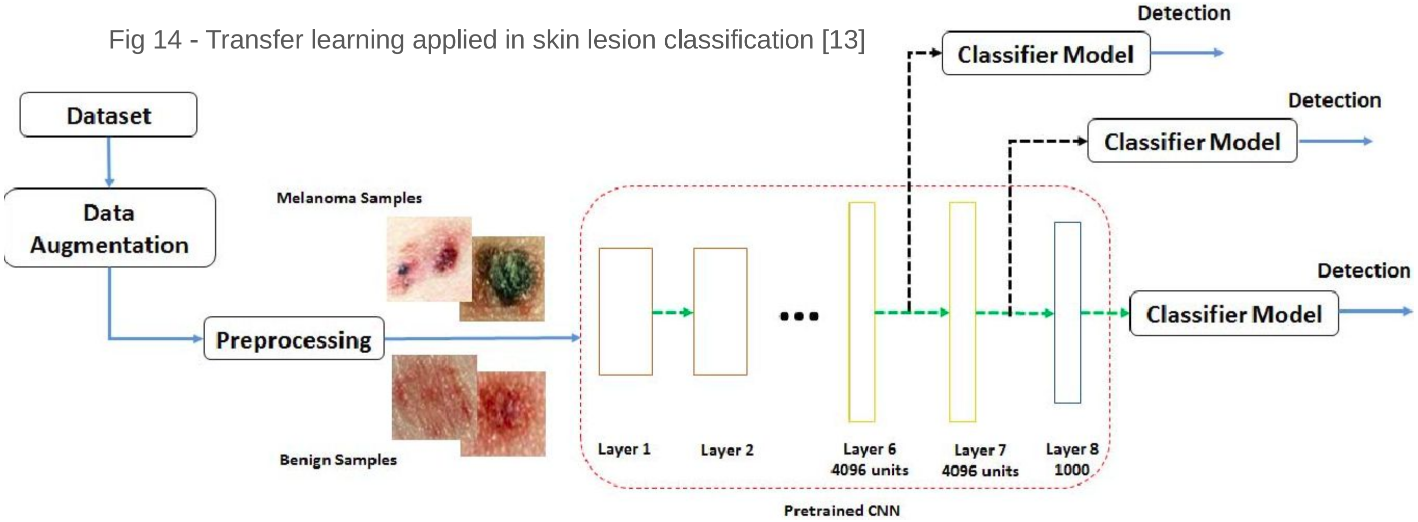


Fig 14- Conventional CNN [12]

Transfer Learning

- Deep networks requires large amounts of data and computational power
- Method of reusing a model or knowledge for another related task
- Very common technique for image recognition problems (most ISIC submissions use this approach)

Fig 14 - Transfer learning applied in skin lesion classification [13]



Training CNN’s: bias and variance trade off

- Oftentimes, models tend to either underfit or overfit while training
- Underfitting means that the model doesn’t perform well on the training data (high bias low variance)
- Overfitting means that the model performs well on the training data but generalizes poorly to new data (low bias high variance)
- Overfitting is very common on small datasets
- One must find the best trade off between bias and variance while training
- Data augmentation:
    - Generative models
    - General transformations

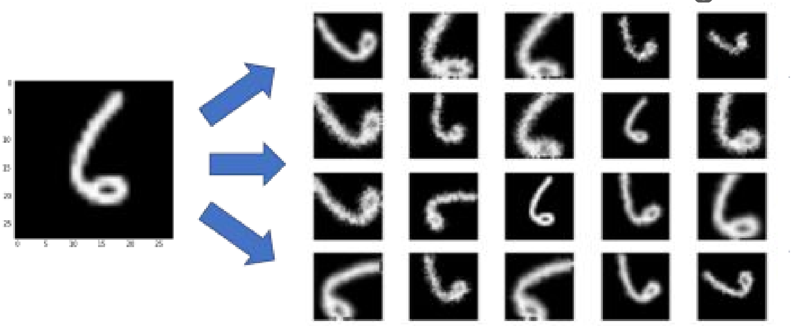


Fig 15 - MNIST data augmentation [14]

Proposed work and next steps

- **Develop a decision support tool for dermatologists:**
    - **Multiclass classifier to ISIC 2020**
    - **Study effects of data augmentation**
    - **Hyperparameter and model optimization**
- **Develop a responsive eHealth app for patients and dermatologists that integrates:**
    - **Change detection tool**
    - **Decision support tool**

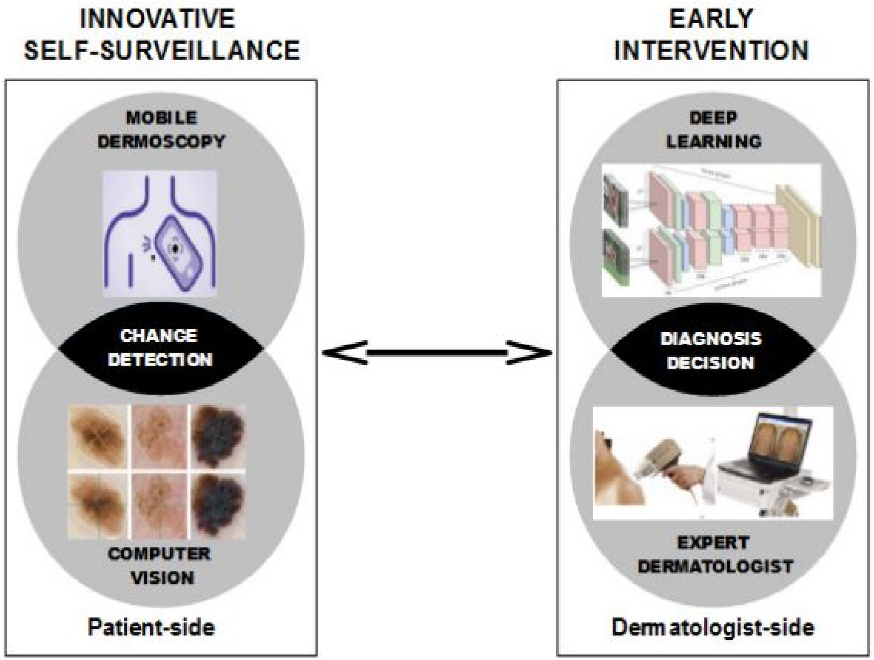


Fig 16 - Dissertation proposed work scheme

Conclusions

- There's a need for a decision support system to help dermatologists in the diagnosis of skin lesions
- CNNs show promising results in automated skin lesion diagnosis
- Techniques such as transfer learning are widely used in skin lesion diagnosis due to lack of data
- Data augmentation can greatly help with the overfitting problem
- Points to consider:
    - Deploying deep learning models into a web app requires orchestration
    - Influence of datasets, skin pigmentation and image quality

References (1)

[1] - Ali, A. A., & Al-Marzouqi, H. (2017, June 28). Melanoma detection using regular convolutional neural networks. 2017 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2017. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICECTA.2017.8252041

[2] - The Surgeon General's Call to Action to Prevent Skin Cancer. (2014). The Surgeon General's Call to Action to Prevent Skin Cancer. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25320835

[3] "SkinVision | Skin Cancer Melanoma Detection App | Check Your Skin." [Online]. Available: https://www.skinvision.com/. [Accessed: 20-Nov-2019]

[4] - Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

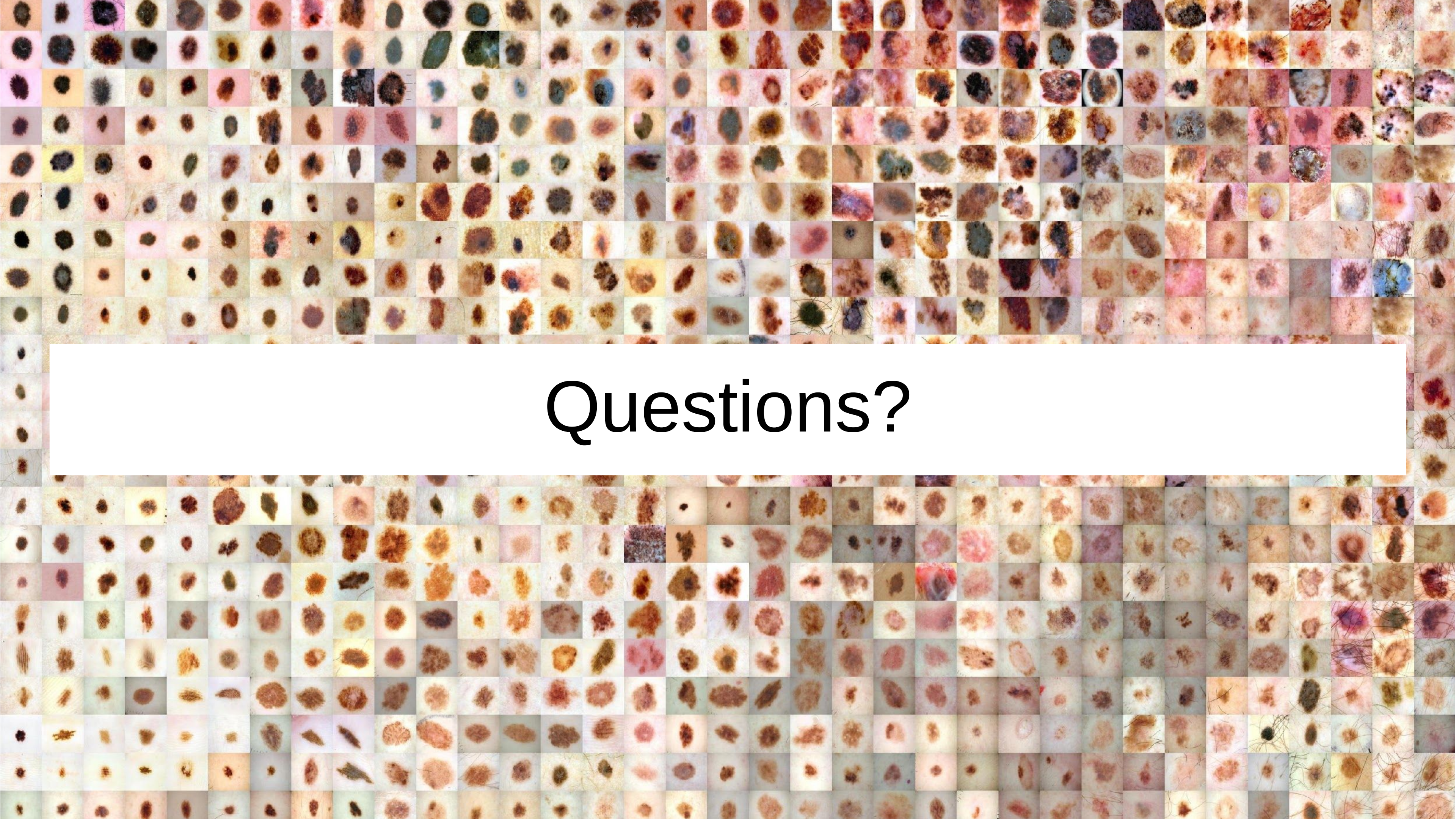
[5] - Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., … Zalaudek, I. (2018). Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology. Oxford University Press. https://doi.org/10.1093/annonc/mdy166

[6] - Ly, P., Bein, D., & Verma, A. (2019, August 15). New Compact Deep Learning Model for Skin Cancer Recognition. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/uemcon.2018.8796628

[7] - Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (n.d.). Skin Lesion Classification Using Loss Balancing and Ensembles of Multi-Resolution EfficientNets.

References (2)

[8] - Nozdryn-Plotnicki, A., Yap, J., & Yolland, W. (n.d.). *Ensembling Convolutional Neural Networks for Skin Cancer Classification*.
[9] - "DermEngine | Visual Search." [Online]. Available: https://www.dermengine.com/en-ca/visual-search. [Accessed: 21-Nov-2019]
[10] - R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Brief. Bioinform., vol. 19, no. 6, pp. 1236–1246, May 2017.
[11] - Canziani, A., Culurciello, E., & Paszke, A. (n.d.). AN ANALYSIS OF DEEP NEURAL NETWORK MODELS FOR PRACTICAL APPLICATIONS.
[12] - Basics of the Classic CNN - Towards Data Science. (n.d.). Retrieved December 11, 2019, from https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add
[13] - Pomponiu, V., Nejati, H., & Cheung, N. M. (2016). Deepmole: Deep neural networks for skin mole lesion classification. In Proceedings - International Conference on Image Processing, ICIP (Vol. 2016-August, pp. 2623–2627). IEEE Computer Society. https://doi.org/10.1109/ICIP.2016.7532834
[14] - Learning to Compose Domain-Specific Transformations for Data Augmentation. (n.d.). Retrieved December 8, 2019, from https://hazyresearch.github.io/snorkel/blog/tanda.html



Questions?