Image Analysis and Recognition using Deep Learning for Dermatological Skin Conditions

Charles A. Hulebak

Table of Contents

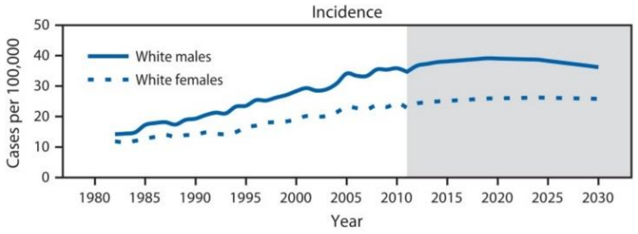
Project Background

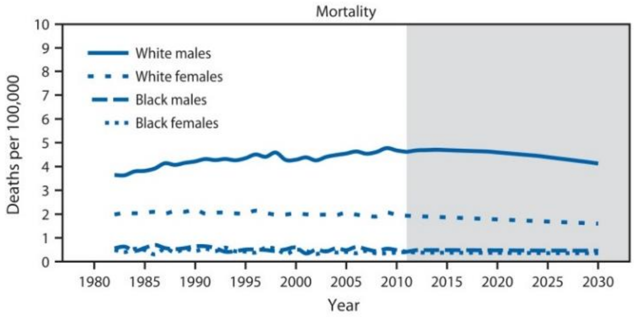
Skin cancer is the most common type of cancer worldwide and affects every race and gender.

Some of these unfortunate statistics in the United States include:

- Each year nearly **5 million people are treated** for all skin cancers combined, with more than **9,500 people diagnosed with skin cancer every day**, and more than **two people dying every hour**. [1,2]
- At least **one in five Americans** will develop skin cancer **by the age of 70**.[3]
- Basal cell carcinoma (BCC) is the most common form of skin cancer, with an estimated **4.3 million cases of BCC are diagnosed in the U.S. each year**. [4]
- The estimated **annual cost of treatment is estimated at $4.8 billion in the U.S.**, with approximately $3.3 billion dollars associated with costs attributable to melanoma.[5]
- **Globally, malignant skin melanoma has increased by 104.3%**, and non-melanoma skin cancer by 211.2% between 1990-2007, and 32.3% and 32.7% between 2007-2017.[6]

Observed and projected age-adjusted melanoma incidence and mortality rates, by sex and race — United States, 1982–2030[1]

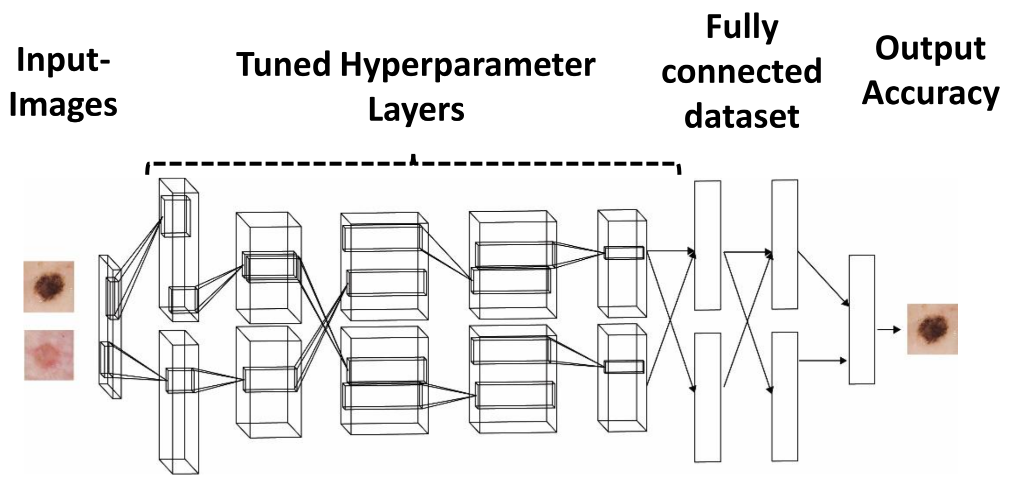




Charles A. Hulebak

Project Background

The use of technology, and specifically machine learning provides the ability to achieve these goals. Machine Learning, deep learning and innovations with convolutional neural networks (CNNs) provide the ability to analyze imagery data, classify, and provide results from inputted data. CNNs require large datasets in order to improve accuracy and provide sensible information.



The practical application of a machine learning program with a high success rate would provide the ability for increasing accurate results from early skin condition problems. Images of skin conditions could be tested with a reliable model, compared and have a probable diagnosis. A healthcare professional could take analyze the data, take this information into account, and treat conditions accordingly.

This project has the ability to influence further development of an initiative to decrease the amount of time required for early analysis of skin conditions and guide healthcare professionals, while increasing the response time for more serious conditions, and reducing the time invested for treating less serious conditions.

Brief Overview - Deep Learning and Convolutionary Neural Networks

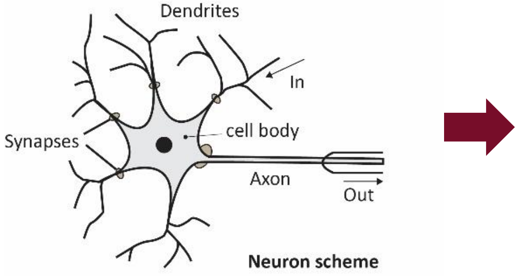
Deep learning is a subset of machine learning, which reflects the human learning process and derived from the biological processing of information. Convolutional neural networks (CNN) are a category of machine learning, in which unique features from a dataset are learned using filtered layers and a backpropagation algorithm to continually learn information about a dataset.

The process of convolution, in which CNNs obtain their name, is analogous to a filter (or stencil) passing over an image which is analyzed (by an activation map). Information is learned and passed to the subsequent layer for further processing, in which several layers may be introduced to more effectively learn information about the dataset.

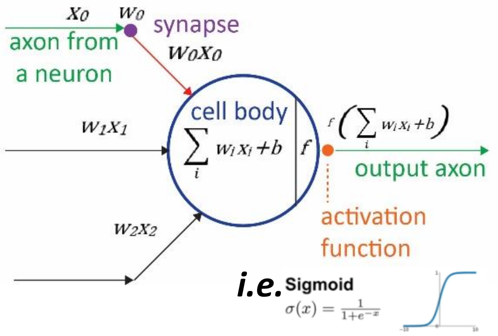
Inputs are introduced with a dataset, which is trained (from layers of programmed code) to make accurate predictions. CNNs are regularized versions of multilayer perceptrons, which are fully connected neuron networks.

Activation Function

Biological Process Diagram



Logical Process Diagram

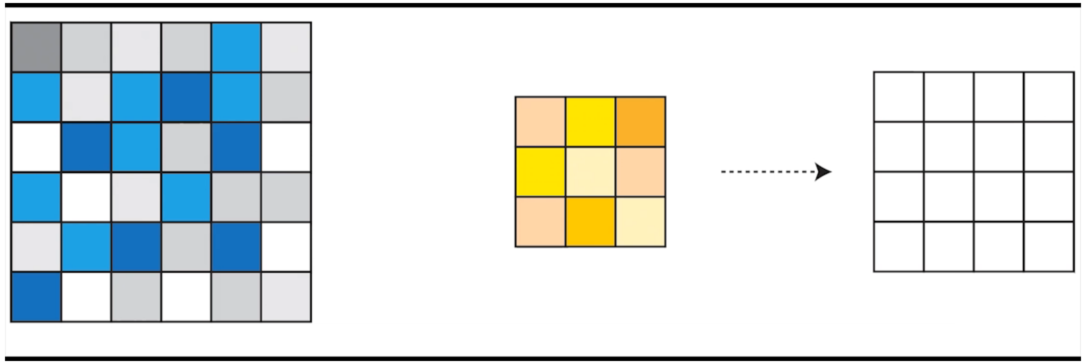


Images recreated from Stanford CS230 course

Convolutional Layer

The image is passed over all of the spatial locations as it is passed along to the subsequent layer(s) toward the output.

Input Layer **Hidden Layers** **Output Layers**



Brief Overview - Deep Learning and Convolutionary Neural Networks

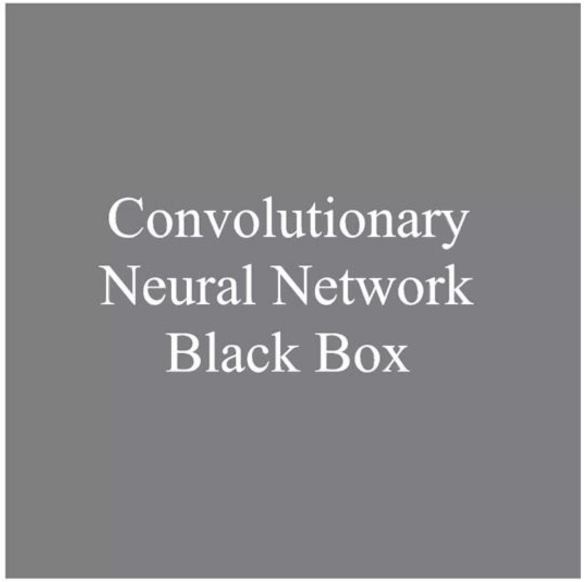
A series of layers are introduced to create a hierarchical structure, in which each layer (controlled by tuning hyperparameters) introduces various filters and increases a level of complexity to analyze the data.

Depending upon the data, and which layers are tested, accuracy may be determined from the test, training, and possible validation set (training and validation terminology may be interchanged). The measurement of accuracy reflects the machine's ability to learn from the data.

A series of epochs can test the data, while increasing the epochs allows the learning algorithm to run until errors have been minimized.

The accuracy of the model is tested and the hyperparameters are tuned for performance.

CNN 3D Walkthrough (animated, press play)



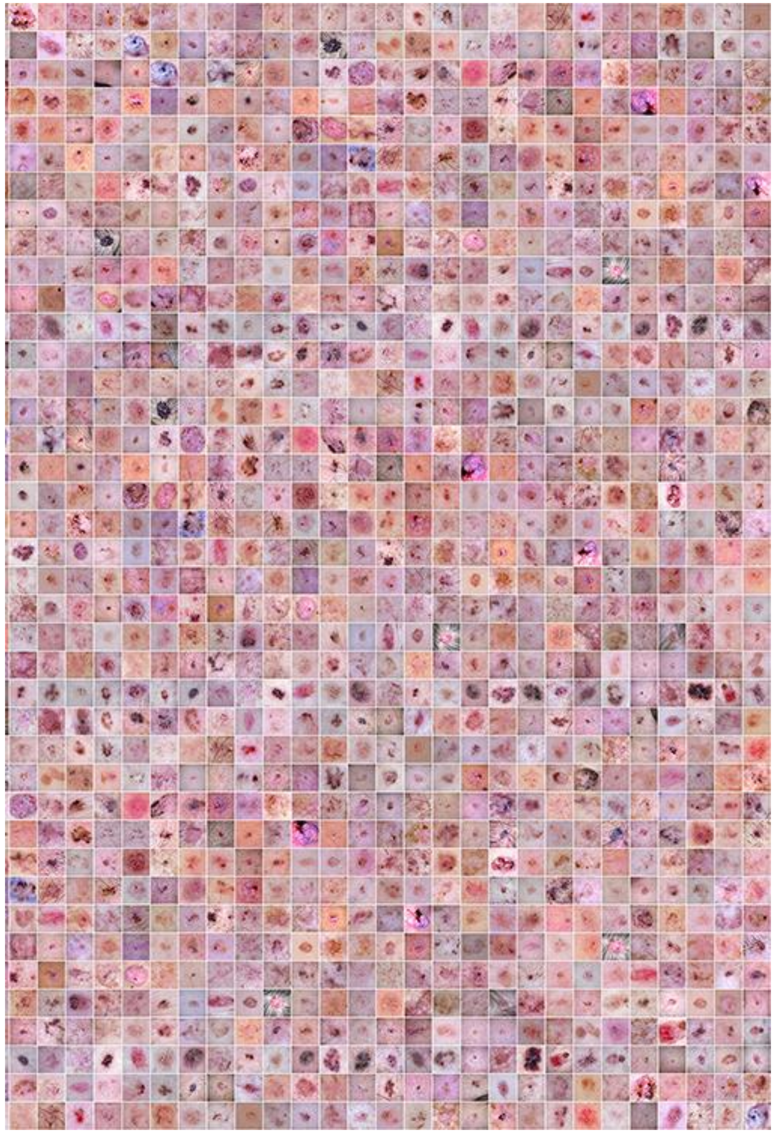
Dataset and Classification

Innovations and initiatives are taking place around the world to address a multitude of healthcare concerns using data science, machine learning, and technology.

This project includes performing machine learning algorithms for image recognition. Utilizing a subset of over 10,000 images from Harvard's Dataverse, and also part of the Kaggle "Skin Cancer MNIST: HAM10000 analysis" competition. [7]

The dataset includes 10,000 images which contain seven categories of skin conditions. Although the CNN is capable of classifying the accuracy between each of these types of conditions, this project will focus on two specific skin conditions, including the *melanocytic nevi, also called nevi (referred to as mole furthermore), and basal cell carcinoma, which is a form of cancerous cell*.

The physical appearance of these two types of skin conditions have many similarities, however the basal cell carcinoma requires immediate attention for clinical treatment.



Clinical Characterization

The clinical characterization of these two types of cells include:

Mole characteristics [8]

- Size: generally less than 5mm diameter, pigmented spots on the skin that usually appear between birth and childhood.
- Shape: round or oval
- Color: brown is the most common; may be tan, black, red, pink, blue, skin-toned, or colorless
- Characteristics: flat or slightly raised

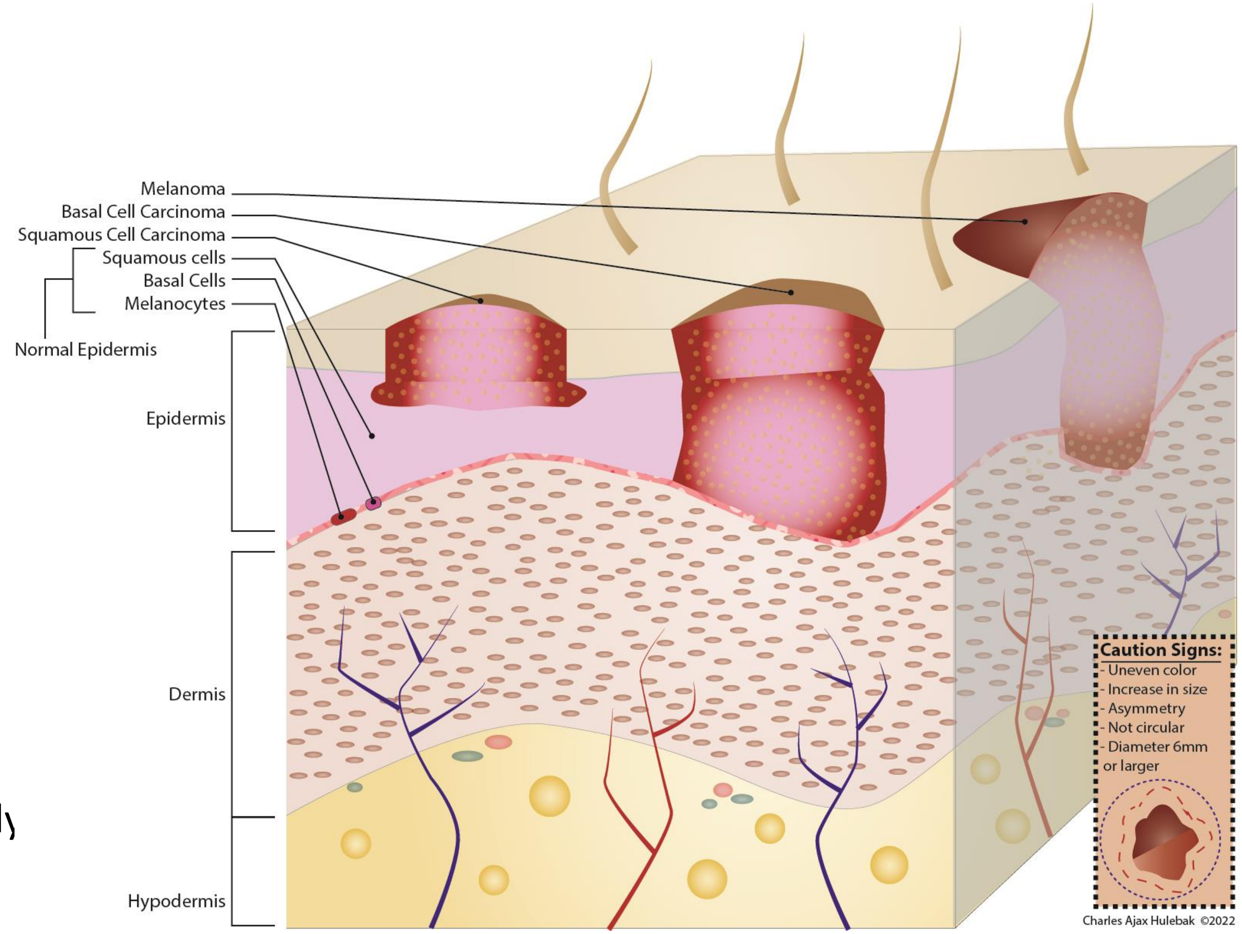
Basal Cell Carcinoma characteristics [9]

- Size: varies
- Shape: varies
- Color: pink or red (most common); brown, black or flecks of these colors; yellowish; white; blue
- Characteristics: may be waxy, dip in the center, scaly patch of skin, sore that may bleed, ooze, or crust over

*There are several stages of this skin condition, which have slightly different characteristics.

There are many similar characteristics, yet distinct differences. How is a deep learning program going to distinguish between them?

Skin Cells Diagram



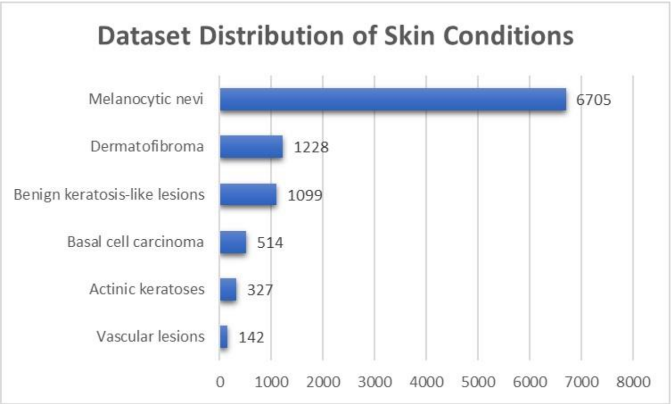
Project Walkthrough: Programming and Tuning the CNN

The initial steps for this project included downloading and reviewing both the metadata and image data for the project.

Following the analysis of the data and determination of the project objective, the intent was to compare two variables, much like the "Cats vs. Dogs" project. In this case, we are assessing the difference between two types of skin cells that are difficult to distinguish with the eye.

The original dataset included 6,705 mole images, and 514 basal cell carcinoma images. Although this was sufficient for preliminary testing, I decided to create more basal cell carcinoma images by rotating and mirroring these images, to create 4,112 images. Increasing this volume of data provides the ability to generate more accurate predictions for the testing and training data. The new combined testing data is 10,817 images.

In addition, the image sizes were very large, amounting to over 2GB of data and lengthy initial testing. The images were reduced from 650px x 450 px at 300dpi to 300px * 225px at 200dpi.



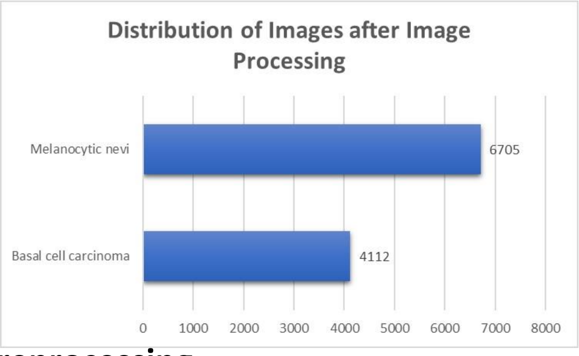
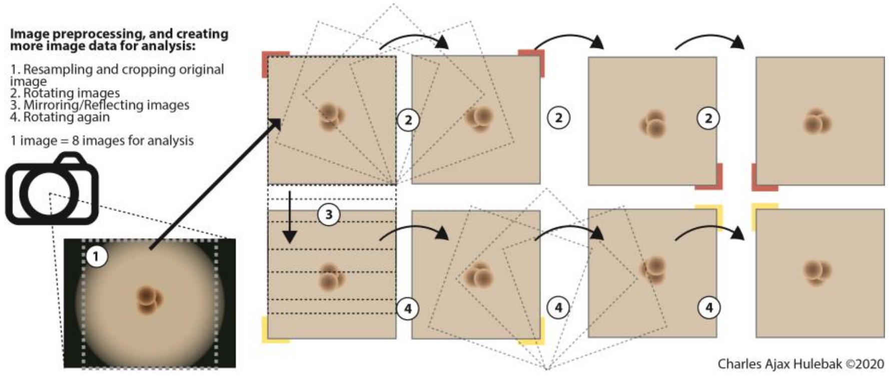


Image preprocessing



Charles A. Hulebak

Project Walkthrough: Programming and Tuning the CNN

The Cats vs. Dogs project provide a great baseline for testing various features within the CNN. I imported programming libraries that were necessary to conduct analysis for the project, including Pandas for data processing, NumPy for linear algebra, the Scikit-image package, TensorFlow, Keras, and a few others shown within the code section.

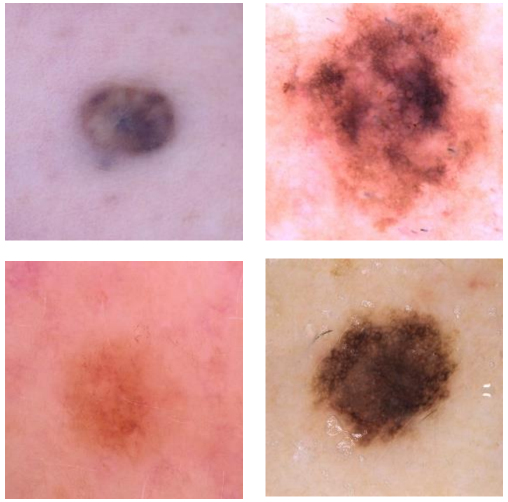
The initial testing results were significantly higher for early models, and it was evident that greater dpi resulted in higher accuracy for models. For example, the image dataset which contained images that were 450px x 450px at 300dpi had initial results around 85% with a single epoch.

Whereas this same image dataset that was reduced to 225px x 225px at 300 dpi were testing with a 65% accuracy rate.

A third dataset was used with smaller images, and resulted with an average of a 55% accuracy rating.

The time required to test the larger dataset was also significantly longer, and I was more interested in testing more models for performance than reaching a very high rating on the large dataset (which might be considered in the future).

Mole Images



Basal Cell Carcinoma Images

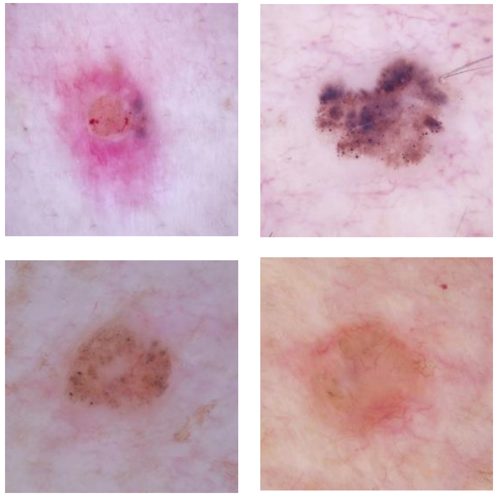
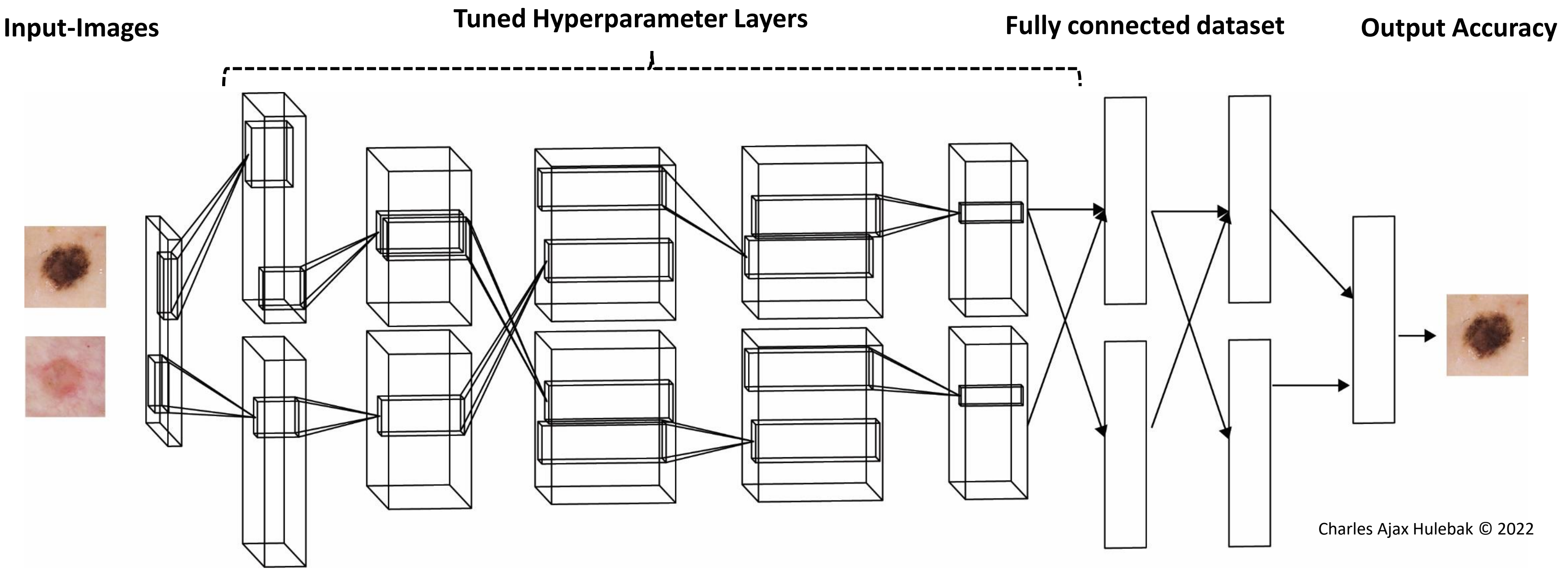


Image Credit: HAM10000 dataset

Project Walkthrough: Programming and Tuning the CNN

Through the process of modifying layer filters and tuning hyperparameters, I initially tested the model to determine the highest accuracy rate on a single epoch.

I continued to tune various layers and chose the best performing model within a single epoch. Then tested these on two epochs, followed by selecting the best performing model for further analysis. This model was tested with 10 epochs and evaluate the test accuracy, which resulted in 92.15% for the test data. It was evident that the learning process during these epochs provided much better results into this fully connected machine learning network.



Charles Ajax Hulebak © 2022

Results and Conclusion

I am very satisfied with the results from this project, as well as confident that the accuracy rating can be increased by analyzing higher resolution photographs, as well as increasing the sample size.

This project has a lot of potential, as well as other affiliated scientific and technology applications.

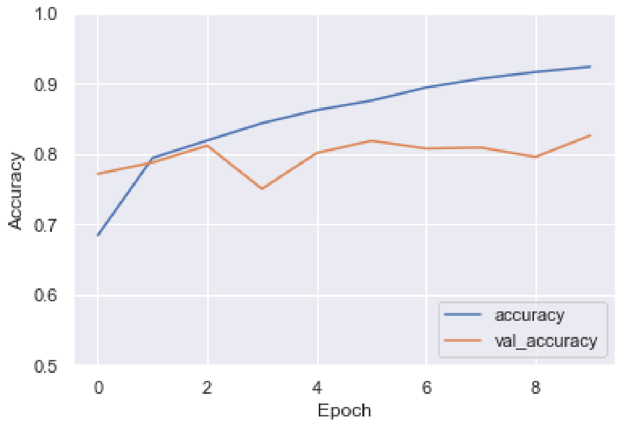
I was originally inspired to perform this project in light of the current COVID-19 pandemic, which is a global challenge and being addressed from many different perspectives to contain this virus, reduce the spread, and treat the those that have been infected. How can technology improve and control a condition such as this?

The CNN for this project has the ability to analyze a multitude of imagery, ranging from X-rays, to biological matter. Analysis can be conducted much faster for detection and containing than some of the other methods currently available.

I am certain that machine learning will continue to positively influence our future.

Testing and Training Data Results

```
Train on 8653 samples, validate on 2164 samples
Epoch 1/10
8653/8653 [==============================] - 1353s 156ms/sample - loss: 22.9174 - accuracy: 0.6845
- val_loss: 0.4868 - val_accuracy: 0.7717
Epoch 2/10
8653/8653 [==============================] - 1329s 154ms/sample - loss: 0.4605 - accuracy: 0.7943
- val_loss: 0.4902 - val_accuracy: 0.7879
Epoch 3/10
8653/8653 [==============================] - 1343s 155ms/sample - loss: 0.4104 - accuracy: 0.8193
- val_loss: 0.4831 - val_accuracy: 0.8119
Epoch 4/10
8653/8653 [==============================] - 1352s 156ms/sample - loss: 0.3609 - accuracy: 0.8436
- val_loss: 0.5081 - val_accuracy: 0.7505
Epoch 5/10
8653/8653 [==============================] - 1300s 150ms/sample - loss: 0.3224 - accuracy: 0.8624
- val_loss: 0.4634 - val_accuracy: 0.8013
Epoch 6/10
8653/8653 [==============================] - 1316s 152ms/sample - loss: 0.2949 - accuracy: 0.8758
- val_loss: 0.4709 - val_accuracy: 0.8189
Epoch 7/10
8653/8653 [==============================] - 1335s 154ms/sample - loss: 0.2599 - accuracy: 0.8945
- val_loss: 0.5123 - val_accuracy: 0.8078
Epoch 8/10
8653/8653 [==============================] - 1358s 157ms/sample - loss: 0.2305 - accuracy: 0.9070
- val_loss: 0.4601 - val_accuracy: 0.8091
Epoch 9/10
8653/8653 [==============================] - 1395s 161ms/sample - loss: 0.2118 - accuracy: 0.9164
- val_loss: 0.5344 - val_accuracy: 0.7957
Epoch 10/10
8653/8653 [==============================] - 1375s 159ms/sample - loss: 0.2015 - accuracy: 0.9240
- val_loss: 0.5419 - val_accuracy: 0.8262
```



References

1. Guy, G. P., Jr, Thomas, C. C., Thompson, T., Watson, M., Massetti, G. M., Richardson, L. C., & Centers for Disease Control and Prevention (CDC) (2015). Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. MMWR. Morbidity and mortality weekly report, 64(21), 591–596.
2. Rogers HW, Weinstock MA, Harris AR, et al. "Incidence Estimate of Nonmelanoma Skin Cancer in the United States", Arch Dermatol. 2010;146(3):283–287. (online) Available: https://doi:10.1001/archdermatol.2010.19
3. "Skin Cancer Facts & Statistics - The Skin Cancer Foundation". (2020). (online) Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/
4. "What are basal and squamous cell skin cancers?" American Cancer Society. (online) Available: http://www.cancer.org/cancer/skincancer-basalandsquamouscell/detailedguide/skin-cancer-basal-and-squamous-cell-what-is-basal-and-squamous-cell.
5. "Medical Expenditure Panel Survey", Agency for Healthcare Research and Quality, 2022. Rockville, MD: Agency for Healthcare Research and Quality; [January 2014]. (online) Available: https://www.ahrq.gov/data/meps.html
6. James, S., Abate, D., Abate, K., Abay, S., Abbafati, C., Abbasi, N. et al. (2018). "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017", The Lancet, 392(10159), 1789-1858. doi: 10.1016/S0140-6736(18)32279-7
7. Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V3, UNF:6:/APKSsDGVDhwPBWzsStU5A== [fileUNF] (online) Available: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T
8. "Moles: Signs and symptoms", American Academy of Dermatology Association, 2022. (online) Available: https://www.aad.org/public/diseases/a-z/moles-symptoms
9. "Basal Cell Carcinoma: Overview", American Academy of Dermatology Association, 2022. (online) Available: https://www.aad.org/public/diseases/skin-cancer/types/common/bcc

Addendum A - Global Prevalence of Skin Conditions

	Prevalence (thousands) 2017 counts	Incidence (thousands) 2017 counts	YLDs (thousands)				
			2017 counts	Percentage change in counts, 1990–2007	Percentage change in counts, 2007–17	Percentage change in age-standardised rates, 1990–2007	Percentage change in age-standardised rates, 2007–17
	7 369 526·2 (7 344 769·0 to 7 392 430·8)	**38 480 253·2 (36 469 390·1 to 40 567 963·0)**	**853 042·6 (642 084·6 to 1 097 347·2)**	**29·8% (28·8 to 30·8)**	**17·0% (16·4 to 17·6)**	**−3·0% (−3·5 to −2·5)**	**−0·9% (−1·4 to −0·4)**
Malignant skin melanoma	2324·4 (1794·8 to 2796·2)	308·7 (237·6 to 365·9)	140·9 (90·8 to 201·6)	104·3% (82·1 to 111·6)	32·3% (27·0 to 36·9)	41·3% (24·4 to 46·9)	4·9% (0·9 to 8·6)
Diagnosis and primary therapy phase of malignant skin melanoma	64·1 (49·0 to 76·8)	308·7 (237·6 to 365·9)	17·5 (11·0 to 25·2)	110·3% (89·4 to 118·3)	32·4% (26·1 to 38·8)	46·4% (30·5 to 52·6)	5·4% (0·5 to 10·6)
Controlled phase of malignant skin melanoma	2208·7 (1709·0 to 2662·4)	..	101·2 (59·9 to 154·2)	113·8% (93·9 to 121·4)	32·9% (27·7 to 37·9)	50·2% (34·9 to 55·4)	6·4% (2·3 to 10·7)
Metastatic phase of malignant	44·3 (32·7 to 49·6)	..	18·4 (11·3 to 25·1)	68·4% (40·0 to 80·0)	29·5% (22·6 to 35·5)	12·5% (−7·3 to 20·9)	−1·6% (−6·9 to 3·2)
Terminal phase of malignant skin	7·3 (5·4 to 8·2)	..	3·9 (2·4 to 5·2)	64·2% (37·2 to 73·7)	28·7% (21·6 to 33·9)	9·0% (−9·7 to 15·6)	−2·6% (−7·8 to 1·3)
Non-melanoma skin cancer	2537·1 (1666·4 to 3696·8)	7663·6 (5251·1 to 10 570·3)	90·2 (49·5 to 149·0)	211·2% (142·3 to 305·0)	32·7% (25·3 to 40·0)	96·8% (46·7 to 162·7)	−2·0% (−7·7 to 3·9)
Non-melanoma skin cancer (squamous-cell carcinoma)	2158·9 (1294·8 to 3255·8)	1778·8 (1068·8 to 2620·9)	87·7 (46·9 to 146·3)	221·7% (148·5 to 334·5)	32·8% (25·0 to 40·0)	102·5% (47·3 to 180·3)	−2·0% (−7·8 to 4·0)
Non-melanoma skin cancer (basal-cell carcinoma)	596·8 (325·9 to 947·0)	5884·8 (3702·9 to 8742·9)	2·5 (0·9 to 5·2)	44·5% (20·3 to 76·5)	30·8% (23·1 to 37·0)	−3·8% (−21·1 to 18·1)	−1·0% (−6·4 to 3·2)
Skin and subcutaneous diseases	1 974 238·4 (1 916 671·8 to 2 034 645·7)	4 185 971·3 (3 971 760·5 to 4 391 218·2)	41 621·9 (27 371·7 to 61 859·5)	24·0% (22·9 to 25·3)	13·0% (12·5 to 13·6)	0·9% (0·4 to 1·4)	0·6% (0·1 to 1·1)
Dermatitis	291 689·4 (276 520·7 to 308 059·0)	274 034·1 (246 120·0 to 302 497·5)	11 128·1 (6484·1 to 17 733·4)	19·3% (18·2 to 20·7)	12·1% (11·3 to 12·8)	−0·1% (−0·7 to 0·5)	1·1% (0·4 to 1·8)
Atopic dermatitis	205 517·2 (193 701·2 to 218 582·4)	27 134·4 (25 282·9 to 29 055·0)	9003·4 (4887·0 to 14 981·0)	17·0% (16·1 to 17·9)	11·6% (10·8 to 12·5)	0·4% (−0·4 to 1·1)	1·7% (0·9 to 2·6)
Contact dermatitis	79 666·7 (70 425·7 to 89 554·9)	221 252·8 (193 775·3 to 249 726·6)	1989·2 (1304·4 to 2950·5)	31·0% (28·8 to 33·2)	14·4% (12·6 to 16·2)	−1·6% (−2·3 to −1·0)	−1·1% (−1·7 to −0·5)
Seborrhoeic dermatitis	10 035·9 (9450·2 to 10 668·4)	25 646·9 (23 981·6 to 27 315·9)	135·6 (78·1 to 215·0)	20·8% (18·5 to 23·3)	8·4% (6·6 to 10·3)	−7·7% (−9·0 to −6·4)	−7·1% (−8·3 to −5·8)
Psoriasis	64 609·6 (62 454·3 to 66 767·5)	7846·6 (7564·9 to 8162·8)	5569·5 (3956·1 to 7354·3)	43·1% (42·0 to 44·2)	21·1% (20·3 to 21·9)	5·1% (4·3 to 5·8)	2·3% (1·6 to 2·9)
Bacterial skin diseases	11 397·9 (11 061·4 to 11 741·7)	266 779·7 (260 229·3 to 273 665·3)	177·9 (112·3 to 274·1)	29·2% (27·2 to 31·0)	15·5% (14·0 to 17·1)	2·0% (−0·5 to 4·4)	0·8% (−1·1 to 2·9)
Cellulitis	2071·5 (1952·2 to 2186·2)	42 958·6 (40 535·7 to 45 172·9)	118·2 (78·7 to 167·8)	26·7% (24·9 to 28·6)	12·9% (11·7 to 14·2)	−2·6% (−3·8 to −1·3)	−2·9% (−3·8 to −2·0)
Pyoderma	10 557·2 (10 245·5 to 10 853·2)	223 821·2 (217 649·0 to 230 259·5)	59·6 (23·9 to 123·5)	35·0% (33·7 to 36·4)	20·9% (19·8 to 22·2)	13·0% (11·9 to 14·2)	8·8% (7·8 to 9·9)
Impetigo	4620·7 (4358·2 to 4870·5)	84 007·5 (79 485·3 to 88 529·9)	26·4 (10·4 to 55·8)	35·7% (33·6 to 37·8)	20·7% (18·7 to 22·7)	24·1% (22·2 to 26·0)	13·4% (11·5 to 15·2)
Abscess and other bacterial skin diseases	5936·4 (5770·4 to 6098·7)	139 813·6 (135 314·1 to 143 965·8)	33·2 (13·4 to 69·5)	34·5% (32·8 to 36·3)	21·1% (19·7 to 22·6)	5·5% (4·2 to 6·8)	5·1% (3·8 to 6·4)
Scabies	175 406·7 (154 517·9 to 198 404·1)	527 476·5 (462 050·9 to 598 087·9)	4528·7 (2506·4 to 7414·6)	16·1% (13·8 to 18·5)	6·6% (5·3 to 8·0)	−4·3% (−5·0 to −3·6)	−3·1% (−3·6 to −2·6)
Fungal skin diseases	743 458·4 (681 568·4 to 808 149·7)	2 126 927·9 (1 917 361·6 to 2 317 274·7)	4154·5 (1651·4 to 8633·2)	21·5% (19·0 to 24·0)	10·9% (9·2 to 12·5)	−3·1% (−3·8 to −2·4)	−4·4% (−5·4 to −3·4)
Tinea capitis	160 239·3 (133 390·6 to 194 439·3)	303 016·6 (245 340·4 to 369 760·2)	916·5 (357·3 to 1965·4)	−5·8% (−7·3 to −4·3)	−13·0% (−14·5 to −11·3)	−13·8% (−15·0 to −12·7)	−19·0% (−20·4 to −17·3)
Other fungal skin diseases	583 219·1 (526 500·8 to 645 363·9)	1 823 911·3 (1 638 297·9 to 2 009 788·9)	3238·0 (1305·4 to 6694·4)	37·1% (35·2 to 39·3)	20·2% (19·0 to 21·6)	1·8% (1·5 to 2·1)	1·3% (1·1 to 1·5)
Viral skin diseases	130 639·2 (125 604·0 to 136 047·9)	116 329·8 (111 012·4 to 121 710·1)	4033·0 (2595·4 to 5995·6)	10·4% (9·8 to 11·1)	6·4% (6·0 to 6·9)	−2·8% (−3·2 to −2·3)	−1·8% (−2·2 to −1·4)
Viral warts	54 309·6 (52 104·2 to 56 403·2)	30 140·5 (29 078·5 to 31 211·0)	1662·2 (1066·0 to 2439·8)	13·7% (12·8 to 14·6)	5·6% (4·8 to 6·3)	−6·8% (−7·3 to −6·3)	−4·8% (−5·4 to −4·3)
Molluscum contagiosum	76 329·6 (71 596·6 to 81 134·1)	86 189·3 (80 984·6 to 91 427·2)	2370·7 (1511·7 to 3525·9)	8·2% (7·5 to 8·9)	7·1% (6·5 to 7·6)	0·2% (−0·3 to 0·8)	0·3% (−0·3 to 0·8)
Acne vulgaris	119 082·3 (107 127·9 to 133 651·4)	60 118·8 (53 260·2 to 68 180·7)	2547·6 (1518·8 to 4056·6)	46·1% (44·6 to 47·6)	16·2% (15·2 to 17·2)	19·8% (18·6 to 20·9)	11·4% (10·3 to 12·5)
Alopecia areata	15 981·0 (15 477·3 to 16 515·9)	28 185·2 (27 302·2 to 29 126·3)	523·1 (334·9 to 774·8)	28·8% (27·6 to 30·1)	12·7% (11·8 to 13·8)	−2·3% (−3·1 to −1·4)	−1·5% (−2·3 to −0·6)
Pruritus	71 224·3 (63 948·2 to 80 034·2)	55 643·1 (49 208·3 to 62 689·4)	755·6 (356·1 to 1433·2)	36·4% (34·1 to 38·8)	18·9% (17·5 to 20·4)	2·9% (2·5 to 3·4)	1·6% (1·2 to 2·0)
Urticaria	83 610·0 (73 335·4 to 95 162·9)	147 198·5 (129 941·2 to 166 345·6)	5014·8 (3321·0 to 7046·4)	19·3% (17·4 to 21·5)	10·8% (9·8 to 11·9)	0·7% (0·2 to 1·2)	0·3% (−0·2 to 0·8)
Decubitus ulcer	1143·7 (1022·6 to 1288·5)	4199·3 (3752·4 to 4741·3)	181·2 (125·8 to 244·3)	45·2% (42·7 to 47·7)	28·9% (26·3 to 31·9)	−3·3% (−4·9 to −1·8)	−0·4% (−2·2 to 1·5)
Other skin and subcutaneous diseases	550 810·3 (538 490·7 to 563 961·3)	571 231·9 (558 726·3 to 584 649·5)	3008·1 (1446·7 to 5557·4)	45·1% (44·6 to 45·6)	25·6% (25·2 to 26·0)	6·6% (6·4 to 6·8)	4·3% (4·1 to 4·5)

Table: Global prevalence, incidence, and YLDs for 2017; percentage change of YLD counts; and percentage change of age-standardized YLD rates for 1990–2007 and 2007–17 for both sexes combined for all Level 5 causes, nature of injury aggregates, and nine impairments (abbreviated content)[6]

Addendum B-Jupyter notebook excerpts (1)

Image Analysis and Recognition using a CNN and Skin Condition Data

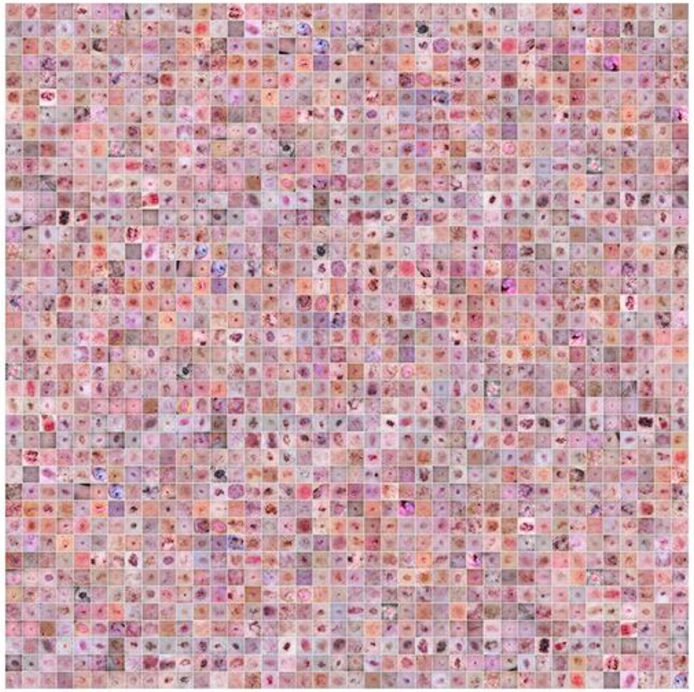
Introduction to Business Applications of Machine Learning (X522)

This project references the Kaggle Skin Cancer MNIST: HAM10000 analysis

Original Data Sources

-Original Challenge: https://challenge2018.isic-archive.com -https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T [1] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, Allan Halpern: "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)", 2018; https://arxiv.org/abs/1902.03368 [2] Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161 (2018).

This project will take a subset of original 10,000 images for the specific classification of one of two skin conditions. These conditions, which can be easily mistaken from one another include the mole and signs of basal cell carcinoma. Utilizing a convolutional neural network (CNN), this machine learning project is seeking to identify physical characteristics from images to determine if an image has a high probability of being one or the other.



In [1]:

```python
#Import libraries

import pandas as pd # data processing, CSV file

import numpy as np # linear algebra

#Scikit-image package
import matplotlib.pyplot as plt
%matplotlib inline
import matplotlib.mlab as mlab
import matplotlib.patches as mpatches

import skimage.io
from skimage import data
from skimage.io import imread
from skimage.io import imread_collection
from skimage.io import imshow

import tensorflow as tf
from tensorflow.keras.layers import Flatten, Dense
import re
import os
from os.path import join, exists, basename, normpath, split

import math
import random

#non-relative paths
import glob2
from glob2 import glob

import cv2
import imutils

from shutil import copyfile
from scipy.stats import norm
import seaborn as sns
sns.set(color_codes=True)

skin_analysis = os.path.join('..', 'input')
```

Load machine learning libraries

Let's take a look at the original dataset:

In [2]:

```python
imageid_path_dict = {os.path.splitext(os.path.basename(x))[0]: x
                     for x in glob(os.path.join(skin_analysis, '*', '*.jpg'))}

lesion_type_dict = {
    'nv': 'Melanocytic nevi',
    'mel': 'dermatofibroma',
    'bkl': 'Benign keratosis-like lesions ',
    'bcc': 'Basal cell carcinoma',
    'akiec': 'Actinic keratoses',
    'vasc': 'Vascular lesions',
    'df': 'Dermatofibroma'
}

#Define path to metadata (.csv file)
tile_df1 = pd.read_csv(os.path.join('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis', 'HAM10000_metadata_orig

tile_df2 = pd.read_csv(os.path.join('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis', 'HAM10000_metadata_subs

tile_df3 = pd.read_csv(os.path.join('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis', 'HAM10000_metadata_fina

#Paths to images
tile_df1['path'] = tile_df1['image_id'].map(imageid_path_dict.get)
tile_df1['cell_type'] = tile_df1['dx'].map(lesion_type_dict.get)
tile_df1['cell_type_idx'] = pd.Categorical(tile_df1['cell_type']).codes

tile_df2['path'] = tile_df2['image_id'].map(imageid_path_dict.get)
tile_df2['cell_type'] = tile_df2['dx'].map(lesion_type_dict.get)
tile_df2['cell_type_idx'] = pd.Categorical(tile_df2['cell_type']).codes

tile_df3['path'] = tile_df3['image_id'].map(imageid_path_dict.get)
tile_df3['cell_type'] = tile_df3['dx'].map(lesion_type_dict.get)
tile_df3['cell_type_idx'] = pd.Categorical(tile_df3['cell_type']).codes

#Metadata information
tile_df1.describe(exclude=[np.number])
```

Associate metadata files

Out [2]:

	lesion_id	image_id	dx	dx_type	sex	localization	path	cell_type
count	10015	10015	10015	10015	10015	10015	0	10015
unique	7470	10015	7	4	3	15	0	7
top	HAM_0003789	ISIC_0034282	nv	histo	male	back	NaN	Melanocytic nevi
freq	6	1	6705	5340	5406	2192	NaN	6705

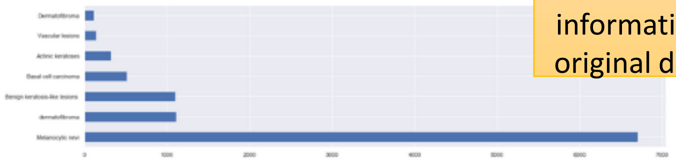
Associate metadata files

Skin Cell Distribution (original dataset)

In [3]:

```python
fig, ax1 = plt.subplots(1, 1, figsize = (20, 5))
tile_df1['cell_type'].value_counts().plot.barh()
```

Out [3]: <matplotlib.axes._subplots.AxesSubplot at 0x1e55e14b320>



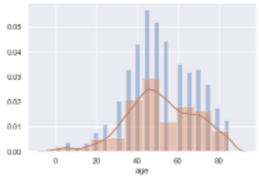
Review statistical information for original dataset

Age Distribution (original dataset)

In [4]:

```python
#Excel file-show statistical information for dataset (Age Distribution)

df = pd.read_csv('HAM10000_metadata_original.csv')
#df = sns.load_csv('HAM10000_metadata.csv')

# Make default histogram of age
sns.distplot( df["age"] )
#sns.plt.show()

# Control the number of bins
sns.distplot( df["age"], bins=10 )
#sns.plt.show()
```

Out [4]: <matplotlib.axes._subplots.AxesSubplot at 0x221ef2502b0>



Review nominal information for original dataset (Age distribution)

Data Preprocessing

There are seven categories of skin conditions, however for this analysis we are interested in the ability to identify the differences between a common mole and a cancerous cell. These skin conditions are refered to as the melanocytic nevi (mole) and the basal cell carcinoma (cancerous cell). Let's take a look at these categories:

In [5]:

```python
#Metadata information
tile_df2.describe(exclude=[np.number])
```

Addendum B-Jupyter notebook excerpts (2)

Data Preprocessing

There are seven categories of skin conditions, however for this analysis we are interested in the ability to identify the differences between a common mole and a cancerous cell. These skin conditions are referred to as the melanocytic nevi (mole) and the basal cell carcinoma (cancerous cell). Let's take a look at these categories:

In [5]:
```
#Metadata information
tile_df2.describe(exclude=[np.number])
```

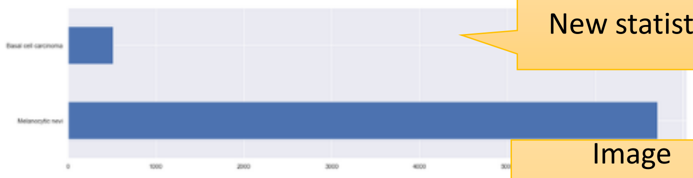
Out[5]:

	lesion id	image id	dx	dx type	sex	localization	path	cell type
count	7219	7219	7219	7219	7219	7219	0	7219
unique	5730	7219	2	3	3	15	0	2
top	HAM_0005263	ISIC_0033512	nv	follow_up	male	back	NaN	Melanocytic nevi
freq	6	1	6705	3704	3735	1613	NaN	6705

Separate data to targeted values (moles and cancerous skin cells)

In [6]:
```
fig, ax1 = plt.subplots(1, 1, figsize = (20, 5))
tile_df2['cell_type'].value_counts().plot.barh()
```

Out[6]: `<matplotlib.axes._subplots.AxesSubplot at 0x221ef11a5f8>`



New statistics

The new dataset includes 6,705 melanocytic nevi images, and 514 basal cell carcinoma images. In order to test ... to include more basal cell carinoma images.

Image preprocessing, increasing dataset

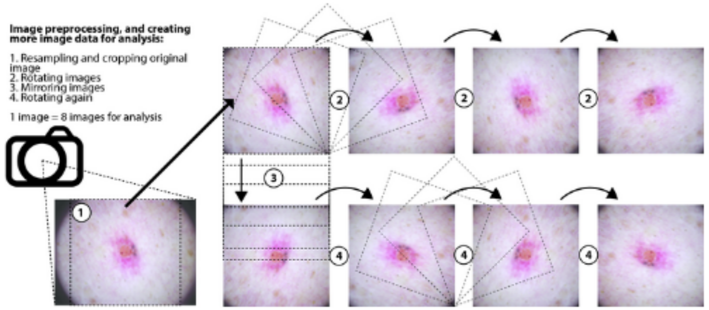
Image Data Preprocessing

The original dataset file size was in excess of 2Gb with images sized at 650 x 450 x 300dpi. By closely examining the dataset, there are several advantages of resizing the images:
-Some of the images contain camera lens boundaries (which are black in the image), which will decrease the testing accuracy.

-The large file size requires a higher level of processing time.

The new file sizes are standardized and cropped to 90 x 90 x 300 dpi.

In addition, we are able to increase the quantity of basal cell carcinoma files from the original 514 images. With the process of rotating (90, 180, and 270 degrees) and mirrowing these images (rotating a second time), we are able to test a total of 4,112 images, which will provide a higher testing accuracy.



The Dataset used for Machine Learning

The final dataset that will be used contains the following information:

In [7]:
```
#Metadata information
tile_df3.describe(exclude=[np.number])
```

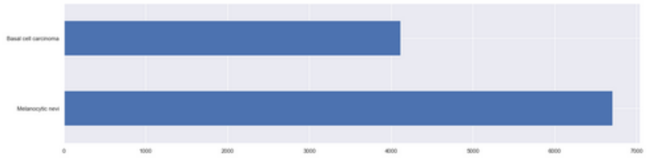
Out[7]:

	lesion id	image id	dx	dx type	sex	localization	orientation	path	cell type
count	10817	10817	10817	10817	10817	10817	10817	0	10817
unique	5730	10817	2	3	3	15	9	0	2
top	HAM_0004621	nv3125	nv	histo	male	back	original	NaN	Melanocytic nevi
freq	32	1	6705	6610	5957	2915	6705	NaN	6705

Final data values used for analysis

In [8]:
```
fig, ax1 = plt.subplots(1, 1, figsize = (20, 5))
tile_df3['cell_type'].value_counts().plot.barh()
```

Out[8]: `<matplotlib.axes._subplots.AxesSubplot at 0x221f0475358>`



This proportion of images for testing accuracy is much better.

In [9]:
```
#Self-check, remove for final
tile_df3['path']
```

Out[9]:
```
0        None
1        None
2        None
3        None
4        None
         ...
10812    None
10813    None
10814    None
10815    None
10816    None
Name: path, Length: 10817, dtype: object
```

Choosing the machine learning platform and tuning the Convolutionary Neural Network

This process includes testing different layers and features from the prescribed CNN. In this project, we are using Tensorflow and Keras, and assigning various layers in Keras to determine which model can most accurately predict the differences between these two skin cells.

Implementing the CNN using TensorFlow and Keras

Now, the images will be loaded into memory for analysis

```
#Import libraries and testing with Tensorflow
import tensorflow as tf    # Import tensorflow (Keras included)
import os
import cv2
```

```
#Show color and resampled image in grayscale
img = cv2.imread('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis/images/nv/nv0009.jpg')
plt.imshow(cv2.cvtColor(img, cv2.COLOR_BGR2RGB))
plt.show()

img = cv2.imread('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis/images/nv/nv0009.jpg', cv2.IMREAD_GRAYSCALE)
IMSIZE = 225
new_array = cv2.resize(img, (IMSIZE, IMSIZE))
plt.imshow(new_array, cmap='gray')
```

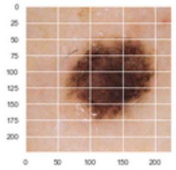
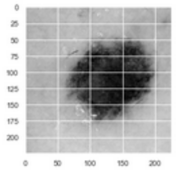


Image preview, filter on secondary image

Out[5]: `<matplotlib.image.AxesImage at 0x1e39a0f96d8>`



In [6]:
```
#Create training data from images, this will take some time depending on size of data
training_data = []

def create_training_data():
    for category in ['bcc', 'nv']:
        path = os.path.join('C:/Users/admin/01_Business_Apps_of_ML(X522)/skin_analysis/images', category)
        class_label = {'bcc':0, 'nv':1}[category]
        for img in os.listdir(path):
            try:
                img_array = cv2.imread(os.path.join(path,img), cv2.IMREAD_GRAYSCALE)
                new_array = cv2.resize(img_array, (IMSIZE, IMSIZE))
                training_data.append([new_array, class_label])
            except Exception as e:
                pass

create_training_data()
```

Creating random training data from dataset

In [7]:
```
#Show total quantity of data/images that will be used for analysis
#Shuffle the images randomly for the training data
print(len(training_data))

import random

random.shuffle(training_data)
```

10817

Addendum B-Jupyter notebook excerpts (3)

Show feature set for data

In [8]:
```python
#Show feature set for data
X = []
y = []

for features, label in training_data:
    X.append(features)
    y.append(label)
X = np.array(X).reshape(-1, IMSIZE,IMSIZE,1)
y = np.array(y)

print(X.shape, y.shape)
```

```
(10517, 225, 225, 1) (10517,)
```

Implementing array assigning training and test datasets

In [9]:
```python
#Numpy core array assigned to the dataset for training and test datasets
import pickle
pickle_out = open('X.pickle', 'wb')
pickle.dump(X, pickle_out)
pickle_out.close()

pickle_out = open('y.pickle', 'wb')
pickle.dump(y, pickle_out)
pickle_out.close()
```

In [10]:
```python
#Tensorflow and Keras Convolutionary Neural Network with imported layers for image detection
#Refer to Keras layer information at https://keras.io/api/models/sequential/#sequential-class
#Refer to Keras layer information at https://keras.io/api/layers/core_layers/
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Activation, Flatten, Conv2D, Dense, Dropout, Flatten, MaxPooling2D
import pickle
```

Testing model performance on various hyperparameters

In [9]:
```python
#Running CNN on assigned layers, training data-no test
tf.random.set_seed(0)

# define a sequential neural network
model = Sequential()

model.add(Flatten()) # Flatten image
model.add(Dense(1000, activation='relu'))# Adds a densely-connected layer with
#model.add(Dense(800, activation='relu')) #Omitted for this test
model.add(Dense(10, activation='softmax')) # Number of neurons Final layer

model.compile(optimizer='adam', #'adam'  #'adamax',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

model.fit(X,y, epochs=1)
```

```
Train on 10517 samples
10517/10517 [==============================] - 65s 6ms/sample - loss: 2797.0415 - accuracy: 0.6545
```

Out[9]: <tensorflow.python.keras.callbacks.History at 0x241500266a0>

In [10]:
```python
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
flatten (Flatten)	multiple	0
dense (Dense)	multiple	50626000
dense_1 (Dense)	multiple	10010

Total params: 50,636,010
Trainable params: 50,636,010
Non-trainable params: 0

Testing model performance on various hyperparameters

In [40]:
```python
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 10% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu')) #ReLU - rectified linear unit, computationally effecient
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(64))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=2, validation_split=0.1)
```

```
Train on 9735 samples, validate on 1082 samples
Epoch 1/2
9735/9735 [==============================] - 342s 35ms/sample - loss: 3.7100 - accuracy: 0.7405 - val_loss: 0.5221 - val_a
ccuracy: 0.7532
Epoch 2/2
9735/9735 [==============================] - 387s 40ms/sample - loss: 0.4104 - accuracy: 0.8180 - val_loss: 0.4033 - val_a
ccuracy: 0.8087
```

In [42]:
```python
model.summary()
```

Model: "sequential_5"

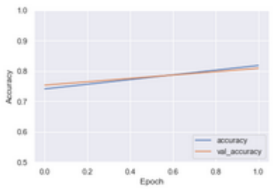
Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 223, 223, 64)	640
activation_6 (Activation)	(None, 223, 223, 64)	0
max_pooling2d_3 (MaxPooling2	(None, 111, 111, 64)	0
flatten_5 (Flatten)	(None, 788544)	0
dense_10 (Dense)	(None, 64)	50466880
dense_11 (Dense)	(None, 1)	65
activation_7 (Activation)	(None, 1)	0

Total params: 50,467,585
Trainable params: 50,467,585
Non-trainable params: 0

Reviewing performance and accuracy results

In [41]:
```python
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```

Out[41]: <matplotlib.legend.Legend at 0x25f4dd0b8d0>



Second test, varying hyperparameters

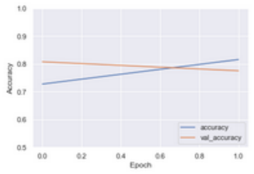
In [19]:
```python
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 20% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu')) #ReLU - rectified linear unit, computationally effi
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(64))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=2, validation_split=0.2)
```

```
Train on 8653 samples, validate on 2164 samples
Epoch 1/2
8653/8653 [==============================] - 265s 31ms/sample - loss: 5.4257 - accuracy: 0.7276 - val_loss: 0.4411 - val_a
ccuracy: 0.8078
Epoch 2/2
8653/8653 [==============================] - 265s 31ms/sample - loss: 0.4069 - accuracy: 0.8159 - val_loss: 0.4491 - val_a
ccuracy: 0.7754
```

Reviewing performance and accuracy results

In [22]:
```python
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```

Out[22]: <matplotlib.legend.Legend at 0x221c7e04630>



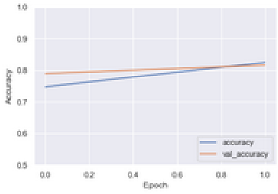
Third test, varying hyperparameters

In [11]:
```python
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 20% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(1000))
#model.add(Dense(32))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=2, validation_split=0.2)
```

```
Train on 8653 samples, validate on 2164 samples
Epoch 1/2
8653/8653 [==============================] - 1394s 161ms/sample - loss: 6.7557 - accuracy: 0.7474 - val_loss: 0.4634 - val
_accuracy: 0.7893
Epoch 2/2
8653/8653 [==============================] - 1344s 155ms/sample - loss: 0.3990 - accuracy: 0.8245 - val_loss: 0.4595 - val
_accuracy: 0.8156
```

Addendum B-Jupyter notebook excerpts (4)

In [12]:
```
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```
Out[12]: `<matplotlib.legend.Legend at 0x2418d52c6a0>`

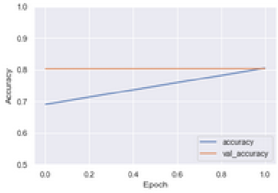


Reviewing performance and accuracy results

In [20]:
```
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 20% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(1000))
model.add(Dense(500))
model.add(Dense(1))
model.add(Activation('sigmoid'))


model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=2, validation_split=0.2)
```
```
Train on 8653 samples, validate on 2164 samples
Epoch 1/2
8653/8653 [==============================] - 1351s 156ms/sample - loss: 11.3758 - accuracy: 0.6905 - val_loss: 0.4843 - va
l_accuracy: 0.8027
Epoch 2/2
8653/8653 [==============================] - 1331s 154ms/sample - loss: 0.4402 - accuracy: 0.8049 - val_loss: 0.4864 - val
_accuracy: 0.8031
```

Fourth test, varying hyperparameters

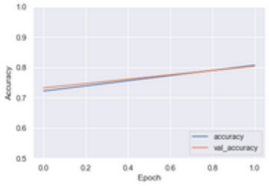
In [21]:
```
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```
Out[21]: `<matplotlib.legend.Legend at 0x2418dd08668>`



In [11]:
```
#Final Model test, chose model 2, now running 10 epochs
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 20% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(500))
#model.add(Dense(64))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=2, validation_split=0.2)
```
```
Train on 8653 samples, validate on 2164 samples
Epoch 1/2
8653/8653 [==============================] - 737s 85ms/sample - loss: 7.8275 - accuracy: 0.7219 - val_loss: 0.5086 - val_a
ccuracy: 0.7320
Epoch 2/2
8653/8653 [==============================] - 728s 84ms/sample - loss: 0.4469 - accuracy: 0.8069 - val_loss: 0.4266 - val_a
ccuracy: 0.8041
```

Fifth test, varying hyperparameters

In [12]:
```
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```
Out[12]: `<matplotlib.legend.Legend at 0x1e5ef672d30>`

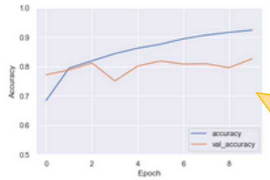


Reviewing performance and accuracy results

In [13]:
```
#Modifying layers to test accuracy on different assigned layers
#Add validation parameter, currently assigned to 20% of images
x = pickle.load(open('X.pickle', 'rb'))
y = pickle.load(open('y.pickle', 'rb'))

X = x/225.0

model = Sequential()
model.add(Conv2D(64, (3,3), input_shape=(225,225,1)))

model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Flatten())
model.add(Dense(1000))
#model.add(Dense(32))
model.add(Dense(1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy']
             )

history = model.fit(X,y, epochs=10, validation_split=0.2)
```
```
Train on 8653 samples, validate on 2164 samples
Epoch 1/10
8653/8653 [==============================] - 1353s 156ms/sample - loss: 22.9174 - accuracy: 0.6845 - val_loss: 0.4868 - va
l_accuracy: 0.7717
Epoch 2/10
8653/8653 [==============================] - 1329s 154ms/sample - loss: 0.4605 - accuracy: 0.7943 - val_loss: 0.4902 - val
_accuracy: 0.7879
Epoch 3/10
8653/8653 [==============================] - 1343s 155ms/sample - loss: 0.4104 - accuracy: 0.8193 - val_loss: 0.4831 - val
_accuracy: 0.8119
Epoch 4/10
8653/8653 [==============================] - 1352s 156ms/sample - loss: 0.3609 - accuracy: 0.8436 - val_loss: 0.5081 - val
_accuracy: 0.7805
Epoch 5/10
8653/8653 [==============================] - 1300s 150ms/sample - loss: 0.3224 - accuracy: 0.8624 - val_loss: 0.4634 - val
_accuracy: 0.8013
Epoch 6/10
8653/8653 [==============================] - 1316s 152ms/sample - loss: 0.2949 - accuracy: 0.8758 - val_loss: 0.4709 - val
_accuracy: 0.8189
Epoch 7/10
8653/8653 [==============================] - 1335s 154ms/sample - loss: 0.2599 - accuracy: 0.8945 - val_loss: 0.5123 - val
_accuracy: 0.8078
Epoch 8/10
8653/8653 [==============================] - 1358s 157ms/sample - loss: 0.2305 - accuracy: 0.9070 - val_loss: 0.4801 - val
_accuracy: 0.8091
Epoch 9/10
8653/8653 [==============================] - 1395s 161ms/sample - loss: 0.2118 - accuracy: 0.9164 - val_loss: 0.5344 - val
_accuracy: 0.7957
Epoch 10/10
8653/8653 [==============================] - 1375s 159ms/sample - loss: 0.2015 - accuracy: 0.9240 - val_loss: 0.5419 - val
_accuracy: 0.8262
```

Sixth test, running on best of previous 5 models

In [14]:
```
plt.plot(history.history['accuracy'], label='accuracy')
plt.plot(history.history['val_accuracy'], label = 'val_accuracy')
plt.xlabel('Epoch')
plt.ylabel('Accuracy')
plt.ylim([0.5, 1])
plt.legend(loc='lower right')
```
Out[14]: `<matplotlib.legend.Legend at 0x1e5ef5c7630>`



Reviewing performance and accuracy results. This model had a 92.4% accuracy rating for predicting the difference between a normal and cancerous skin cell. (slightly better than initial 91.25%)

Charles A. Hulebak

Thank you

Keywords: applications, algorithms, analytics, convolutional neural network (CNN), data science, databases, data processing, dataset, deep learning, dermatology, digital health, health information, image processing, machine learning, model tuning, python, skin cancer, skin cells, skin conditions, tensorflow.

Attribution-NonCommercial-ShareAlike 4.0 International

Creative Commons Corporation ("Creative Commons") is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an "as-is" basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and conditions that creators and other rights holders may use to share original works of authorship and other material subject to copyright and certain other rights specified in the public license below. The following considerations are for informational purposes only, are not exhaustive, and do not form part of our licenses.

Considerations for licensors: Our public licenses are intended for use by those authorized to give the public permission to use material in ways otherwise restricted by copyright and certain other rights. Our licenses are irrevocable. Licensors should read and understand the terms and conditions of the license they choose before applying it. Licensors should also secure all rights necessary before applying our licenses so that the public can reuse the material as expected. Licensors should clearly mark any material not subject to the license. This includes other CC-licensed material, or material used under an exception or limitation to copyright. More considerations for licensors:

wiki.creativecommons.org/Considerations_for_licensors

Considerations for the public: By using one of our public licenses, a licensor grants the public permission to use the licensed material under specified terms and conditions. If the licensor's permission is not necessary for any reason—for example, because of any applicable exception or limitation to copyright--then that use is not regulated by the license. Our licenses grant only permissions under copyright and certain other rights that a licensor has authority to grant. Use of the licensed material may still be restricted for other reasons, including because others have copyright or other rights in the material. A licensor may make special requests, such as asking that all changes be marked or described. Although not required by our licenses, you are encouraged to respect those requests where reasonable. More considerations for the public:

wiki.creativecommons.org/Considerations_for_licensees

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License
By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 -- Definitions.

a. Adapted Material means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synced in timed relation with a moving image.

b. Adapter's License means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.
c. BY-NC-SA Compatible License means a license listed at creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License.
d. Copyright and Similar Rights means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.
e. Effective Technological Measures means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.
f. Exceptions and Limitations means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.
g. License Elements means the license attributes listed in the name of a Creative Commons Public License. The License Elements of this Public License are Attribution, NonCommercial, and ShareAlike.
h. Licensed Material means the artistic or literary work, database, other material to which the Licensor applied this Public License.
i. Licensed Rights means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.
j. Licensor means the individual(s) or entity(ies) granting rights under this Public License.
k. NonCommercial means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this Public License, the exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.
l. Share means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the may access the material from a place and at a time individually chosen by them.
m. Sui Generis Database Rights means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.
n. You means the individual or entity exercising the Licensed Rights under this Public License. Your has a corresponding meaning.

Section 2 -- Scope.

a. License grant.
Subject to the terms and conditions of this Public License, Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
a. reproduce and Share the Licensed Material, in whole or in part, for NonCommercial purposes only; and
b. produce, reproduce, and Share Adapted Material for NonCommercial purposes only.

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, making modifications authorized by this Section 2(a) (4) never produces Adapted Material.

5. Downstream recipients.

a. Offer from the Licensor -- Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

b. Additional offer from the Licensor -- Adapted Material. Every recipient of Adapted Material from You automatically receives an offer from the Licensor to exercise the Licensed Rights in the Adapted Material under the conditions of the Adapter's License You apply.

c. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. Other rights.

Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.

2. Patent and trademark rights are not licensed under this Public License.

3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties, including when the Licensed Material is used other than for NonCommercial purposes.

Section 3 -- License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. Attribution.

If You Share the Licensed Material (including in modified form), You must:

a. retain the following if it is supplied by the Licensor with the Licensed Material:

i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);

ii. a copyright notice;

iii. a notice that refers to this Public License;

iv. a notice that refers to the disclaimer of warranties;

v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

b. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

c. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.

b. ShareAlike.

In addition to the conditions in Section 3(a), if You Share Adapted Material You produce, the following conditions also apply.

The Adapter's License You apply must be a Creative Commons license with the same License Elements, this version or later, or a BY-NC-SA Compatible License.

2. You must include the text of, or the URI or hyperlink to, the Adapter's License You apply. You may satisfy this condition in any reasonable manner based on the medium, means, and context in which You Share Adapted Material.

3. You may not offer or impose any additional or different terms conditions on, or apply any Effective Technological Measures to, Adapted Material that restrict exercise of the rights granted under the Adapter's License You apply.

Section 4 -- Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database for NonCommercial purposes only;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material, including for purposes of Section 3(b); and

c. You must comply with the conditions in Section 3(a) if You Share or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 -- Disclaimer of Warranties and Limitation of Liability.

a. UNLESS OTHERWISE SEPARATELY UNDERTAKEN BY THE LICENSOR, TO THE EXTENT POSSIBLE, THE LICENSOR OFFERS THE LICENSED MATERIAL AS-IS AND AS-AVAILABLE, AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE LICENSED MATERIAL, WHETHER EXPRESS, IMPLIED, STATUTORY, OR OTHER. THIS INCLUDES, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OR ABSENCE OF ERRORS, WHETHER OR NOT KNOWN OR DISCOVERABLE. WHERE DISCLAIMERS OF WARRANTIES ARE NOT ALLOWED IN FULL OR IN PART, THIS DISCLAIMER MAY NOT APPLY TO YOU.

b. TO THE EXTENT POSSIBLE, IN NO EVENT WILL THE LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY (INCLUDING, WITHOUT LIMITATION, NEGLIGENCE) OR OTHERWISE FOR ANY DIRECT, SPECIAL, INDIRECT, INCIDENTAL, CONSEQUENTIAL, PUNITIVE, EXEMPLARY, OR OTHER LOSSES, COSTS, EXPENSES, OR DAMAGES ARISING OUT OF THIS PUBLIC LICENSE OR USE OF THE LICENSED MATERIAL, EVEN IF THE LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH LOSSES, COSTS, EXPENSES, OR DAMAGES. WHERE A LIMITATION OF LIABILITY IS NOT ALLOWED IN FULL OR IN PART, THIS LIMITATION MAY NOT APPLY TO YOU.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 -- Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or

2. upon express reinstatement by the Licensor.

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Attribution-NonCommercial-ShareAlike 4.0 International

Section 7 -- Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.
b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 -- Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.
b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.
c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.
d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

===

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the “Licensor.” The text of the Creative Commons public licenses is dedicated to the public domain under the CC0 Public Domain Dedication. Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark "Creative Commons" or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

This data may also be protected by the California Consumer Privacy Act (CCPA) and other legal restriction depending upon your jurisdiction.

The data is publicly available and shared within the works cited section. Please download the data for yourself and create meaningful value, while respecting the investment of countless hours, training, education, time, and processes used to develop this publication as a digital product. Thank you for taking time to read, learn, and hopefully inspiration to contribute towards a better future for all of us.