Skin Cancer Classification

Introduction:

Skin cancer is the most common cancer in the world. It's not the most deadly but the most common.

In the U.S., we find 5.4 million new cases of skin cancer every year. They come in different types;

Some are called Carcinomas, some are called Melanomas.

Melanomas are the ones that typically kill people; it's called the black cancer. Twenty percent of Americans will eventually get skin cancer, in most cases, benign cancer.

Pre-cancer also causes an Actinic Keratosis, affects 58 million Americans and many more in the world.

In the US, there are 76,000 Melanomas each year and 10,000 deaths.

To put this in perspective, traffic accidents in this country are 40,000, so, you are about four times as likely to die in a traffic collision as a skin cancer. The cost for United States per year is 8.1 billion in treatment and diagnostics.

Skin Cancer stages:

The curves I'm showing you here at this moment show the difference in survival outlook for the different stages.
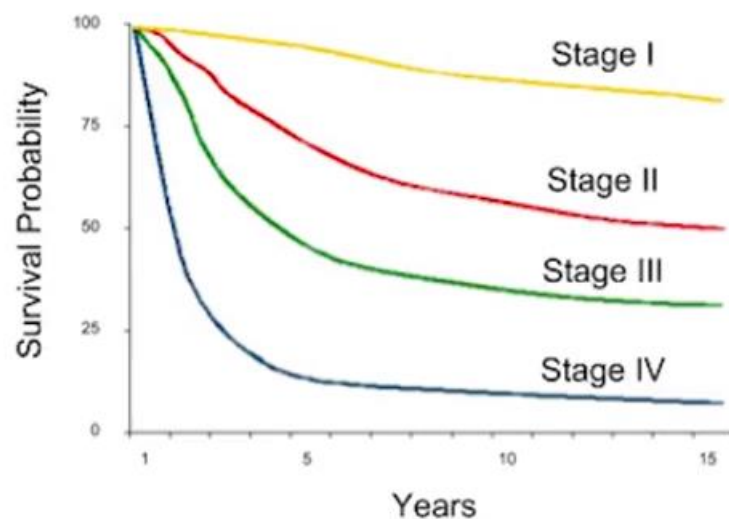

What's even more shocking is how similar stage zero and stage IV is.
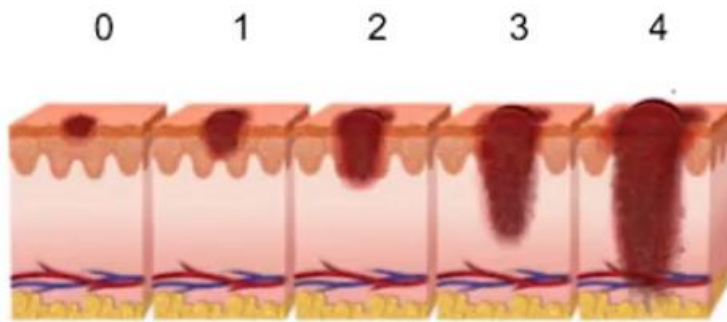
A stage zero cancer has barely penetrated the surface, but it is very visible. But a stage IV cancer has just grown below the skin layer and touched the underlying tissue and vessels. It hasn't even spread yet, and yet, with this kind of cancer, which can develop in less than a year, your survival chances would be under 20 percent.

Therefore, early detection is paramount to not dying from skin cancer.

Here, you see the natural evolution of a melanoma over the years from 2003 to 2006.
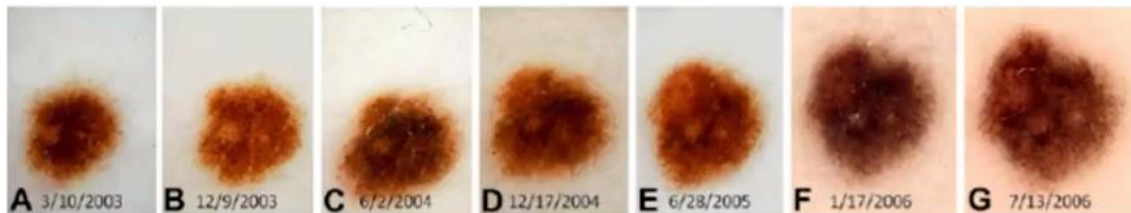
And there are characteristics that doctors use for finding melanomas. They look at the fuzziness of the border, the asymmetry, the coloration, the growth rate if it is accessible.

But still, classifying melanomas, as I've shown you, is very hard problem. Partially, it's hard because some people have lots of moles. These are clinical images of two backs and you can see how many moles they are, and finding the one that's malignant is really, really hard.

Copyright © Dr Eric Ehrsam, dermatologist

So, dermatologists are extremely highly trained to find melanomas and carcinomas. There are about 10,000 dermatologists and they are incredibly hard to train because they need to make life and death decisions in diagnosing your skin.

Just to show how difficult it is to find skin cancer, I challenge you to look at this image, the reader is invited to look at this image of correctly classified or diagnosed images (some are irrelevant cases but the point is how similar they are) You can see different types of lesions, and you can see benign and malignant lesions. And these are all correctly classified.
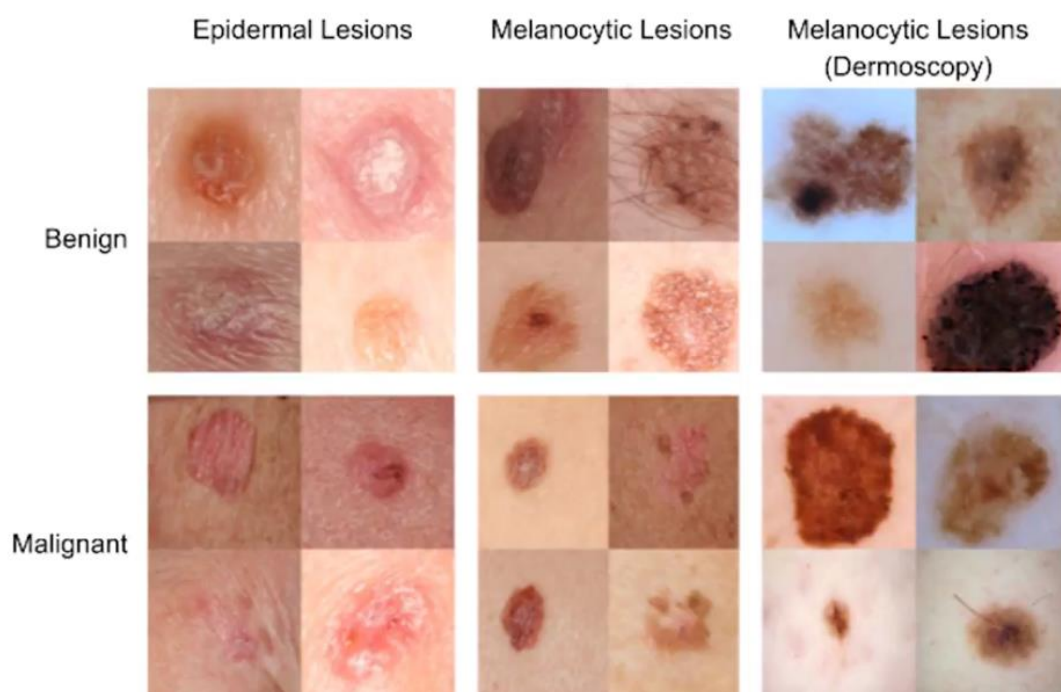

That's how a neural network is a good candidate because neural networks can process hundreds of thousands of images as a human doctor can never study as many as you can feed into a neural network.
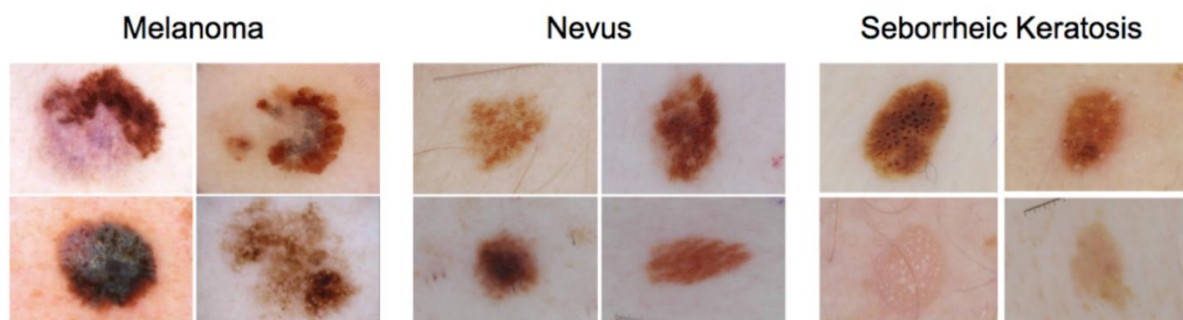
Training the neural network:

We didn't train a neural network from scratch; we alternatively used the transfer learning technique to train our neural network this technique can be summed as follows:

We use a pretrained network that has been trained on the famous ImageNet dataset, which is a huge dataset of more than 1.4 million images with 1000 different classes. And we use this network to classify our data.
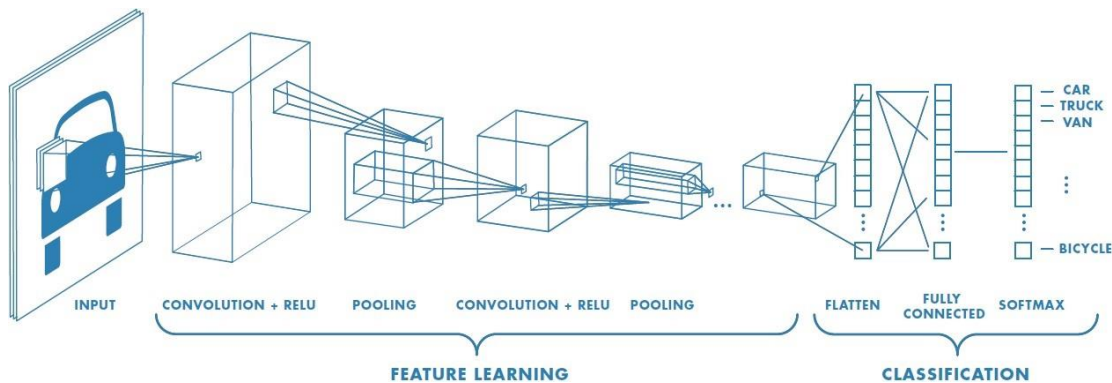
The catch here is our classes are not among those 1000 classes but it is fine. We only need the first part of the deep neural network which is responsible for detecting features out of an image, these features can be color, shape, edges or anything really that humans automatically can detect but don't know they distinguish according to.

Then we attach a classification layer to this neural network and only train the classifier part; so we basically do not train the whole network as this is inefficient in terms of power and time.

The results we obtained are the same as Prof.Sebastian Thrun 72% accuracy
Test Loss: 0.660876
Test Accuracy (Overall): 72% (435/600)

at the first glance this doesn't seem so high, but when compared to certified board dermatologists at Stanford result. It's much higher.

Validation set

Classifier	Three-way accuracy
Dermatologist 1	65.6%
Dermatologist 2	66.0%
CNN	69.5%
CNN - PA	**72.0%**

**Disease classes:
three-way classification**

0. Benign single lesions
1. Malignant single lesions
2. Non-neoplastic lesions

The Classification Pipeline:

- First we applied our needed transforms on the data which ensures data augmentation and proper sizes for feeding into network
- We then divided our data into training, validation and testing sets
- Then we fed our data to a train loader that ensures the data is fed to the neural network into batches to save memory.
- We then got our pretrained model which is a ResNet 152 with pretrained weights and turned off their gradients which mean its parameters won't be updated.
- We attached our classifier and omitted the ImageNet classifier.
- Then we started training for around 60 epochs of forward and back propagation.
- We used the early stop technique to ensure that our model doesn't over fit.
- After finishing we ran our test set and got our 72% accuracy
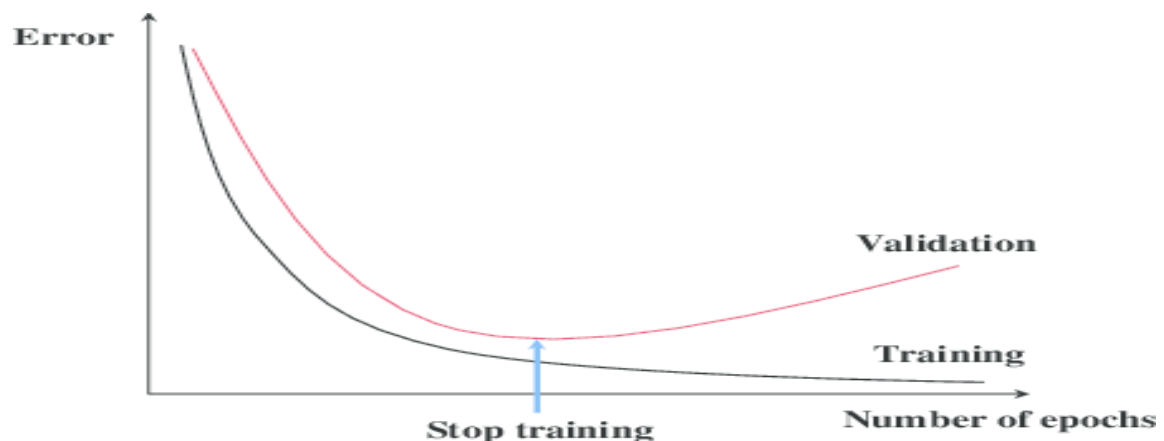
Yasser Al-Barbary
Yasser.m.albarbary@gmail.com
 linkedin.com/in/yasser-albarbary/
github.com/YasserAlBarbary