# **Machine Learning Engineer Nanodegree**

## Capstone Proposal

Mark Wright

January 25th 2018

## Proposal

### Domain Background

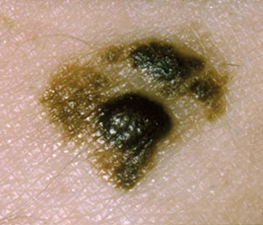
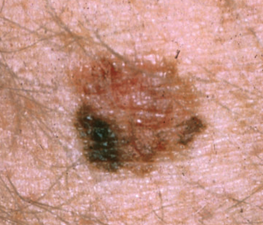
Since a young age I have been fascinated by how things work, in particular biology and nature. When deciding on careers I initially wanted to go into the field of medicine, but decided after some work experience at a local hospital that I wasn't cut out for all the long hours, so followed my next love technology and computing. It is because of this interest in biology that I want to focus my capstone project on the healthcare sector and tie this in with advances in computer vision.

Skin cancer rates globally are rising, according to the the WHO between 2-3 million non melanoma skin cancers and 132,000 melanoma skin cancers (a more dangerous form that is more easily spread to other parts of the body) occur every year, and 1 in every 3 cancer diagnoses are a skin cancer [1]. Incidence rates for melanoma skin cancer are projected to rise by 7% in the UK between 2014 and 2035 [2]. The rise appears mainly due to recreational exposure to the sun and the use of tanning equipment driven by modern culture, and the media. The depletion of earth's ozone layer, thought to be due to the release of harmful pollutants is also allowing more ultraviolet B rays to reach Earth, which can cause skin cancer. [3]

### Problem Statement

Skin cancers are usually diagnosed by a clinical specialist after a referral by a GP by visually inspecting the potential cancer, or an image of the potential cancer, followed by a biopsy to confirm a classification if they think the specimen likely cancerous.

There are several visual features that may indicate that a melanoma is cancerous. For example it could be unsymmetrical, have irregular borders, have uneven color, and/or be of a large size (at least the size of a pencil tip) [4].

Symmetry Border Color Size

Images from CancerResearch UK

Since there are visual clues in the diagnosis of skin cancer it is a suitable subject to utilise machine learning and computer vision to help specialists to make decisions. A classifier could be written that given an image of a specimen would indicate the likelihood of it being cancerous.

In the future if a classifier like this was utilised with a smartphone application it could greatly increase the speed at which a referral to a specialist was made, removing the patients GP from the diagnosis process, this is important because the earlier a cancer is diagnosed the easier it is to treat [5], and the better the potential outcome.

### Datasets and Inputs

The International Skin Imaging Collaboration: Melanoma Project have accumulated a large labelled image archive of skin cancer images that are publicly available for teaching and the testing of automated diagnostic systems [6]. With the aim to reduce melanoma mortality.

There are a total of 13,786 images in the archive [7]. 12668 are classified as benign, and 1084 are malignant. The archive is accompanied by a gallery navigation tool (powered by girder [8]) which allows for the images to be browsed and filtered by various facets. Each image is paired with a json file which stores metadata about the image including features such as whether the specimen was found to be benign/malignant, some data about the patient such as sex, age and family history. There are many features within the metadata for images which is not recorded for each image. I have decided for the course of this project to ignore all metadata apart from the target variable ‘benign\_malignant’.

I intend to utilise a subset of the dataset that was partitioned for the 2017 ISBI challenge. It contains a total of 2000 training images, 150 validation images and 600 testing images. [9]

### Solution Statement

I intend to make use of deep Convolutional Neural Networks to train a classifier that will be able to identify if an image of a mole is benign or malignant. I will use transfer learning to make use of an existing general image processing model that have already been trained to significantly reduce the training time required. I will remove the final fully connected layers of the model and replace with a new fully connected layer configured to classify images as benign or malignant. I also intend to experiment to see if an ensemble of trained CNN models, using the technique detailed above, achieves better or worse performance than the individual models.

### Benchmark Model

Last year Esteva et al, trained a CNN using transfer learning that is able to perform at a similar level to dermatologists [11]. For melanomas the average dermatologist in their trial classified around 95% of malignant lesions, and 76% of benign moles correctly. By comparison their algorithm was capable of classifying 96% of malignant lesions and 90% of benign lesions.

Whilst I am unlikely to be able to get close to this level of accuracy, given the size of my dataset and computational power I have access to, what I will hope to be able to show it that, for a relatively novice data scientist with access to a laptop and open source machine learning libraries such as keras [12], tensorflow [13] and scikit learn [10], that a classifier that operates at a comparable level to a human is possible.

### Evaluation Metrics

I will use Binary cross entropy to assess the performance of each CNN classifier since there are only two classes we want to predict (malignant/benign).

For the final Random Forest classifier I will assess the accuracy of predictions, and also experiment at maximising recall since it is more important for a medical test that we do not miss true positives. I will produce a ROC curve to show how the recall/precision rate changes by altering the decision boundaries.

### Project Design

1. I will install all software that may be required for the project on my laptop, to make this repeatable I will store the configuration as a conda env. I will make use of keras, tensorflow and scikit learn.
2. I will download all images from the dataset specified in early sections on to my local macbook, and ensure the training, validation and testing datasets are stored in separate locations. I will then conduct some eda over the images to ensure that they are all the same size, and if they are not produce a function to normalise them.
3. I will then download several existing image classifier models [14], for example, ResNet50, Exception, InceptionV3, that will be the base of each of my CNN classifiers. For each I will repeat the process to train them described below in step 4 - 5.
4. I will remove the final layer of the model, and add a new fully connected layer with 2 outputs, the same as the number of classes I want my model to be able to predict (benign, malignant). I will set the initial state of the model to have the default weights of the pre-trained classifier. I will experiment with the number of training epochs that the model is trained over. I will utilise the training and validation sets that were pre-split in the challenge during the training of the model. I will use checkpointing so that the weights of the best epoch of training/validation will be stored and that is what will be used when the model is assessed using the testing dataset.
5. The hidden testing set will be used to assess the final performance of the best checkpointed classifier to reduce the risk that the model is over fit to the training/validation set.
6. Once I have a suitable set of trained classifiers I will see if using an ensemble technique such as Random Forest [15] can be used to improve the overall performance by combining and weighting the output of each classifier. I will process the output from each classifier so it is of the form:

*Image 1 : CNN\_1\_Malignant | CNN\_2\_Malignant | CNN\_n\_Malignant*

*…. …. …. …. …. …. ….*

*Image n : CNN\_1\_Malignant | CNN\_2\_Malignant | CNN\_n\_Malignant*

1. I will then use scikit learn to train a Random Forest classifier that combines these results from the training set. I will experiment with tuning the classifier, by using grid search to alter parameter.
2. I will test the performance of the classifier by getting a prediction from each of the CNN’s using the test data set, converting to the data format above, and use the random forest classifier to make a final prediction. I will compare the predictions to the the actual target labels of the test set to work out the final accuracy of the ensemble to see if it is better than the individual CNN trained via transfer learning.
3. I will generate a ROC curve for the classifier so that it is possible to see the recall and precision tradeoff and how it is affected by altering the decision threshold [14].

### References

[1] <http://www.who.int/uv/faq/skincancer/en/index1.html>

[2] <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer#heading-Zero>

[3] <https://www.nationalgeographic.com/environment/global-warming/ozone-depletion/>

[4] <https://www.nhs.uk/be-clear-on-cancer/symptoms/skin-cancer#QxQ74ksgVC7py2mo.97>

[5] <http://www.cancerresearchuk.org/about-cancer/skin-cancer/getting-diagnosed/seeing-your-gp>

[6] <http://isdis.net/isic-project/>

[7] <https://isic-archive.com/#images>

[8] <http://girder.readthedocs.io/en/latest/user-guide.html>

[9] <https://challenge.kitware.com/#phase/5840f53ccad3a51cc66c8dab>

[10] <http://scikit-learn.org/>

[11]] <https://www.nature.com/articles/nature21056>

[12] <https://keras.io/>

[13] <https://www.tensorflow.org/>

[14] <https://keras.io/applications/#available-models>

[15] <http://scikit-learn.org/stable/modules/ensemble.html#forest>

[16] <https://towardsdatascience.com/fine-tuning-a-classifier-in-scikit-learn-66e048c21e65>