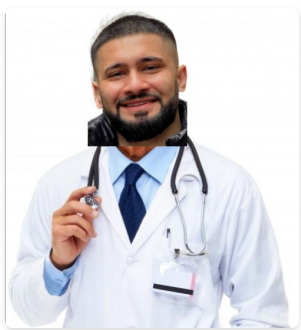
How to become something you’re not - my dermatologist story

Hasib Ahmed

Have you ever wanted to become something you're not?

- 🥼 **Doctor**
- 📈 **Stock trader**
- 👨‍🚀 **Astronaut***

With code & a magic thing called ✨ **Machine Learning**✨ you can! You can even do it part time, just like how I became a part time (unpaid) dermatologist 🔬



*probably can't be an astronaut with just code

What is Machine Learning?

Reality

What is Machine Learning?

machine learning

noun

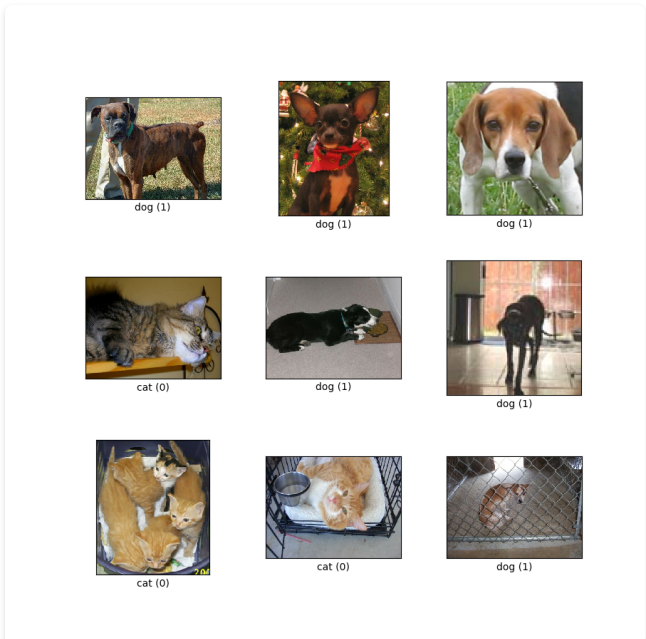
1. A field of study concerned with the design and development of **algorithms** and techniques that allow computers to **learn**.

We teach machines to guess what things are, or will, be based on lots and lots of data

How do machines learn? 🧠💻

The machine or algorithm can understand whatever we tell it to understand. We pick out features, which is known as feature engineering, to help the algorithm to figure out what a piece of data is. A piece of data could be numbers, images, video, etc

Common example is classifying cats and dogs



Why I become a part time Dermatologist

A family member was diagnosed with Alopecia Areata which is a disease that happens when the immune system attacks hair follicles and causes hair loss.

What serious skin conditions are out there?

- **Melanoma**
- Benign keratosis
- Melanocytic nevi
- Basal cell carcinoma
- Actinic keratoses
- Vascular lesions
- Dermatofibroma

Now I need data & lots of it

HAM10000 dataset: a large collection of multi-source dermatoscopic images of common pigmented skin lesions

10, 000 images 🤔

Explore the dataset, with the power of code

```
# Categories of the diferent diseases
lesion_type_dict = {
    'nv': 'Melanocytic nevi',
    'mel': 'Melanoma',
    'bkl': 'Benign keratosis ',
    'bcc': 'Basal cell carcinoma',
    'akiec': 'Actinic keratoses',
    'vasc': 'Vascular lesions',
    'df': 'Dermatofibroma'
}
```

image_id	lesion_id	dx	dx_type	age	sex	localization
ISIC_0027419	HAM_0000118	bkl	histo	80.0	male	scalp
ISIC_0025030	HAM_0000118	bkl	histo	80.0	male	scalp
ISIC_0026769	HAM_0002730	bkl	histo	80.0	male	scalp
ISIC_0025661	HAM_0002730	bkl	histo	80.0	male	scalp
ISIC_0031633	HAM_0001466	bkl	histo	75.0	male	ear

Explore the dataset, with the power of code

```
df.lesion.value_counts()
```

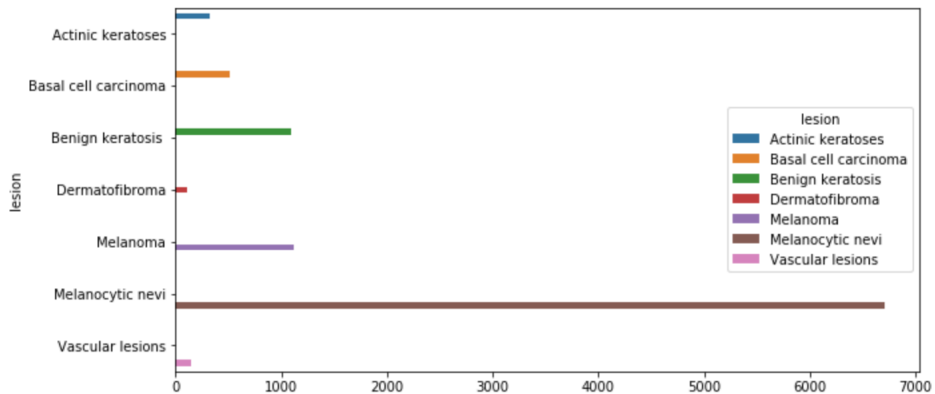
```
Melanocytic nevi          6705
Melanoma                  1113
Benign keratosis          1099
Basal cell carcinoma       514
Actinic keratoses          327
Vascular lesions           142
Dermatofibroma             115
Name: lesion, dtype: int64
```

Something doesn't look right

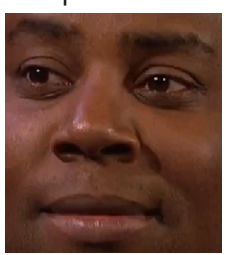
Explore the dataset, with the power of code

Let's visualise that data

```
fig, ax1 = plt.subplots(1, 1, figsize = (10, 5))
sns.countplot(y = 'lesion', data = df, hue = "lesion", ax = ax1)
```



Suspicions confirmed 🤨



Now I need a machine learning algorithm to use

My requirements:

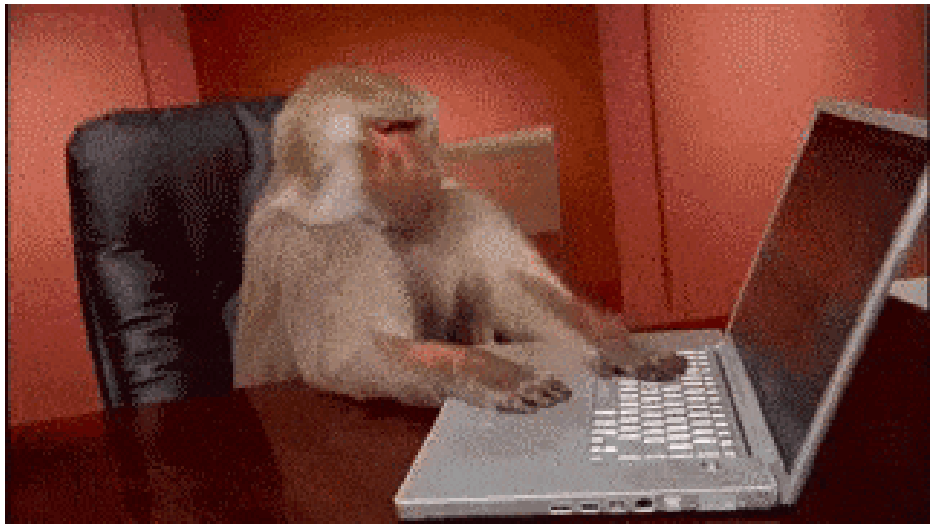
- Lightweight for a mobile application
- Ideally pretrained to save me some time & additional work
- A cool name

MobileNet Deep Neural Network

Pretrained refers to a machine learning model created by someone else and trained on a large dataset to solve a similar problem.

Time to train the model

Time to train the model



Steps to train

- Read and pre-process the data

```
df = pd.read_csv(os.path.join(base_directory, 'HAM10000_metadata.csv'))
# creation of new columns for readability later
df['path'] = df['image_id'].map(img_path_dict.get)
df['cell_type'] = df['dx'].map(lesion_types_dict.get)
df['cell_type_idx'] = pd.Categorical(df['cell_type']).codes
df['age'].fillna((df['age'].mean()), inplace=True) # because None values not supported.
```

Steps to train

- Data Augmentation - Artificially expand the HAM10000 dataset in order to avoid the model overfitting

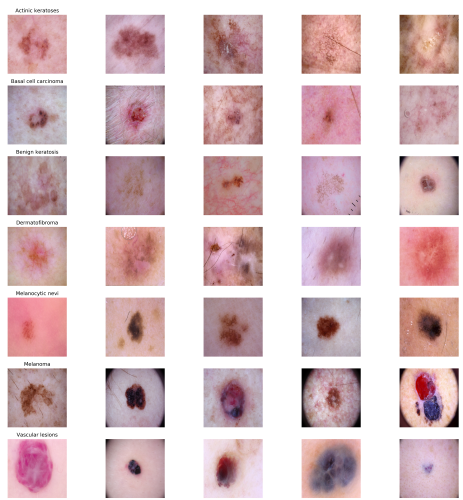
```
datagen = ImageDataGenerator(
        rotation_range=20,  # rotate images between 0-180 degrees randomly
        zoom_range = 0.1, # zoom image randomly
        width_shift_range=0.1,  # shift images horizontally (the input is a fraction of total width) randomly
        height_shift_range=0.1,  # shift images vertically (thei nput is a fraction of total height) randomly
        horizontal_flip=True,  # flip images horizontally randomly
        vertical_flip=False)  # flip images vertically randomly
```

This gave me a lot more images than the 10,000 I started with

Steps to train

- Resize the images to 224x224 as MobileNet cannot handle the original image dimesions (450x600)

```
df['image'] = df['path'].map(lambda x: np.asarray(Image.open(x).resize((224,224))))
```

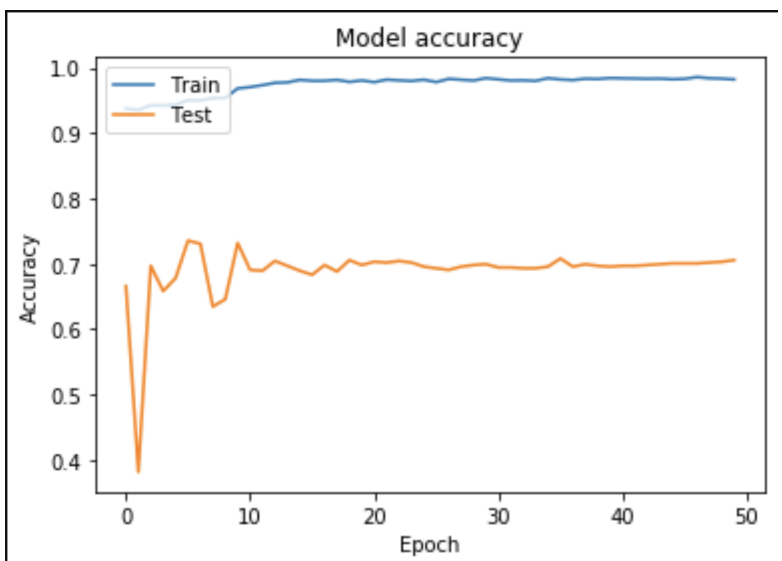


Test the model’s accuracy

- 80/20 split for training and testing data - mock test
- 10% split of training data for a validation set - actual exam

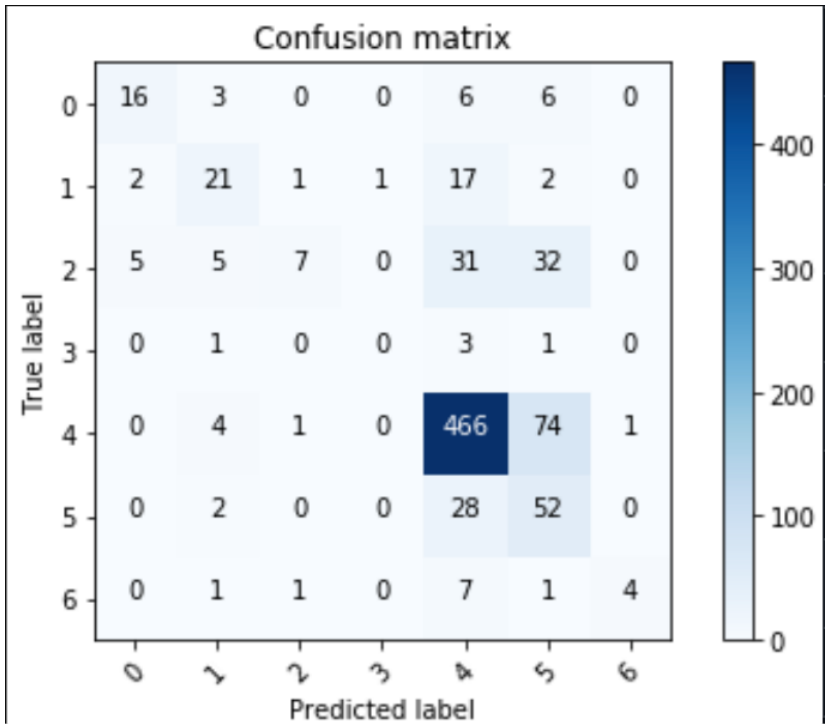
Test the model's accuracy

```
# Plot the training and validation accuracy
plt.plot(history.history['accuracy'])
plt.plot(history.history['val_accuracy'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper left')
plt.show()
```



Confusion Matrix

```
def plot_confusion_matrix(...): ...
```



Lesion Name	Code
Actinic keratoses	0
Basal cell carcinoma	1
Benign keratosis	2
Dermatofibroma	3
Melanoma	4
Melanocytic nevi	5
Vascular lesion	6

Results

Training accuracy of 98.2% - Overfitting has occurred because...
Overall accuracy of 70.3% across 7 classification classes (types of skin lesion) - Beating dermatologists with 3-5 years of experience (1)
85.35% accuracy for Melanoma classification - Beating dermatologists with over 10 years experience (1)

Real Life Application

Convinced friends or family members to seek medical advice using an image taken of their skin and running it through the machine learning model

Ideal situation would be to improve the model & have it in an accessible mobile app so everyone can have this technology at their fingertips

Reference and Links

- (1) Morton, C.A and Mackie, R.M, 1998, Clinical accuracy of the cutaneous malignant melanoma. The British journal of dermatology, 132(2), pp.283-287
- https://github.com/1solation/ham10000_exploratory_data_analysis
- https://github.com/1solation/dermoscopic-skin-cancer-image-classifier-WTL

Questions?