AI皮膚疾病偵測服務

Skin Disease & Cancer Detection AI Services

吳智鴻

目的

- 以網路公開之兩個皮膚疾病圖片資料庫，以深度學習網路，開發AI皮膚病自動偵測系統。
- 具有：1.判斷疾病良惡性。 2. 判斷疾病種類（7種）

解決痛點

- 皮膚為常見之疾病，解決
  - 1. 延誤就醫問題
  - 2. 醫師推薦
- 解決方式
  - 透過手機/電腦，拍攝皮膚狀況照片，上傳AI程式。AI程式透過電腦視覺技術，分析照片，提供皮膚病的疾病辨識。

馬斯洛的需求層級理論

解決 身體安全的需求

SWOT分析

- Strength
  - AI技術
  - 擁有醫師團隊支援
- Weakness
  - 尚未建立自己的人工標註資料集
- Opportunity
  - 提供線上AI服務尚不多見
- Trend
  - 大廠的資源優勢

產業分析（同性質、類似產品分析）

開發「痣能達人MoleMe」時，就以超過5000筆手機拍攝的痣醫療影像進行模型的訓練與測試，建立的模型正確率超過93%。

從今年三月推出至今僅三個月餘，已經累積超過兩萬次的風險判斷，隨著未來數據持續增加，正確率有再升高的可能性。

民眾只要回答五個問題並上傳一張照片，就可以馬上獲得評估結果。該軟體最大的好處，在於提醒較高風險的民眾趕快就醫、以及對於較低風險的民眾提供定期提醒的服務，提昇民眾自我健康識能。

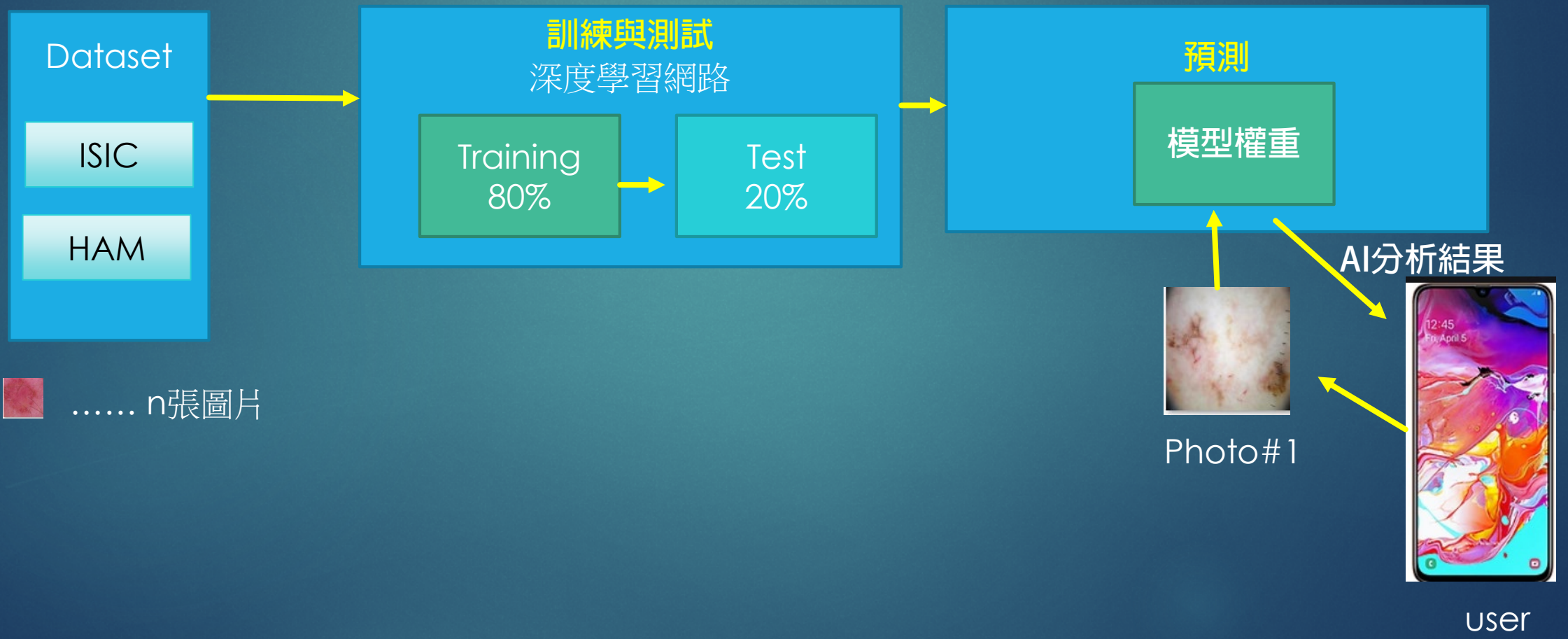
資料來源：https://www.gbimonthly.com/2019/08/50439/

WHY
民眾不直接去看皮膚科而需要AI服務？

- 不知道這各疾病的嚴重程度，延誤就醫。
- 比較私密的位置，比較避諱就醫。
- 當地可能就醫較不方便

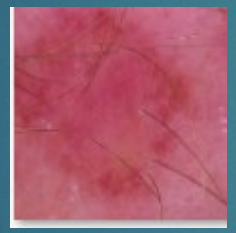
系統流程

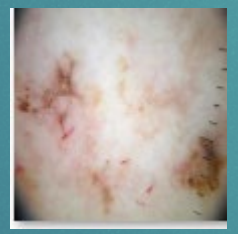
AI Model Development



API#1 皮膚良/惡性判斷
預計輸入輸出結果

- 輸入：based64編碼的圖片 （手機/PC）尺寸 224*224 彩色圖檔
- 處理：CNN卷積神經網路模型
- 輸出：

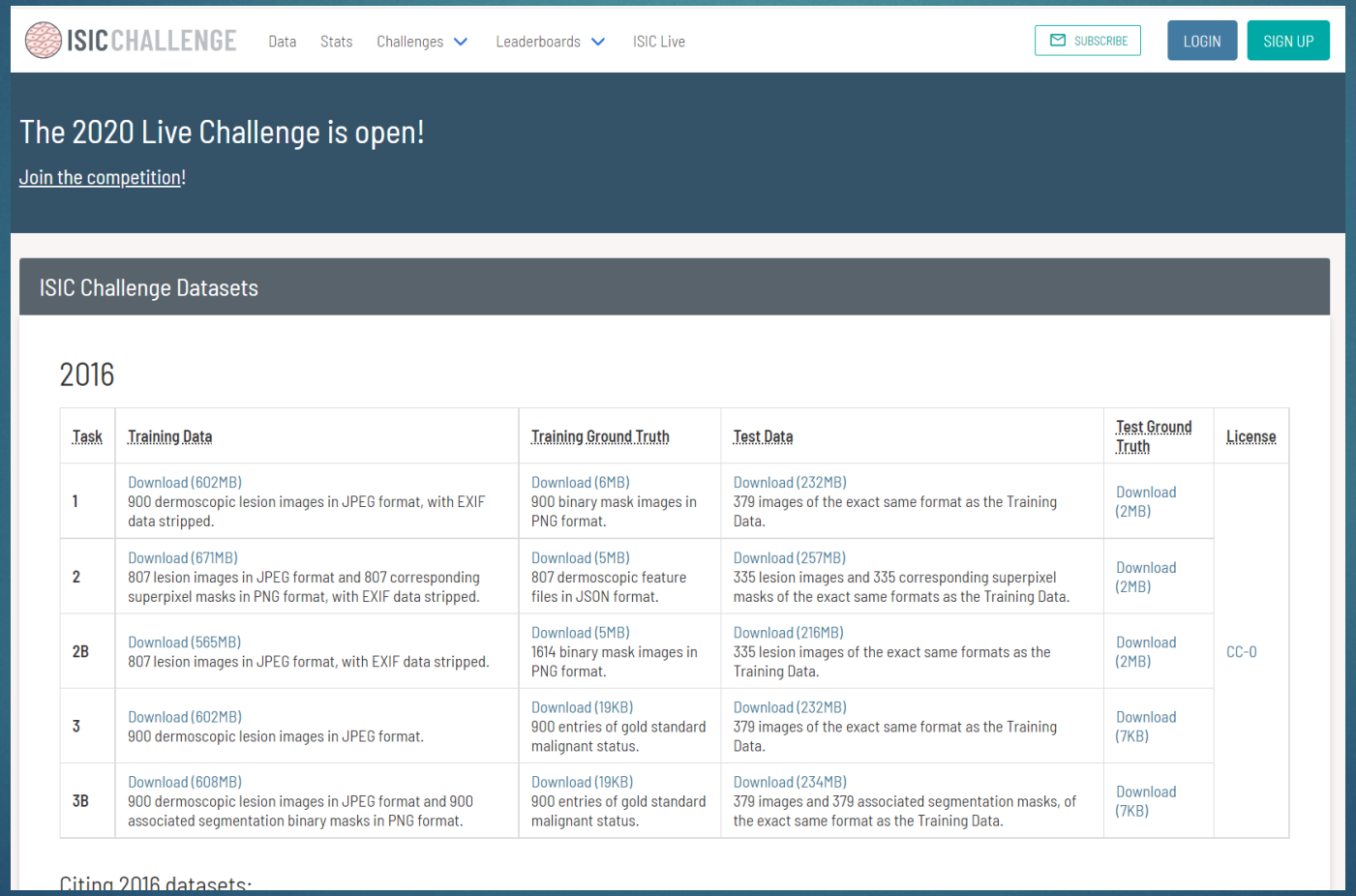




- 良性　　惡性

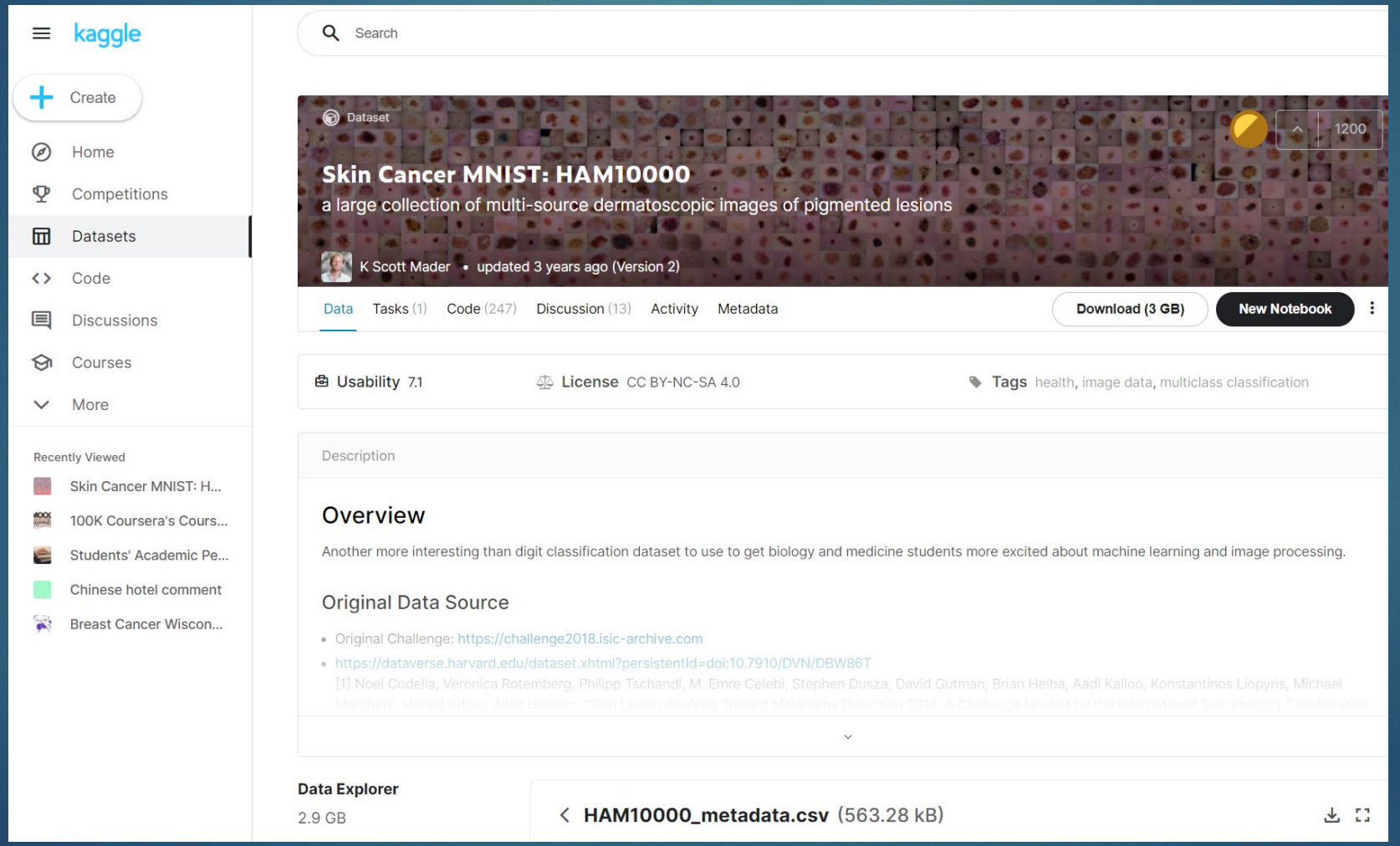
資料集來源#1

- International Skin Image Collaboration (ISIC)
- https://challenge.isic-archive.com/data/



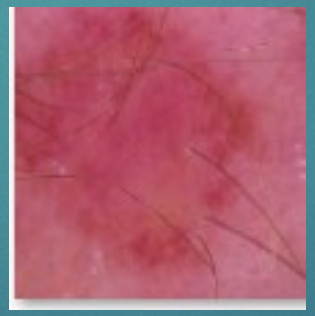
資料集來源#2

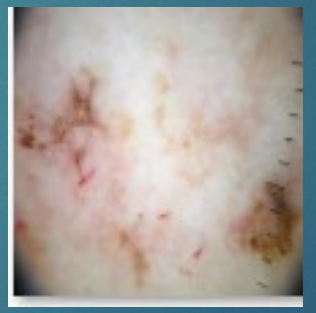
- HAM
- https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000



API#1 皮膚良/惡性判斷 模型建立

- 輸入：
- 模型資料集 人工標註後的資料集。 訓練與測試比率為 8:2
- The dataset is taken from the ISIC (International Skin Image Collaboration) Archive. It consists of 1800 pictures of benign moles and 1497 pictures of malignant classified moles.
- Train (2637張）
  - 良性Benign：1440張
  - 惡性Malignant：1197張
- Test (660張）
  - 良性Benign：360張
  - 惡性Malignant：300張
- Total: 3297
- 計算：CNN卷積神經網路
- 正確率: 約80%

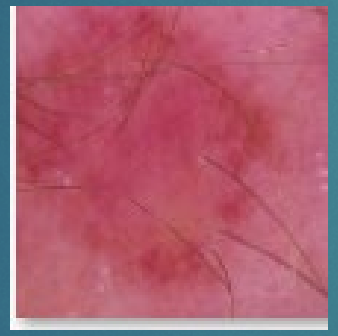




尺寸 224 x 224 RGB 3維彩色圖片

圖片來源：ISIC

API#1 皮膚良/惡性判斷
預計輸入輸出結果

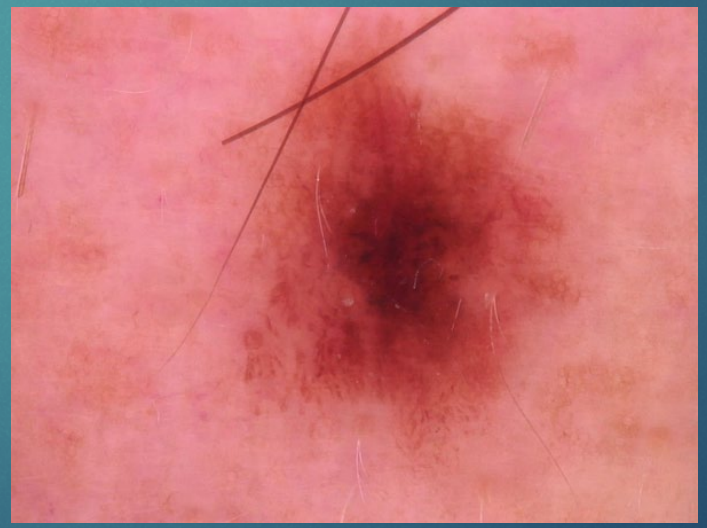
- 輸入：based64編碼的圖片 （手機/PC）尺寸 224*224 彩色圖檔
- 處理：CNN卷積神經網路模型
- 輸出：
- 
  良性
  
  惡性

圖片來源：ISIC

API#2 皮膚疾病判斷模型建立

- 輸入：
- 模型資料集 HAM

  More than 50% of lesions are confirmed through histopathology (histo)
- 訓練與測試比率 8:2
  - Train (80012張）
  - Test (20003張）
  - Total: 10,015張
- 計算：CNN卷積神經網路
- 正確率: 約70%

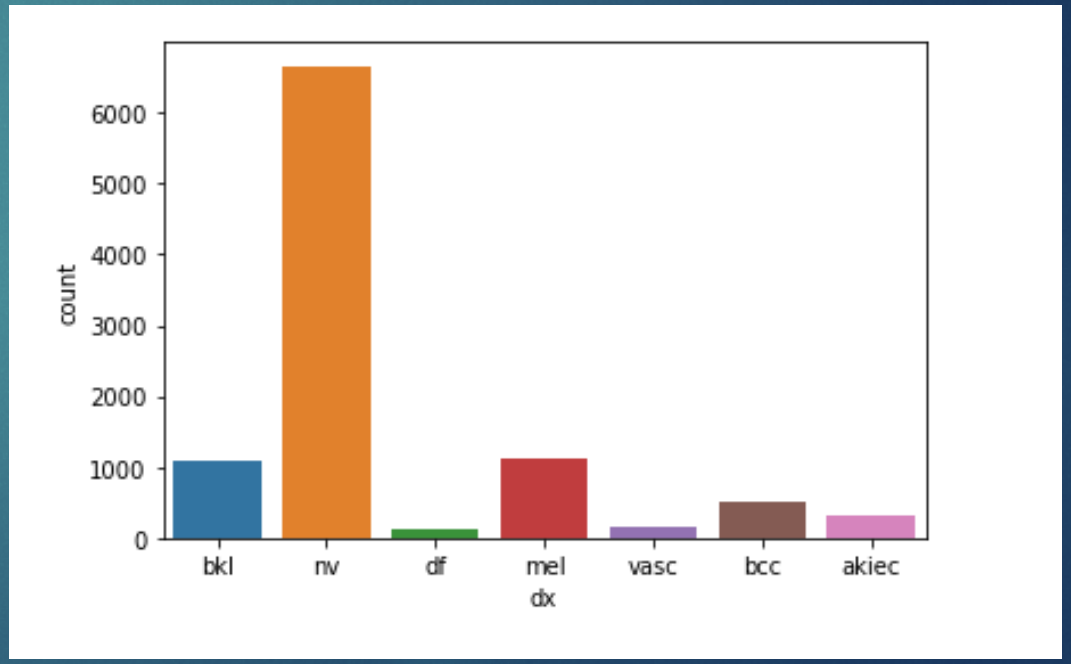


尺寸 224 x 224 or 100 x 100

圖片來源：Ham1000

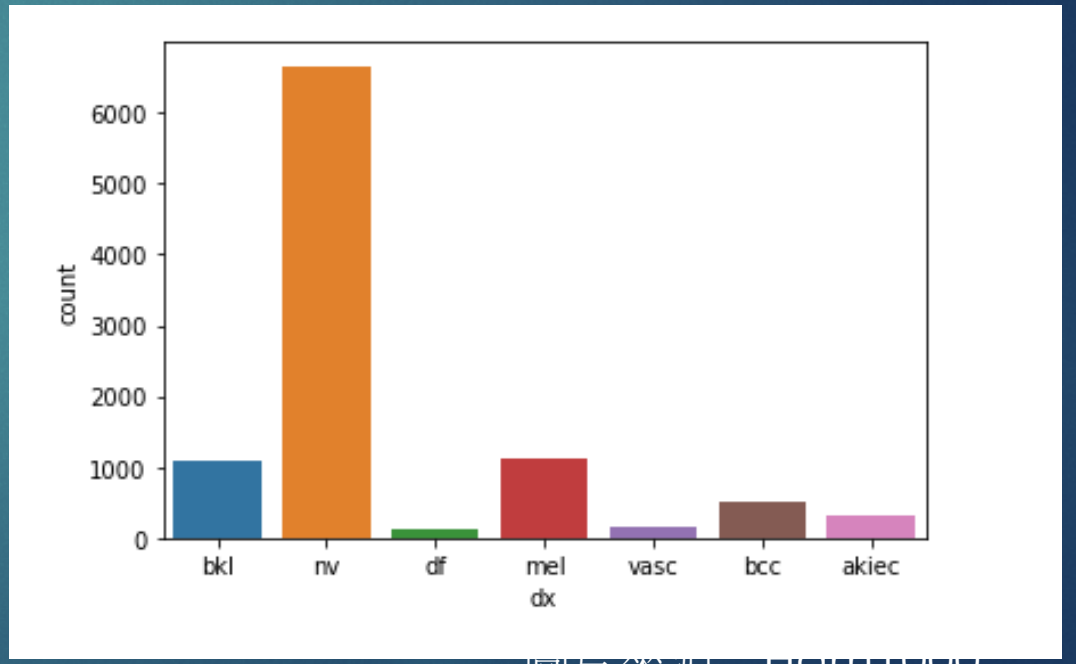
疾病分布比例

- 七種疾病
- 1. Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec),
- Actinic keratoses （鱗狀細胞癌）、
- 2. basal cell carcinoma (bcc), 基底細胞癌
- 3. benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl),
  良性角化病變
- 4. dermatofibroma (df), 皮膚纖維瘤
- 5. melanoma (mel), 黑色素細胞癌/黑色素瘤
- 6. melanocytic nevi (nv) 黑色素細胞痣
- 7. vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc). 血管病變



API#2 皮膚疾病判斷
預計輸入輸出結果

- 輸入：based64編碼的圖片 （手機/PC）尺寸 224*224 （或100 * 100) 彩色圖檔
- 處理：CNN卷積神經網路模型
- 輸出： bkl / nv / df / mel dx / vasc /bcc / akie



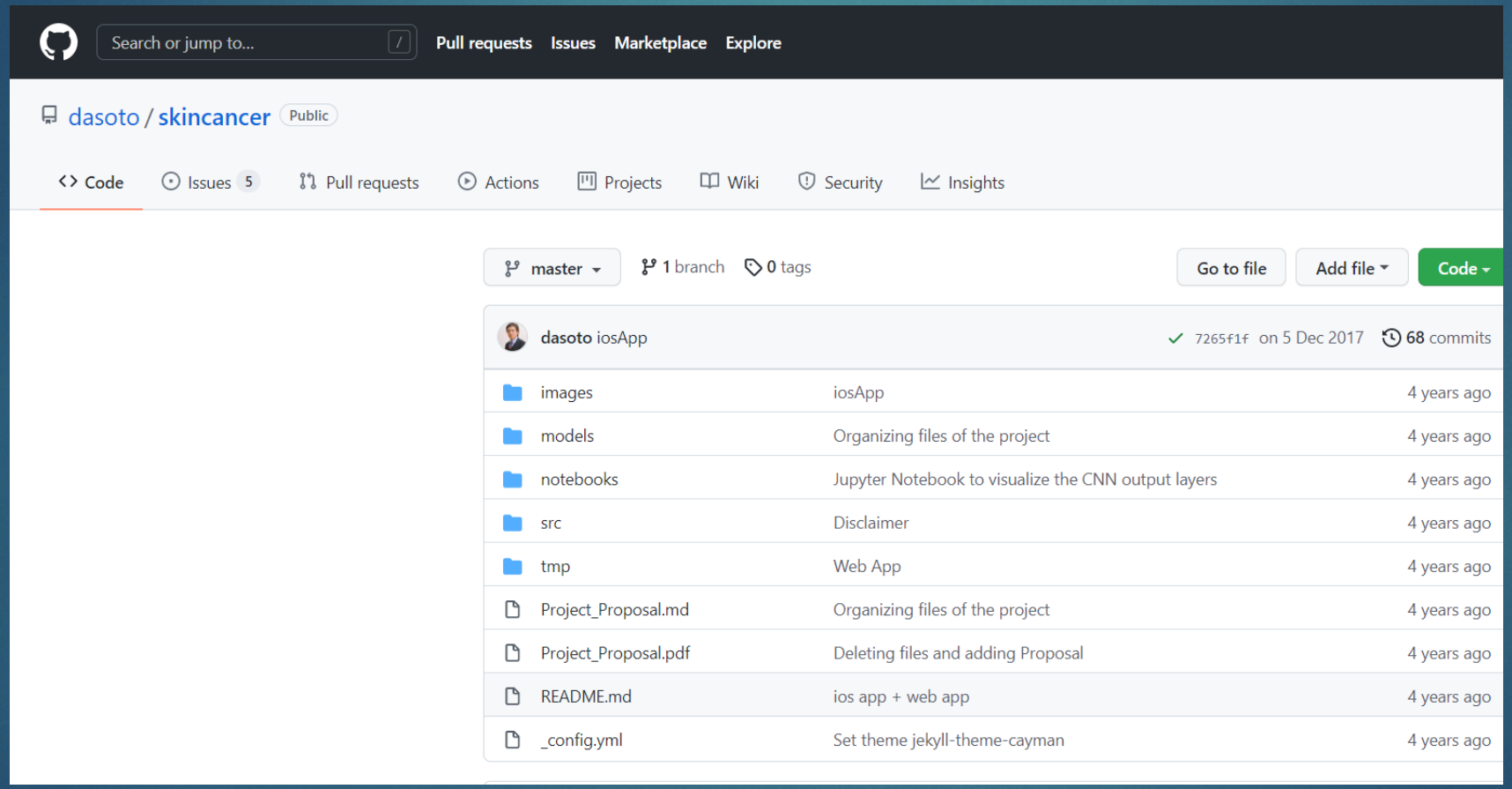
圖片來源：Ham1000

開發環境與工具

- Python
- Anaconda
- Tensorflow
- Keras
- CNN
- Flask
- HTML5

程式資源

- https://github.com/dasoto/skincancer



Prepare data

2. Development process and Data

The idea of this project is to construct a CNN model that can predict the probability that a specific mole can be malign.

2.1 Data:

To train this model the data to use is a set of images from the International Skin Imaging Collaboration: Mellanoma Project ISIC https://isic-archive.com.

The specific datasets to use are:

- ISIC_UDA-2_1: Moles and melanomas. Biopsy-confirmed melanocytic lesions. Both malignant and benign lesions are included.
  - Benign: 23
  - Malign: 37
- ISIC_UDA-1_1 Moles and melanomas. Biopsy-confirmed melanocytic lesions. Both malignant and benign lesions are included.
  - Benign: 398
  - Malign: 159
- ISIC_MSK-2_1: Benign and malignant skin lesions. Biopsy-confirmed melanocytic and non-melanocytic lesions.
  - Benign: 1167 (Not used)
  - Malign: 352
- ISIC_MSK-1_2: Both malignant and benign melanocytic and non-melanocytic lesions. Almost all images confirmed by histopathology. Images not taken with modern digital cameras.
  - Benign: 339
  - Malign: 77
- ISIC_MSK-1_1: Moles and melanomas. Biopsy-confirmed melanocytic lesions, both malignant and benign.

資料處理方式

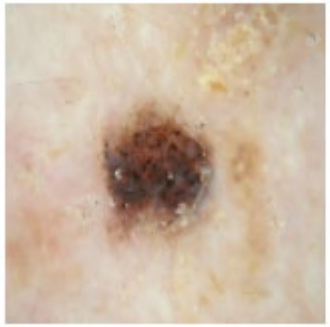
Sample images

As summary the total images to use are:

Benign Images	Malignant Images
1208	849

1. Sample images of benign moles:

- Sample images of malign moles:



Tools

- Tensorflow (GPU High performance computing - NVIDIA)
- keras
- Python
- matplotlib
- scikit-learn
- AWS (EC2 - S3)
- IoS swift + core ML
- Flask

Results

Classification Report CNN From scratch, CV Folder.

- Model_name = models/cnn-scratch-cv.hdf5
- 110 epochs. No early stop.
- AUC: 0.9164

class	precision	recall	f1-score	support
0.0	0.86	0.88	0.87	50
1.0	0.88	0.86	0.87	50
avg / total	0.87	0.87	0.87	100

Second Model: VGG16 + Dense Layer

Classification Report VGG16 + Dense Layer.

- Model_name = models/VGG-Full.hdf5
- 100 epochs. No early stop.
- AUC: 0.9496

class	precision	recall	f1-score	support
0.0	0.87	0.92	0.89	50
1.0	0.91	0.86	0.89	50
avg / total	0.89	0.89	0.89	100

Third Model: CNN + Data Augmentation

Classification Report CNN Scratch with Data Augmentation.

- Model_name = models/CNN-Scratch-DA
- 100 epochs.ModelCheckpoint. Best Val Accuracy
- AUC: 0.9444

class	precision	recall	f1-score	support
0.0	0.81	0.96	0.88	50
1.0	0.95	0.78	0.86	50
avg / total	0.88	0.87	0.87	100

Fourth Model: VGG16 + Dense Layer + Data Augmentation

Classification Report VGG16 with Data Augmentation.

- Model_name = models/BM_VA_VGG_FULL_DA.hdf5
- 100 epochs.ModelCheckpoint. Best Val Accuracy
- AUC: 0.9612

class	precision	recall	f1-score	support
0.0	0.88	0.88	0.88	50
1.0	0.88	0.88	0.88	50
avg / total	0.88	0.88	0.88	100

CNN Architecture:

報價

AI 皮膚偵測系統報價	
功能	
AI 系統： ▶ API#1 皮膚良惡性分析 ▶ API＃2 皮膚病種類分析 7 種 ▶ 具有再增加新的圖片訓練功能 ▶ 具有圖片裁切功能 ▶ 移植 GCP 雲端主機 ▶ 製作成 Web Services。 AI 網頁後台系統： ▶ AI 醫師標註網頁系統，供人工（醫師）標註圖片，以提高 AI 判斷正確性。	
	Total：

驗收

- 完成兩個AI皮膚病檢測程式
- 製作成Web Services方便串接
- 製作一個網頁後台，提供後續人工標註。
- 程式移植放置GCP雲端主機（主機由衆匯提供）
- 網站後台系統與資料庫放置GCP雲端主機（主機同上）

- 一次性收費。程式保固一年。若有新增功能與重新訓練模型再另行收費。
- 競業條款2年
- 預估工作時間： 3個月