



# 東華AI通訊報

雙月刊

發行 | 人工智慧辦公室  
連絡電話 | (03)890-3792

發行人 | 林信鋒  
傳真號碼 | (03)890-0225

主編 | 江政欽  
網址 | <https://aioffice.ndhu.edu.tw>

## 【AI 科技萬花筒】

達人分享：「數據科學分析的土壤：數據品質評估」 ----1

「AI 和電腦視覺如何成為農業最強收割機？農業機器人一秒辨識雜草和葛苳！」 -----3

「1秒辨識車牌！嘉義市警車升級「嘉e智能巡邏車」上路2個月破案29件」 -----3

「刷牙好懶？未來可以靠「奈米機器人」清潔！連漱口和牙線都到位」 -----3

「AI 繪圖「Midjourney」作品參加美術展獲獎，引發藝術家不滿」 -----3

## 【AI 元宇亮東華】

亮點研發：「植物病蟲害辨識技術在智慧農業的應用」 ----4

## 【AI 小學堂開講】

《Teachable Machine 無程式碼機器學習工具》 ----5

## 【AI 跨域徵人帖】

《史上最大缺工 101.7 萬人！日月光人資主管：退役軍人、運動員都是戰力》 -----8

《PropTech 數位人才招募》 -----8

《2022 AppWorks x Yourator 科技新創聯合大徵才》 --8

## 【AI 科技萬花筒】—達人分享】

### 《數據科學分析的土壤：數據品質評估》

<sup>1</sup> 國立臺灣大學資訊管理系 李家岩教授

<sup>2</sup> 國立成功大學製造資訊與系統研究所

楊舒惠<sup>2</sup>、歐子毓<sup>2</sup>、盧宣文<sup>2</sup>

人工智慧與大數據分析是現今的主流之一，希望透過足夠多的數據和完善的資訊基礎設施，提供分析人員建構出先進的人工智慧或深度學習模型，為日常生活帶來許多便捷的應用。例如購物網站的推薦系統、疫情時代下所誕生的人臉口罩辨識系統等。然而，並非所有實證場域都能搜集到足夠多的數據來訓練模型，也並非所有的裝置都能裝設高成本的 GPU (Graphics Processing Unit) 提供運算資源。因此，如何在有限的數據與計算資源中，能讓模型從歷史數據裡學習資料間的關聯性，在這過程，數據品質扮演關鍵角色。

俗語說「垃圾進，垃圾出」(garbage in garbage out)。若使用品質差的數據，除了可能會誤導訓練過程，還可能使得訓練時間延長並導致預測不準確。事實上，在大多數資料科學家的經驗中，數據準備、清理和轉換佔據了模型建構的大部分時間。近年來，各家企業導入自動化機器學習(AutoML) 相關技術，讓資料科學家、分析師在建構模型時可以減少反覆性的工作流程。雖然不需像以往花費大量的時間在重新訓練並比較不同模型間的表現，但相對取而代之的是將時間投入於**廣泛的數據品質檢查**。

儘管針對數據品質所執行的檢查可能非常主觀，但從過去文獻中可以整理出評估數據品質的一些關鍵因素與衡量方法。一般來說，數據品質的評量是對數據可用程度以利決策支援和適合其服務環境的評估。現有許多要素能用來衡量數據品質，包含**數據一致與格式、數據來源與時效性、數據準確與完善程度、與資訊安全** ([1], [2])。

如圖 1 所示，分述如下。

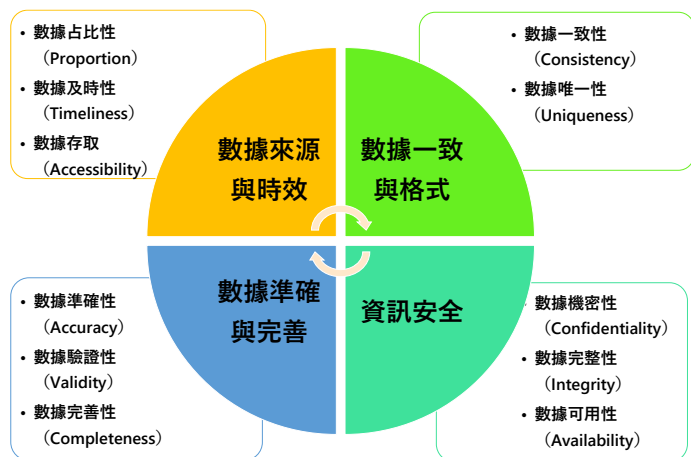


圖 1 數據品質的評量構面

在數據一致與格式的方面，說明數據格式主要著重在資料庫的儲存格式，如文字型(character)或數值型(integer)，甚至是關聯式資料庫中定義的正規化(normalization)，主要用來消除資料儲存的異常，例如第一正規化(1NF)主要用來確定主鍵(primary key)一辨識每筆資料的關鍵欄位；第二正規化(2NF)則是用來消除重複欄位，其餘正規化也皆有相對應功能，若有需要可以自行查閱。唯一性的部分主要是針對不同資料庫串接的過程，確保資料不會重複。

### 數據一致與格式

- 數據一致性 (Consistency)：數據不可違反語義規則的定義或資料庫資料型態的定義。
- 數據唯一性 (Uniqueness)：對於特定字串、記錄或數據集，系統內部或跨系統間不得有重複的存在。

數據來源與時效性的面向說明了樣本對於母體的百分比，以評估資料特性是否足以反映母體、資料蒐集的時間、範圍甚至容易程度來進行評估。

### 數據來源與時效

- 數據占比性 (Proportion)：收集數據樣本相對於母體所佔的百分比。
- 數據及時性 (Timeliness)：數據的年齡或收集時間是否適用於分析。
- 數據存取性 (Accessibility)：是否使用者有適當權限存取數據且容易可快速檢索。

數據準確性及完善程度也是評估數據品質很重要的一部份，如何確保所蒐集的資料足夠代表問題，或是針對數據進行

驗證確認準確性 (Accuracy) 及完善性 (Completeness) 的清楚定義。定義準確性可為變數之間的相關係數或可解釋迴歸模型的係數是否為同向，以驗證資料與物理背景知識的相符，而後者則是參考論文利用隨機新增欄位的方式確認資料的完善程度。

### 數據準確及完善

- 數據準確性 (Accuracy)：從資料庫中收集到的數據是否是正確、可靠且驗證過，能夠真實地反應實際狀況。
- 數據驗證性 (Validity)：對於輸入資料正確性的檢查工作，目的在確定資料處理時所使用資料正確無誤且符合需求。
- 數據完善性 (Completeness)：表達所收集數據其深度與廣度對於特定問題提供資訊量的完整度。

關於資訊安全的部分，由於現在數據的取得比較容易，如何能確保資料在安全的環境下存取及使用變得十分重要。機密性、完整型及可用性是資訊安全的鐵三角(CIA Triad)，任何違反這些指標的行為都可能會減低資安的防護強度，甚至對公司重要資產或機密資料造成威脅。簡單來說機密性是確保未經授權的人無法查閱資訊；完整性則是資訊無法輕易被竄改或刪除；最後可用性的部分則是定義不同使用者可以使用不同的資料內容。

### 資訊安全

- 數據機密性 (Confidentiality)：確保資料傳遞與儲存的隱密性，避免未經授權的使用者有意或無意的揭露資料內容。
- 數據完整性 (Integrity)：在傳輸、儲存資訊或資料的過程中，確保資訊或資料不被未授權的篡改或在篡改後能夠被迅速發現。
- 數據可用性 (Availability)：當使用者需透過資訊系統進行操作時，資料與服務須保持可用狀況並能滿足使用需求。

在不同的產業中，對於數據品質的評估指標與項目會隨著不同的需求而有所調整。以國際貿易資料為例，聯合國商品貿易統計數據庫 (UN Comtrade) 是進行實物貿易分析的原始數據來源之一，也是最廣泛被使用的數據來源。然而，若資料庫中的數據產生差異且未及時進行檢視與處理，可能會導致完全矛盾的結論。因此在進行數據品質評估時，異常值與缺失值的



偵測與填補會是考量之一。而在製造業中，預測產品品質或應用虛擬量測時，數據的即時性可能會是除了數據準確性之外的另一個評估重點。零組件的剩餘壽命估計與保養排程建立則與數據占比性與完善性有關，當沒有完整的零組件壽命週期資料，對於剩餘壽命的估計很容易失真。正如同夏蟲不可語冰，不同的健康狀態下是急速衰退或是穩定衰退對於生產調度與保養策略的制定影響甚大。

良好的數據品質對於任何數據分析是不可缺少的一環，在過去文獻中存在多種數據品質的評估方法，但真正導入工廠或組織中卻有一定的難度。儘管我們知道數據不正確和延遲的數據搜集會導致不精確的預測，進而影響生產計劃和機台控制，但數據評估方法的複雜度及文獻中評估指標多元，是數據品質評估難以落地的兩大主因。

綜合上述，對於數據品質機制設計上我們提供另外兩個面向—**建立存取機制與培訓人員、與實施數據質量解決方案**。前者主要說明盡量減少組織內人為活動導致的數據品質問題，應設計系統存取與權限等機制或審核流程。此外，所有開發人員和資料庫管理員都必須對業務流程有很好的詮釋與認知，並且在開發和設計資料庫和應用程序有統一且一致的參考模式。後者主要說明開發數據品質分析工具或應用程序與建立評估準則，使數據有基礎品質，接著針對資料在不同指標的不足之處進行改善，循序漸進，使整體數據品質提升。

#### 參考文獻

- [1]製造數位轉型學院，數據品質 Data Quality，2022。  
[https://www.facebook.com/permalink.php?story\\_fbid=132736309422240&id=103233739039164](https://www.facebook.com/permalink.php?story_fbid=132736309422240&id=103233739039164)
- [2]Fadlallah, H., "An introduction to data quality," TowardsDataScience, 2018.  
<https://towardsdatascience.com/an-introduction-to-data-quality-951cc6fe0274>

### 【AI 科技萬花筒】-AI 科技動態

「AI 和電腦視覺如何成為農業最強收割機？農業機器人一秒辨識雜草和萵苣！」

<https://pse.is/4g4ajr>

隨著科技快速進步，不只製造業工廠講求全自動化，這樣的技術對於農業來說也是相當重要。機器人市場已經蓬勃發展了一段時間，預計到 2026 年將達到 740 億美元，且其中一部分成長是因為農業機器人的需求增加。(邵元婷, TechOrange, 2022/08/23)

「1 秒辨識車牌！嘉義市警車升級「嘉 e 智能巡邏車」上路 2 個月破案 29 件」

<https://pse.is/4frbde>

嘉義市警察局今年在 12 個派出所及保安隊、交通隊 14 輛巡邏車上，裝載 AI 車辨系統打造「嘉 e 智能巡邏車」，可讓警車在時速 40 公里以上，1 秒自動辨識車牌號碼，白天辨識率達 95%，夜間也有紅外線辨識功能，有別以往查詢失竊車輛，員警須用 M-Police 行動電腦一筆一筆輸入，快速又有效率，成軍上路 2 個月，已經破獲失竊車輛(牌)、竊盜、酒駕、侵占等 29 件案件，好用程度讓員警驚呼「真是太神奇！」(王善嫻，自由時報，2022/08/26)

「刷牙好懶？未來可以靠「奈米機器人」清潔！連漱口和牙線都到位」

<https://pse.is/4ej8ue>

近期，美國研究團隊開發出一種不需要自己動手刷牙的新科技—奈米機器人，能一次執行刷牙、漱口和牙線的三種功能，為口腔護理保健帶來新進展。(傅珮晴、錢玉紘，數位時代，2022/07/29)

「AI 繪圖「Midjourney」作品參加美術展獲獎，引發藝術家不滿」

<https://pse.is/4qlzqt>

先前曾在網路引發話題的 AI 繪圖工具「Midjourney」，成為了美國紐約時報與泰晤士報的焦點，因為一名男子使用「Midjourney」人工智能創作的繪圖參加科羅拉多州博覽會舉辦的「Fine Arts Exhibition」美術展獲獎，引發了藝術圈的嘩然。(歪力，4GAMERS，2022/09/03)

## 【AI 元宇亮東華】- 亮點研究

## 《植物病蟲害辨識技術在智慧農業的應用》

資訊工程學系 張意政教授

根據行政院農委會的「植物疫情監測與通報系統之現況與展望」[1]，台灣地處亞熱帶地區，氣候高溫多濕，地形複雜多變，農作物產區涵蓋高山溫帶氣候至熱帶雨林氣候，加上農作物複種指數高，各種疫病蟲害極易繁衍為害。當農作物受到病蟲害感染時，除了影響到果實的品質及收成外，也會波及到經濟發展與人口成長等問題。農藥的使用是病蟲害防治中常用的方法；然而，依據「環境資訊中心」2020 年的報導[2]：農地長期施用化學農藥不僅對農民身體健康造成風險、潛藏食安危機，農田生態系也受到化學製劑影響，不利環境和農業永續發展。為推廣友善農業、減少農產品安全風險，農委會 2017 年宣布「農藥 10 年減半」政策，預計要在 2027 年達成農藥減半的目標。台灣近年來因經濟發展以科技產業為主，農耕土地面積大幅降低，但農藥銷售量卻反增，根據調查，農地施用農藥量至 2018 年台灣每公頃農藥用量已是全球第一名。如何有效控制農藥用量，進而保護環境與食安是一個重要的課題。

花蓮瑞穗、台南麻豆與雲林斗六是台灣文旦三大產區，而位於花蓮縣瑞穗鄉東側的鶴岡村，地處大片山坡地，擁有東部地區無汙染的環境與土壤，因此孕育出香甜且味美的「鶴岡文旦」。文旦的成長容易受病蟲害影響，而花蓮農友在對治病蟲害時的農藥選擇常依從農藥販賣商的建議，而販賣商所銷售的農藥通常會將多種不同農藥混成的產品銷售給農友，因此容易產生農藥過度使用的問題。近年來不少年輕人回返家園投入農業發展，這些青農因經驗不足無法正確地辨識植物所感染之病蟲害，此外，某些病蟲害在葉片上的病徵是相似的，也會讓農民產生辨識上的困擾。

過去主要的作物病蟲害檢測方式是透過肉眼辨識並判斷，這需要一定的農業經驗才能夠做出正確的判斷，而在瞭解病蟲害後也需要有相關藥物的知識才能夠選擇合適的農藥進行噴灑。為將人工智慧科技運用於柚樹病蟲害防治及農藥精準施用，東華資工透過執行科技部「東台灣數位綠色經濟提升之混合實境創新技術與應用計畫」與花蓮縣「地方產業創新研發推動計畫」（地方型 SBIR 計畫）為在地柚農成功研製一個以深度學習（Deep Learning）技術為基礎的雲端文旦病蟲害辨識

與農藥推薦系統，藉由此系統，即使經驗不足的農友也隨時可利用手機拍攝柚葉以識別柚樹遭受何種病蟲害並及時進行適當防治。此外，系統還能提供各種病蟲害的農藥訊息及用藥方式建議，引導農友精確選藥與施藥，進而有效降低農藥過度使用的問題。

近年來，植物病蟲害辨識領域開始採用深度學習技術，因為要收集足夠的數據並不容易，訓練數據集的建構是應用深度學習技術的重要問題之一。比較著名的公共數據集 PlantVillage 乃透過網路收集並整理圖像數據而建立。PlantVillage 資料庫包含 54,306 張圖片，包含 19 種作物的 38 種健康及受害葉片類別，很多研究的分析以此資料庫作為測試標準。2021 年在知名數據競賽網站 kaggle 上發布了一個新的公開資料集 Plant Pathology 2021，該資料集包含 19 種蘋果的不同疾病共 23249 張圖片，也是一個數據相對較多的資料庫。此外，一些研究因個別的應用需求多半會構建自己的數據集。因為柚子的病蟲害症狀一般來說會顯示在葉子上，因此針對花蓮瑞穗文旦，東華資工系研究團隊費時一年蒐集了一套遭受病蟲的文旦葉片的資料庫作為訓練資料庫來訓練人工智慧系統辨識文旦常見的病蟲害。

我們所建立的文旦葉片病蟲害辨識及農藥推薦系統（圖 2），可分為四個部分進行建構，分別為：資料庫建置、辨識系統建置、客戶端介面建置與農藥資料庫建置。

農友在使用本系統時，可在農園現場以手機拍攝文旦葉片或上傳已拍攝之相片讓雲端的人工智慧病蟲害辨識系統進行識別，圖 3 為手機上呈現的前端介面。辨識系統在識別出病蟲害的種類後會立刻在手機上顯示識別結果，同時也提供該病蟲害相關背景知識並建議適當的用藥及施用方式，農友可再予以確認後根據建議作為選擇農藥和引導施藥的參考。

資料庫中包含五種類別的文旦葉片：健康及四種常見的病蟲害分別是黑點病、缺鎂、油胞病和潛葉蛾。被黑點病感染的植株，其果實外觀可觀察到細小黑點的分布，而葉片上也有近似的黑點分布；缺鎂的文旦葉片上會呈現不同大小的黃綠色區塊分布，文旦在缺鎂的狀況下會引起葉片葉綠素的缺乏，進而降低光合作用效果，而長期缺鎂的植株也會因抵抗力下降，增加受到其他病蟲害感染的風險。



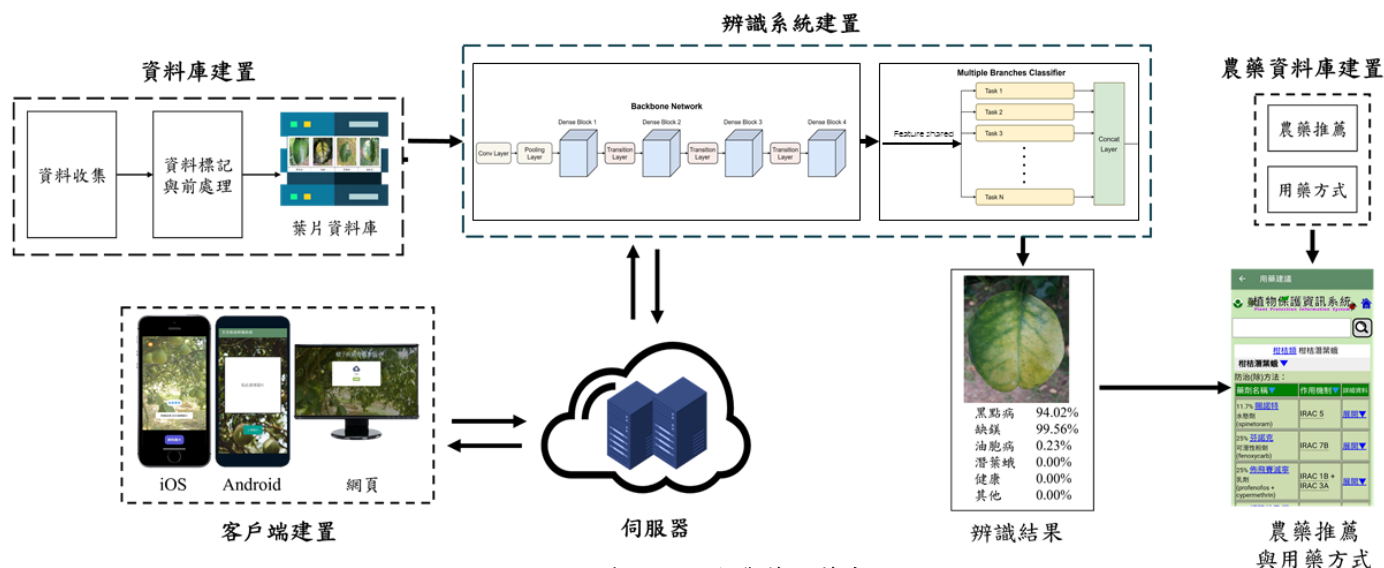


圖 2 病蟲害辨識與農藥推薦系統



圖 3 前端介面在手機上之呈現畫面

文旦受到油胞病感染後，其果實外表或葉片上會產生不規則的黑色斑點，而斑點的周圍通常也伴隨黃褐色的病斑。潛葉蛾感染的植株其葉片上會呈現蜿蜒曲折之隧道圖案，而其受感染新葉常呈現捲縮樣態，無法伸展的葉片也容易成為不同蟲類的寄生場所，如粉介殼蟲、葉蟬等，而受到潛葉蛾感染的植株受其他病害侵害的機率相對提高。圖 4 顯示病蟲害葉片病徵與健康葉片。



圖 4 文旦病蟲害在葉片上呈現之病徵

本系統的辨識系統核心選用深度較深的網路模型來建置，在影像辨識技術中，較淺的網路模型並不適合辨識過於複雜的圖片。經過實驗測試後，我們選擇以 Densely Connected Convolutional Networks (DenseNet)來架構本系統的深度學習

網路。DenseNet 是一個密集連接的深度學習網路模型，通過特徵的重複使用來得到網路的隱含信息，其組成架構是由多個 Dense block 與 Transition layer 連接而組成，每個 Dense block 中又包含多個 conv 層，其中 conv 層為一個 Convolution + batch normalization + relu 的三層式組合。DenseNet 中每個 conv 層都能得到前面層參數計算後的結果，因此訓練時能夠保持前面層的資訊，也能夠降低梯度消失問題的影響。除此之外，因為同一個植株可能會罹患多種病蟲害，我們也提出一個多標籤的辨識架構，來處理同一葉片呈現多重病徵的辨識問題。

病蟲害的防制是農業發展上重要課題之一，我們以深度學習網路為核心建構一個病蟲害辨識與農藥推薦系統，系統在輸出所辨識病蟲害種類的同時，也提供農友病蟲害的知識，且進一步連結農藥資料庫，提供治療病蟲害所需要的農藥種類與其使用方式。系統提供正確的用藥知識與種類可讓農友精準用藥，避免農藥過度使用，在達到降低農損、提高獲益同時，更可兼顧自然環境與食品安全保護議題。

#### 參考文獻

- [1] 農業委員會資訊：  
<https://www.coa.gov.tw/ws.php?id=1985>
- [2] 環境資訊中心資訊：  
<https://e-info.org.tw/node/228366>

#### 【AI 小學堂開講】

#### 《Teachable Machine 無程式碼機器學習工具》

通識中心 陳文盛助理教授

第一期中我們稍微介紹到 Teachable Machine，它是由 Google 推出的網頁工具，可讓所有人都能快速建立機器學

【版權所有，未經同意請勿轉載】

習模型，所以本期再進一步跟大家介紹此工具的實際使用，希望讓更多讀者能更真實體驗到即使沒有程式設計的專業背景者也能自己建構頗具智慧的系統。

如圖 5 所示，Teachable Machine 在 首頁中 (<https://teachablemachine.withgoogle.com/>) 清楚標示“輕鬆快速地建立機器學習模型，以使用於網站、應用程式和其他地方，不需要編寫程式或具備專業知識”，利用這個工具，初學者學習機器學習的難度與門檻可以大大降低。Teachable Machine 提供三種不同深度學習模式：圖片辨識（圖 6）、聲音辨識（圖 7）、人體姿勢追蹤（圖 8）。



圖 5 Teachable Machine 首頁

Teachable Machine 深度學習模式-圖片辨識結果如圖 6。

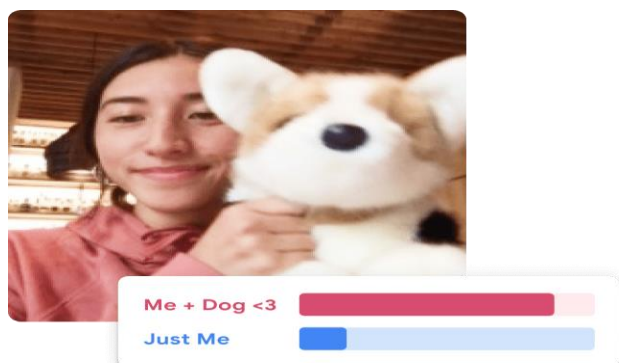


圖 6 Teachable Machine 圖片辨識結果

Teachable Machine 深度學習模式-聲音辨識結果如圖 7。



圖 7 Teachable Machine 聲音辨識結果

Teachable Machine 深度學習模式-姿勢辨識結果如圖 8。

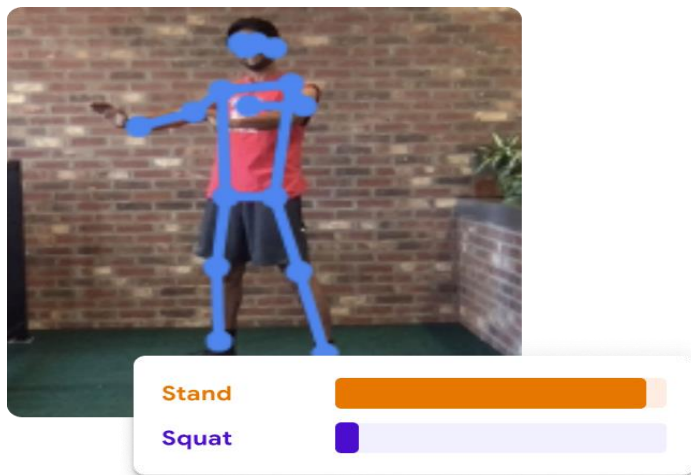


圖 8 Teachable Machine 姿勢辨識結果

Teachable Machine 的辨識流程分為下列三步驟：

### 1. 收集範例

收集範例，並將範例分為你想讓電腦學會的類別。

### 2. 訓練模型

訓練模型並立即進行測試，看看它是否能將新範例正確分類。

### 3. 預覽訓練結果和匯出模型

預覽可以測試模型辨識的正確性，匯出模型可為你的網頁、應用程式專案匯出模型，可以下載模型或線上使用。

下面將帶領大家使用 Teachable Machine 作圖片辨識。

步驟 1 進入 Teachable Machine，點選「開始使用」（圖 9）。



圖 9 開始使用 Teachable Machine 畫面

步驟 2 在「新增專案」頁面中點選「圖片專案」(圖 10)。



圖 10 新增圖片辨識專案

步驟 3 再點選「標準圖像模型」，進入模型訓練流程(圖 11)。



圖 11 選擇標準圖像模型

在訓練圖像模型的流程中共有三個階段：

- ❶ 上傳你想讓電腦學會的類別內容，可以透過攝影鏡頭或上傳的方式增加圖片，至少需要準備好兩個類別的圖片(圖 12)，分類建立完成後，點選「訓練模型」。
- ❷ 點選訓練模型，就會開始進行圖片模型的訓練。
- ❸ 預覽訓練結果和匯出模型，出現「模型已訓練完成」的文字表示訓練完成。



圖 12 準備分別上傳待辨識的各類別物品圖片

以辨識水仙花和向日葵為例，我們可於 google 瀏覽器上找尋花卉資料庫，找到牛津大學工程科學系 Andrew Zisserman 和 Maria-Elena Nilsback 兩人所提供的 17 種花卉資料集 (<https://www.robots.ox.ac.uk/~vgg/data/flowers/index.html>)，我們使用其中的水仙花和向日葵圖片作為展示。

步驟 4 上傳我們所下載的第一種類(水仙花)圖片樣本，點選「Class1」修改圖片樣本名稱，再點選「上傳」(圖 13)。然後出現如圖 14 的畫面中選擇「從檔案中選擇圖片或是將圖片拖曳到這裡」選擇所需要的圖片樣本上傳。完成 Class1 圖片上傳後依同樣方式再點選「Class2」進行第二種類(向日葵)圖片的上傳，兩種花卉圖片樣本上傳完成後如圖 15 所示。

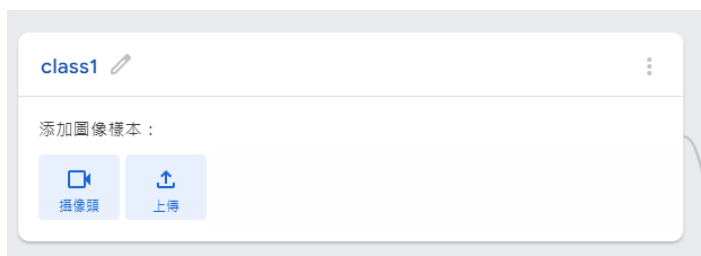


圖 13 上傳第一種類(水仙花)圖片

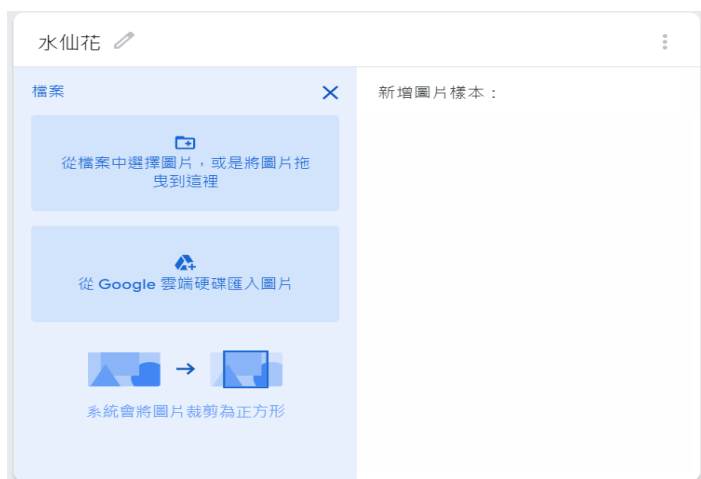


圖 14 從電腦的檔案中選擇各類別圖片資料集上傳



圖 15 完成各類別圖片資料集的上傳



步驟 5 點選「訓練」中的「訓練模型」，Teachable Machine 就會開始訓練模型(圖 16)。

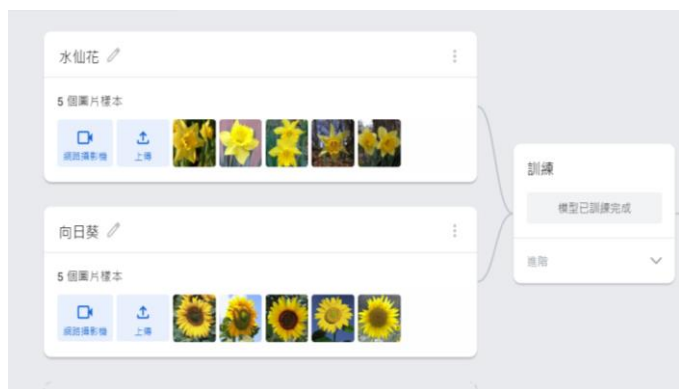


圖 16 開始訓練模型

步驟 6 訓練完成後，點選預覽中的輸入開啟以「檔案」預覽測試結果(圖 17)，再用「從檔案中選擇圖片或是將圖片拖曳到這裡」上傳測試圖片樣本，測試模型辨識的正確性(圖 18)。



圖 17 完成模型的訓練可以匯出模型並預覽測試結果

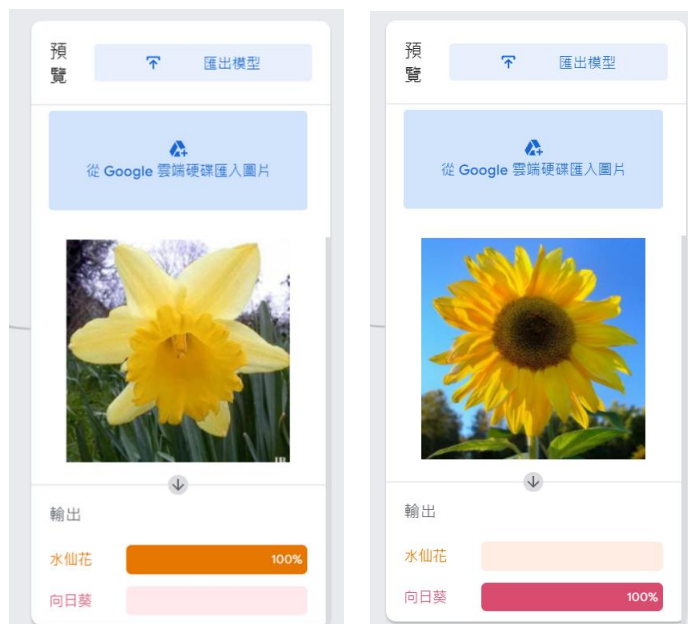


圖 18 分別以水仙花和向日葵圖片測試模型得到正確的辨識結果

以上就是 Teachable Machine 作圖片辨識的操作流程，我們目前只有預覽模型辨識的正確性，進一步可下載 Teachable Machine 所訓練模型應用於網頁或應用程式中，即可建立智慧系統供使用者使用。

### 【AI 跨域徵人帖】

【史上最大缺工 101.7 萬人！日月光人資主管：退役軍人、運動員都是戰力】

<https://pse.is/4gac33>

104 人力銀行 9 月全站工作機會數 101.7 萬個，其中電子資訊、軟體、半導體人才缺口最大，日月光人力主管分享，當百萬缺工潮已屬長期現象，企業嘗試以 AI 自動化解決缺工問題，AI 工作機會 5 年成長 72%。

【PropTech 數位人才招聘】

<https://hr.sinyi.com.tw/proptech>

**PropTech**  
**數位科技 人才招聘**

強力徵求 **熱愛數位科技 x 居住生活應用** 的一群人

【2022 AppWorks x Yourator 科技新創聯合大徵才】

<https://appworks.yourator.co/>



Yourator 希望打造數位工作者的職涯基地，提供多元化工作機會與職涯內容，在資訊紊亂的大時代，引導你走出璀璨未來。

為此，Yourator 致力成為企業與人才間溝通媒合的橋樑，全力進化服務，幫助企業找到融入團隊的人才，幫助人才找到創造正向價值的工作。