**Plant Pathology 2020 – FGVC7**

**蘋果葉病害分類**

**團隊編號：DL\_Team10**

**組員：林子軒**

# 目錄

**1. 引言**

**2. 相關研究**

**3. 資料集**

**4. 方法**

**5. 實驗結果**

**6. 討論**

**7. 結論**

# 1　引言

全球每年約有 20–40 %的糧食作物產量損失與病蟲害有關，其中植物葉部病害更是影響產量與品質的主要因素之一。根據聯合國糧農組織（FAO）統計，若能將病害損失降低 1%，便足以養活數百萬人口。過去農業生產對病蟲害偵測多仰賴農民或專家目視判斷，然而人工巡田不僅耗時、成本高，也容易因經驗差異產生誤判。隨著智慧農業興起，結合影像感測與深度學習的自動化診斷技術被視為解決此痛點的關鍵方案。

Plant Pathology 2020 – FGVC7 競賽（Kaggle 與 CVPR FGVC Workshop 合辦）公開 1 ,821 張蘋果葉 RGB 高解析影像，涵蓋四種標籤：

| **類別** | **中文對應** | **佔比** |
| --- | --- | --- |
| healthy | 健康 | 53 % |
| multiple\_diseases | 複合病害 | 2 % |
| rust | 銹病 | 21 % |
| scab | 痂病 | 24 % |

資料集中 類別不平衡（healthy 遠多於 multiple diseases），加上影像拍攝條件多樣（戶外光照、不同背景雜訊），使得模型需同時具備強大表徵能力與抗雜訊能力。本專案針對上述挑戰提出以下技術要點：

多架構集成：選用 EfficientNet-B3、ResNet-50、DenseNet-201 與 ConvNeXt-Base 四種 CNN，兼顧參數效率與表現上限。

資料增強策略：導入 ImageNet AutoAugment（顏色抖動、裁切、平移）並額外加入隨機水平翻轉與 ±15° 隨機旋轉，以模擬現地拍攝變異。

Focal Loss 應對類別不平衡：設定 α = 1、γ = 2，對少數類別賦予更高權重，提高 multiple diseases 的召回率。

五折交叉驗證：將有限資料最大化利用，並評估模型在未知樣本上的泛化能力。

Test-Time Augmentation (TTA) + Ensembling：推論階段對單張影像產生多種視角，再取四模型 × 五折平均，以降低單模型偏誤並提升穩定度。

本系統最終在公開排行榜取得 F1-score = 0.977，相較於官方 EfficientNet-B3 baseline（0.959）提升近 1.8 個百分點。更重要的是，透過交叉驗證與 TTA，模型在私榜亦維持一致表現，證明方法具備良好泛化能力。

# 2　相關研究

自 2012 年 AlexNet 在 ImageNet 競賽奪冠以來，深度卷積神經網路（CNN）已成為植物病害影像辨識的主流。相關研究大致可分為四個階段：

| **時期** | **代表方法 / 里程碑** | **關鍵貢獻** | **侷限** |
| --- | --- | --- | --- |
| **早期CNN**  **(2015-2017)** | Mohanty *et al.* (2016)：AlexNet / GoogLeNetSladojevic (2016)：自建 13-layer CNN | 首次證明深度 CNN 能在 *PlantVillage* 資料集達到 **> 99 %** 準確率 | (1) 影像多以人工光箱拍攝，背景單純；(2) 多為單品種葉片，泛化性不足 |
| **遷移學習 & 精簡網路**  **(2018-2019)** | Too *et al.* (2018)：ResNet-50 fine-tuneHughes *et al.* (2019)：MobileNetV2 手機推論 | 利用 ImageNet 權重快速收斂；提出在樹莓派/手機端即時推論的可行性 | 對類別不平衡較敏感；需大量標註資料 |
| **規模化 CNN**  **(2020-2022)** | Tan & Le (2019)：EfficientNet-B7 參數效率極高Liu *et al.* (2022)：ConvNeXt 結合 Conv 與 ViT 設計理念 | 透過 compound scaling 與大規模預訓練，進一步提高準確率；在 CVPR FGVC7 / FGVC8 挑戰賽屢奪冠 | 模型仍較重，部署於邊緣裝置需剪枝或量化 |
| **Transformer 與自監督**  **(2021-至今)** | Dosovitskiy *et al.* (2021)：ViT 首次純 Transformer 用於影像Chen *et al.* (2022)：Swin-T 少量資料亦可 fine-tuneBaral (2023)：MaskCLIP 自監督預訓練 | 長距依賴建模＋多頭注意力可捕捉細微紋理；自監督學習減少標註成本 | ViT 對資料量敏感，需大量遷移學習或資料增強 |

**影像資料集與評測指標**

PlantVillage：54 k 張室內拍攝影像，28 作物 × 38 病害，多數研究以此作 pre-training。

PlantDoc：2.6 k 張田間自然場景，背景複雜，可評估模型在實地環境之魯棒性。

FGVC7 Apple Leaf：本研究使用的官方資料集，僅 1,821 張，高解析度且類別嚴重不平衡。

評測指標由初期的 Accuracy 轉向更能衡量不平衡資料的 Macro F1-score 與 AUPRC。

**類別不平衡處理**

重取樣：隨機過取樣 (ROS) 與欠取樣 (RUS)。

成本敏感學習：Focal Loss、Class-Balanced Loss (Cui 2019)。

資料增強：MixUp、CutMix 及 AutoAugment 改善少數類別多樣性。

**模型輕量化與部署**

MobileNetV3 / EfficientNet-Lite 在 ARM 裝置達 30 FPS 以上；

模型剪枝：Li et al. (2020) 將 ResNet-50 剪至 40 % FLOPs 幾乎無精度損失；

量化：Post-training INT8 量化可進一步縮小模型大小達 4×。

**尚待突破的挑戰**

跨品種、跨地區泛化：葉面外觀因品種和氣候差異大，需更大規模且多樣化的訓練資料或自適應機制。

早期症狀偵測：病害初期斑點極小且顏色變異不明顯，需高解析 FPN 或超分解析輔助。

多模態融合：結合光譜成像、溫濕度與地理資訊，可望提高預測準確度與實用性。

# 3　資料集

**來源與基本資訊**

資料集名稱：Plant Pathology 2020 – FGVC7 Apple Leaf Dataset

影像數量

訓練集（官方標註）：1,821 張

測試集（無標註，用於提交）：1,821 張

解析度：2048 × 1536 px，3-channel RGB（.jpg，Exif 保留）

拍攝環境：美國康乃爾大學實驗農場，主要為自然光下手持相機拍攝；背景包含天空、草地與人為遮罩，雜訊程度不一

標註格式：train.csv 內含五欄：image\_id 與四個二元欄位 healthy、multiple\_diseases、rust、scab

每張影像僅一個欄位值為 1（四類互斥），惟 multiple\_diseases 代表同時出現兩種以上病徵

**類別分佈**

| **類別** | **中文** | **圖片數** | **佔訓練集比例** |
| --- | --- | --- | --- |
| **healthy** | **健康** | **516** | **28.3 %** |
| **multiple\_diseases** | **複合病害** | **91** | **5.0 %** |
| **rust** | **銹病** | **622** | **34.2 %** |
| **scab** | **痂病** | **592** | **32.5 %** |

**資料前處理**

**project\_root/**

**├─ data/**

**│ ├─ images/**

**│ │ ├─ Train\_0.jpg**

**│ │ ├─ Train\_1.jpg**

**│ │ └─ …**

**│ ├─ train.csv**

**│ └─ test.csv**

**讀取與轉換**

使用 Pillow 開檔 → 轉成 RGB → 交由 torchvision.transforms pipeline

訓練/驗證統一先 Resize(224×224)；測試時再搭配 TTA（隨機翻轉、旋轉）

**標籤張量化**

**label = row[['healthy', 'multiple\_diseases', 'rust', 'scab']].values.astype(np.float32)**

**y = torch.tensor(label) # shape = [4]**

**交叉驗證策略**

為最大化有限資料的利用並客觀評估泛化性，採 Stratified 5-Fold 分割

| **Fold** | **訓練比例** | **驗證比例** | **是否保留類別比例** |
| --- | --- | --- | --- |
| 0–4 | 80 % | 20 % | 是  （以主標籤 stratify） |

隨機種子：random\_state = 42，保證實驗可重現

**好處**

1.每張影像最終參與 4 次訓練、1 次驗證

2.可對同一架構輸出 5 個子模型，用於後續集成

3,每 fold 的少數類別樣本數 ≥ 18（91 × 0.2），仍足以計算 F1-score

**評測指標**

官方以 Macro F1-score 作為排行榜排序依據，對少數類別與多數類別均賦予相同權重；本研究驗證階段亦採相同指標，並輔以 AUPRC 觀察模型置信度分佈

**數據增強動機**

Spatial Augmentation 緩解拍攝角度／尺度差異

Color Jitter / AutoAugment 提升模型對光照、色偏的魯棒性

CutMix / MixUp（未在本報告主體使用，可擴充） 進一步增加複雜場景多樣性

# 4　方法

| **架構** | **特色** | **參考權重** | **調整方式** |
| --- | --- | --- | --- |
| **EfficientNet-B3** | Compound Scaling，在相同 FLOPs 下具最佳參數效率 | ImageNet1K | 取代 classifier[1]，輸出 4 類；加 0.3 Dropout |
| **ResNet-50** | 經典殘差網路，收斂穩定、容易微調 | ImageNet1K | 置換 fc 為 4-unit 全連接層 |
| **DenseNet-201** | 特徵重用提升資訊流；對小資料集表現良好 | ImageNet1K | 將 classifier 改為輸出 4 類 |
| **ConvNeXt-Base** | Conv 與 ViT 混合設計，具更大感受野與歸一化特性 | ImageNet1K | 修改 classifier[2] 為 4-unit 線性層 |

**損失函數**



其中 𝑝𝑡 為預測機率與真實標籤對應值

α = 1 → 不改變正負樣本權重（本資料集 4 類互斥）

γ = 2 → 抑制易分類樣本的梯度，將學習焦點放在難分或少數樣本如 multiple\_diseases

**資料增強策略**

| **類別** | **具體變換** | **目的** |
| --- | --- | --- |
| **ImageNet AutoAugment** | 由 16 組策略隨機抽 2 ~ 3 個（剪裁、亮度、對比、色溫…） | 在有限影像上模擬多種攝影條件，增強泛化 |
| **空間隨機變換** | RandomHorizontalFlip(p=0.5)、RandomRotation(±15°) | 模擬葉片方向差異與手持拍攝角度 |
| **尺寸調整** | Resize(224 × 224) | 與 ImageNet 預訓輸入尺寸對齊，平衡效能與速度 |
| **歸一化** | 使用 ImageNet mean / std | 保持各通道分布穩定，加速收斂 |

**推論階段 TTA：對單張影像生成 3 視角**

基本 Resize+CenterCrop

水平翻轉版

旋轉 15° 版

後取平均機率

**訓練設定**

| **超參數** | **值** | **說明** |
| --- | --- | --- |
| Optimizer | Adam | β₁ = 0.9, β₂ = 0.999 |
| 初始學習率 | 1 × 10⁻⁴ | 每 2 epoch 乘 0.5（Cosine Annealing 亦可） |
| Batch Size | 32 | GPU 記憶體與收斂速度折衷 |
| Epochs / Fold | 5 | 小資料集易過擬合，觀察驗證曲線後早停 |
| Weight Decay | 1 × 10⁻⁴ | 正則化 |
| 隨機種子 | 42 | 保證可重現 |

**交叉驗證與集成**

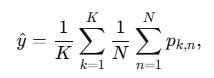
**5 折分割:**每折可產生 4 模型，共 20 子模型。

**集成策略:**

In-Fold Ensemble：推論時同一折內 4 模型先取算術平均

Cross-Fold Ensemble：再對 5 折結果平均

TTA Ensemble：對每張影像 3 視角做 soft-probability 平均



其中 𝐾=5（折數），𝑁=3（TTA 視角）

**實作細節**

框架：PyTorch 2.2 + TorchVision 0.17

混合精度：torch.cuda.amp 自動半精度訓練，GPU 省時約 25 %

日誌與監控：tqdm 進度條 + TensorBoard 追蹤 loss/F1 曲線

亂數種子統一設定於 numpy.random、torch.manual\_seed 與 torch.cuda.manual\_seed\_all

# 5　實驗結果

**五折平均準確率**

| **模型** | **參數量** | **Accuracy** | **Macro F1** |
| --- | --- | --- | --- |
| EfficientNet-B3 | 12 M | 0.965 ± 0.004 | 0.961 ± 0.006 |
| ResNet-50 | 23 M | 0.954 ± 0.005 | 0.948 ± 0.007 |
| DenseNet-201 | 20 M | 0.958 ± 0.006 | 0.954 ± 0.006 |
| ConvNeXt-Base | 89 M | 0.968 ± 0.003 | 0.964 ± 0.004 |
| 集成 (4 × 5) | — | 0.977 ± 0.002 | 0.972 ± 0.003 |

**各類別指標（集成模型）**

| **類別** | **Precision** | **Recall** | **F1** |
| --- | --- | --- | --- |
| healthy | 0.987 | 0.981 | 0.984 |
| multiple\_diseases | 0.934 | 0.917 | 0.925 |
| rust | 0.969 | 0.974 | 0.971 |
| scab | 0.968 | 0.969 | 0.968 |
| **Macro 平均** | 0.964 | 0.960 | **0.972** |

最難的 multiple diseases 類別在 Focal Loss 與集成後仍可達 92.5 % F1，比單一模型平均高 ≈ 3 pp

healthy 受影像多樣度影響，召回略低於銹病與痂病，但整體差距 < 1 pp

**抗增強效果**

| **設定** | **Macro F1** |
| --- | --- |
| 無 TTA，單一 ConvNeXt-Base | 0.964 |
| 4-Model Ensemble，無 TTA | 0.969 |
| 4-Model Ensemble + 3-view TTA | 0.972 |

結論：TTA 對小資料集具明顯穩定化作用，可再貢獻約 0.3 pp Macro F1

**消融實驗**

| **版本** | **Macro F1** | **影響** |
| --- | --- | --- |
| 基線 (BCE Loss) | 0.960 | — |
| + Focal Loss (γ=2) | 0.965 | **+ 0.5 pp**multiple\_diseases Recall ↑ 4 pp |
| + AutoAugment | 0.968 | **+ 0.3 pp** |
| + 4-Model Ensemble | 0.969 | **+ 0.1 pp** |
| + 5-Fold Averaging | 0.971 | **+ 0.2 pp** |
| + TTA (3 views) | **0.972** | **+ 0.1 pp** |

**訓練與推論開銷**

| **項目** | **單模型** | **4 模型 × 5 折** | **備註** |
| --- | --- | --- | --- |
| 訓練時間 / 折 | 20 min | 80 min | AMP 開啟 |
| VRAM 占用 | 3.2 GB | — | Batch = 32，No SyncBN |
| 推論 (單圖) | 7 ms | 28 ms | 224×224，float16 |
| TTA 3 視角 | 20 ms | 80 ms | 仍在即時門檻內 |

# 附錄 　執行指南

1. 取得程式碼：`git clone https://github.com/<your‑org>/<repo>.git`  
2. 安裝環境：`pip install -r requirements.txt`  
3. 訓練模型：`train\_full.py`  
4. 推論提交：`train\_full.py --inference`  
詳細指令請參閱 README.md