

基于生成式人工智能（GenAI）的合成和实地图像，通过定制的 EfficientNetV2-L 模型改进西瓜（*Citrullus lanatus*）疾病分类

Nitin Rai¹, Nathan S. Boyd¹, Gary E. Vallad²,
Arnold W. Schumann^{3*}

¹Gulf Coast Research and Education Center (GCREC), University of Florida, Wimauma, FL, USA.

²Department of Plant Pathology, Gulf Coast Research and Education Center (GCREC), University of Florida, Wimauma, FL, USA.

³Citrus Research and Education Center, University of Florida, Lake Alfred, FL, USA.

*Corresponding author(s). E-mail(s): schumaw@ufl.edu;

Abstract

The current advancements in generative artificial intelligence (GenAI) models have paved the way for new possibilities for generating high-resolution synthetic images, thereby offering a promising alternative to traditional image acquisition for training computer vision models in agriculture. In the context of crop disease diagnosis, GenAI models are being used to create synthetic images of various diseases, potentially facilitating model creation and reducing the dependency on resource-intensive in-field data collection. However, limited research has been conducted on evaluating the effectiveness of integrating real with synthetic images to improve disease classification performance. Therefore, this study aims to investigate whether combining a limited number of real images with synthetic images can enhance the prediction accuracy of an EfficientNetV2-L model for classifying watermelon (*Citrullus lanatus*) diseases. The training dataset was divided into five treatments: H0 (only real images), H1 (only synthetic images), H2 (1:1 real-to-synthetic), H3 (1:10 real-to-synthetic), and H4 (H3 + random images to improve variability and model generalization). All treatments were trained using a custom EfficientNetV2-L architecture with enhanced fine-tuning and transfer learning techniques. Models trained on H2, H3, and H4 treatments demonstrated high precision, recall, and F1-score metrics. Additionally, the weighted F1-score increased from 0.65 (on H0) to 1.00 (on H3-H4) signifying

that the addition of a small number of real images with a considerable volume of synthetic images improved model performance and generalizability. Overall, this validates the findings that synthetic images alone cannot adequately substitute for real images; instead, both must be used in a hybrid manner to maximize model performance for crop disease classification.

Keywords: Watermelon diseases, Classification, Convolutional neural network, Generative AI, Synthetic data, Precision agriculture.

1 介绍

西瓜 (*Citrullus lanatus*) 被认为是葫芦科中最重要的作物之一，特别是其丰富的营养成分，包括高含水量和维生素。根据美国农业部 (USDA) 经济研究服务的数据，佛罗里达州在 2021 年生产了 10.2 亿磅的西瓜，超过了德克萨斯州和加利福尼亚州的产量 (Weber and Kramer 2022)。如此大规模的生产需要持续的作物保护措施，以尽量减少因生物胁迫造成的产量损失。据报道，种植者通常在西瓜生产期间每 7 到 14 天施用一次杀菌剂，每个季节平均喷洒大约四次 (Mossler 2005)。持续的杀菌剂施用有助于整个季节管理疾病，例如树胶菌核病、白粉病和炭疽病等 (Keinath and Miller 2022)。智能喷洒技术是一种应能减少杀菌剂用量而不损失效果的工具，从而提供早期检测和农民成本节约 (Ampatzidis 2022; Durham 2016)。

在佛罗里达，西瓜受到多种真菌病害的影响，如黑斑病 (*Colletotrichum obiculare*)、霜霉病 (*Pseudoperonospora cubensis*)、茎溃病 (*Didymella bryoniae*)、尾孢菌病 (*Cercospora citrullina*) 和链格孢叶斑病 (*Alternaria cucumerina*) (Roberts et al. 2023)。此外，多种类型的病毒也影响这一作物的整体生长 (Ravikumara et al. 2025)。这些病原体影响作物的不同部位，受影响的区域在营养组织上显示出一系列可见症状。例如，黑斑病蔓延至叶、茎甚至果实，并可能导致果实在采收后几天内漏水或腐烂。相反，霜霉病仅影响叶片，并可能在作物后期引起叶片脱落。除了这些病原体，病毒，如西瓜花叶病毒可以通过蚜虫传播，叶片表现出多种症状，如浅绿色、矮化，甚至显示光暗斑块的花叶图案。虽然种植者倾向于施用杀菌剂来减轻这些疾病对西瓜的影响，但尤其是在季节初期，两种或多种疾病的症状可能看起来相似。例如，黄瓜斑潜蝇 (*Liriomyza sativae*)，西瓜中的一种常见害虫，造成的叶损伤可能与黑斑病或茎溃病混淆。

智能喷洒技术利用计算机视觉方法，该方法由经过深度学习 (DL) 模型训练的大规模植物病害图像数据集提供支持，以识别感兴趣的目标。这些计算机视觉系统与杀虫剂喷洒装置集成，使其能够在行进过程中做出决策，并根据需要进行目标喷洒 (Li et al. 2022; Oberti et al. 2016)。一个可靠的计算机视觉系统需要大规模的数据集和基于人工智能 (AI) 的分类算法在这些数据集上进行训练，以实现在动态和不受控制的环境中对疾病的分类。关于为模型训练生成大规模数据集，目前的研究工作集中在使用生成式人工智能 (Gen AI) 来创建农作物或杂草数据集的合成图像 (Lu et al. 2022; Pallottino et al. 2025)。例如，基于稳定扩散的文本到图像方法被用来生成自然场景中的害虫图像 (Wang et al. 2024)。大多数研究人员依赖于使用生成对抗网络 (GAN) 进行图像到图像转换以生成合成图像。在图像增强的一部分中，已有超过 15 种不同种类的 GAN 网络被用来生成合成图像，以训练用于作物健康评估的 AI 模型 (Lu et al., 2022)。深度卷积 GAN (DCGAN) 和循环一致 GAN (CycleGAN) 是合成

图像生成中最广泛采用的方法 (Douarre et al. 2019; Gomaa and Abd El-Latif 2021; Zeng et al. 2020; Sun et al. 2020)。

除了将生成式人工智能应用于大规模合成图像生成外，还投入了大量精力开发用于疾病分类的人工智能模型。从简单的分类到实例分割，各种方法已被用于开发智能计算机视觉系统 (Li et al. 2022; Yang et al. 2025)。这些计算机视觉模型随后可以与地面制图或喷洒系统集成。在疾病上进行目标喷洒在决策实时喷洒策略方面带来了若干挑战。首先，有些疾病表现出相似的症状学，需要进一步的实验室检测以验证。这一过程需要时间并需要植物病理学家的专业技能。其次，疾病症状通常随机散布在叶片上，这使得模型定位和绘制边界框成为一项计算密集型任务。在这种情况下，最终目标仅仅是在全局图像上识别疾病并执行喷洒操作，而无需考虑病变级别的定位。在许多研究中，基于深度学习的卷积神经网络 (CNN) 分类方法在分类害虫和植物疾病时，其表现优于更复杂的检测或实例分割技术。例如，与基于边界框的 YOLO 模型相比，某 CNN 架构在分类害虫和植物疾病方面达到了 84.9 % 的准确率 (Shoaib et al. 2025)。另一项研究表明，使用 CNN 模型可在早期阶段检测植物疾病并具有高准确率 (Venkateswara and Padmanabhan 2025)。考虑到在大规模合成农作物图像生成和基于 CNN 的分类方法领域为增强实时目标喷洒技术所取得的进展，目前在将真实田间图像与合成图像相结合来提升农作物疾病分类效果的研究投入尚少。该研究将回答一个重要问题：将真实田间图像与合成数据融合，是否能够更好地进行疾病分类？因此，本研究的目标如下：

1. 评估仅使用合成图像作为训练深度学习模型的替代方法，而不使用手动收集的真實场景图像的有效性。
2. 确定模型准确性是否会随着包含合成图像的真实图像而提高。
3. 研究添加无关 (未知) 类别对模型预测准确性和分类性能的影响。

总体而言，Section 2 讨论了从 GenAI 应用生成合成图像开始到 EfficientNetV2-L 架构定制进行训练和测试的数据分析流程的方法。随后，Section 3 按照本研究的目标报告了在测试数据集上部署模型后的相关结果。Section 4 提供了关于计算机视觉应用于疾病分类方面的关键讨论。此外，Section 5 强调了研究的局限性以及未来研究方向。最后，Section 6 对研究工作进行了总结。

2 材料和方法

图 1 显示了进行本研究所涉及步骤的概述。此外，在每个部分中还分别解释了模型指标。

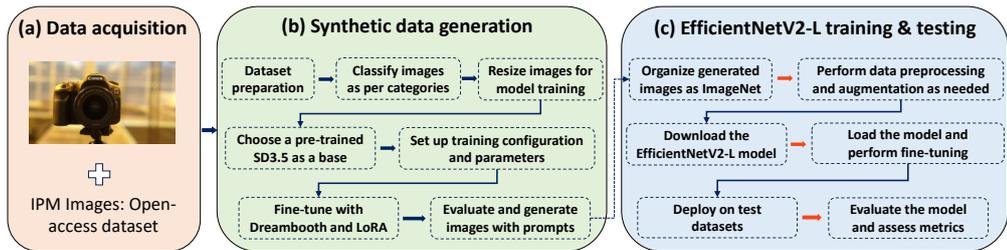


Fig. 1: 从图像数据生成到模型训练，整体说明执行西瓜疾病分类所采取的系统方法。

2.1 结合专家建议的图像数据采集

用于训练和测试模型的疾病图像数据集被标记为三类：真菌、健康和病毒。整个训练集包含五种处理，其中第五种处理包括来自 ImageNet 数据集的第四类称为未知类 (Deng et al. 2009)。添加该未知类是为了使模型专注于植物疾病的特征，并忽略不相关的模式 (Robinson et al. 2021)。在真菌类别中，使用了两种疾病，炭疽病和霜霉病。对于病毒类，训练集中加入了西瓜花叶病毒的图像。真菌类别的数据集，即炭疽病和霜霉病，是在 2025 年的夏天使用索尼的手持单反相机拍摄的。数据采集从 5 月 5 日th 到 6 月 15 日th 进行，分辨率为 2048 × 1536。为了准备分析的测试集，进行了多次实地考察以拍摄所有类别的图像。这些图像在植物病理学专家的指导下进一步验证了特定疾病。图 2 显示了合成生成的疾病图像与在真实田间条件下拍摄的图像之间的区别。

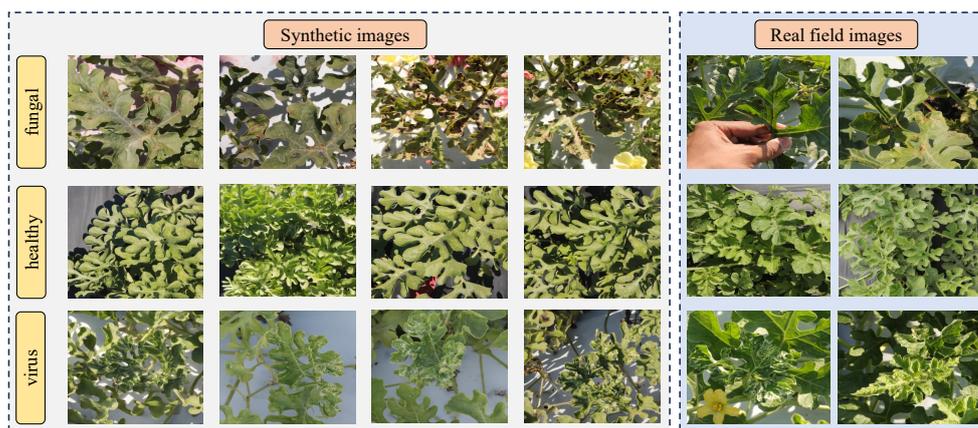


Fig. 2: 用于模型训练和测试的样本图像由合成图像和真实图像组成。合成图像中的炭疽病和霜霉病分别被黄颜色和橙颜色圈出。

2.2 用于模型训练的整体图像数据构建

2.2.1 用于合成数据生成的多模态生成式 AI

生成性人工智能 (GenAI) 是一种人工智能 (AI)，其不局限于学习单一形式模式，如图像、音频或视频。相反，GenAI 方法能够从多个来源开发的数据集中提取模式，以根据学习到的特征生成“前所未见”的样本。例如，在智能农业中，GenAI 可以用来从极为有限的疾病图像样本中生成合成图像。它能够从一组样本图像中学习或模仿疾病特征，然后生成具有不同背景和环境条件的合成版本。它能够为生成的数据集提供多样性，从而迅速生成数千张合成图像。这在农业领域尤其有用，因为互联网上的训练样本多样性较低，且通常不是开放访问的，导致生成的分类模型聚集且不够健壮 (Han et al. 2025; Lu et al. 2022; Davies and Bouldin 2009)。

在本研究中，稳定扩散模型 (SD 3.5M) 通过训练了少量由真菌和病毒病原体引起的多种疾病的西瓜真实图像。此外，使用了两种方法，Learn on Reconstruction and Attention (LoRA) (Hu et al. 2022) 和 DreamBooth (Ruiz et al. 2023)，对 SD 3.5M 模型进行了微调，以生成在不同背景和光照条件下的逼真图像。使用了一组三十张从特定类别中挑选出的最佳代表性真实图像，通过相关提示工程方法生成多张合成图

像。对于真菌类别，分别使用了 32 张炭疽病和 35 张霜霉病的真实图像。同样，使用了 34 张健康叶片和 30 张花叶病毒的图像来生成合成图像。在训练 SD 3.5M 模型时，单个类别的疾病没有被混合。例如，只有显示炭疽病症状的图像用于训练和生成带有炭疽病的西瓜叶子的合成图像。采取这种策略是因为在 SD 3.5M 模型训练中混合多种疾病症状可能导致生成不自然的合成图像。模型训练完成后，提供了特定的提示语来生成合成图像（图 3 c）。图 3 显示了所使用真实图像的样本，一些微调方法，以及用于生成多样性合成数据集的提示语。有关使用 SD 多模态模型生成合成样本的更多信息，参阅 Rai et al. (2025)。

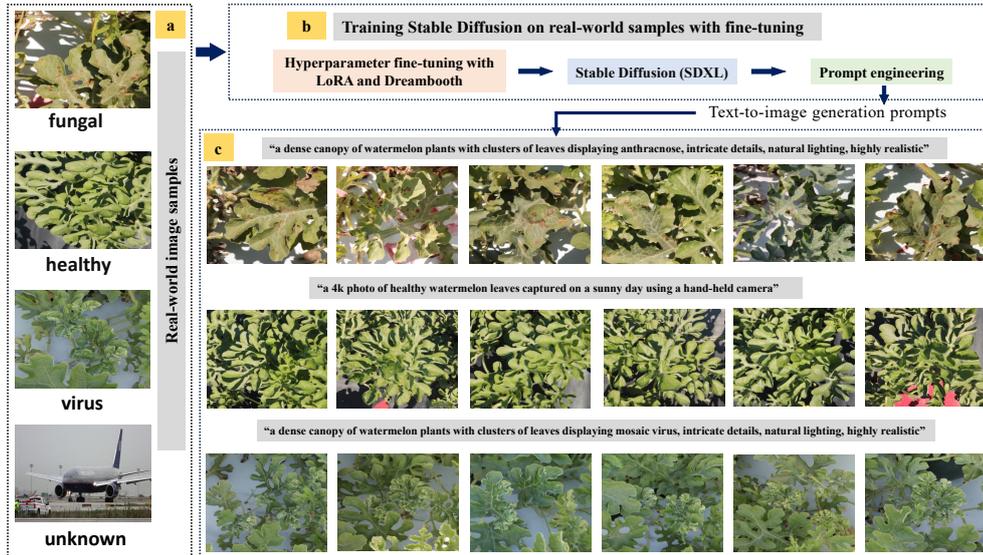


Fig. 3: 对 Stable Diffusion (SD) 生成模型进行训练的步骤示例，使用的是真实世界的样本。它包括：(a) 收集真实世界的样本，(b) 使用低秩适应 (LoRA) 和 Dreambooth 微调平台训练 SD 3.5M，以及 (c) 应用提示工程以生成所需数量的合成样本。

2.2.2 疾病数据集分类和训练平台

该数据集被分为五种处理：(a) 仅真实图像 (H0)，(b) 仅合成图像 (H1)，(c) 真实图像 + 合成 1 图像 (1:1) (H2)，(d) 真实图像 + 合成 2 图像 (1:10) (H3)，以及 H3+ 随机类 (H4)。在最终处理步骤中，加入了一个随机类（包含来自 ImageNet 的图像），以防止在异常图像（如塑料覆盖物、杂草或土壤）输入模型进行分类时产生虚假结果。在这种情况下，模型将其分类为“未知”，而不是标记为真菌、健康或病毒。为了防止偏倚和过拟合，每个处理类别中的图像数量都是均衡的。每个处理开始时都使用固定的 750 张真实图像集，然后进一步分为训练集 (70%)、验证集 (15%)、测试集 (15%)。在这个划分中，112 张测试图像被预留出来，并在所有处理步骤中保持不变，而剩余的图像则进一步划分为训练和验证集 (80%: 20%)。表格 1 显示了用于模型训练和测试中每种处理的图像的精确数量。总的目标是通过控制总体数据集大小，以实现公平的比较，从而可以在没有较大数据集规模影响的情况下，孤立地研究真实图像的包含对模型性能和泛化的影响。此外，在真菌类中，将两种不同的真菌病，炭疽病和霜霉病，合并为一个类别，并使用 EfficientNetV2-L 进行模型训练。这种分组是

为了简化分类任务，并反映真实世界中的疾病诊断情景，在这种情况下，精确识别特定病原体可能并不总是可行的。这种方法使模型能够区分真菌症状和非真菌症状，这通常足以在实际环境中及时进行杀菌剂施用。整体数据分析是在一个高性能计算系统上进行的，HiPerGator，该系统配备了一个具有 180 GB VRAM、64 GB RAM 和 10 个 CPU 核心的 Nvidia B200 Quadro GPU。此外，使用支持 Python (v3.11.11)、torchvision (v0.20.0)、torch (v2.5.1) 和 CUDA 12.4 的 Jupyter Notebook 来执行模型的训练和测试。

2.3 提出的 EfficientNetV2-L 分类架构，结合分类器头部微调和迁移学习

EfficientNetV2-L 分类架构是一个高性能卷积神经网络 (CNN)，针对高效的训练和测试任务进行了优化。与其他最先进的模型 (SoTA) 相比，如“只需看一眼” (YOLO)，EfficientNetV2-L 模型在实时疾病分类中具有四大优势：(a) 模型速度，(b) 在理解全局特征而不是单个疾病症状方面的鲁棒性，(c) 无需手动标注数据，(d) 无需位置 (坐标或边界框) 特定信息以实现更快的喷洒和驱动。然而，来自互联网的 EfficientNetV2-L 预训练模型是基于 ImageNet 数据集训练的，该数据集只适合高精度、高准确率地分类一般领域。因此，有必要在农业领域用疾病数据集来训练模型，并进行额外的微调。作为一个总体方法，本研究使用了迁移学习技术，该技术分两步完成：(a) 用额外的层训练一个新的分类器头，和 (b) 通过选择性冻结基础层微调剩余层 (更深层)。此外，应用了若干超参数来增强训练策略，例如优化器和回调技术。

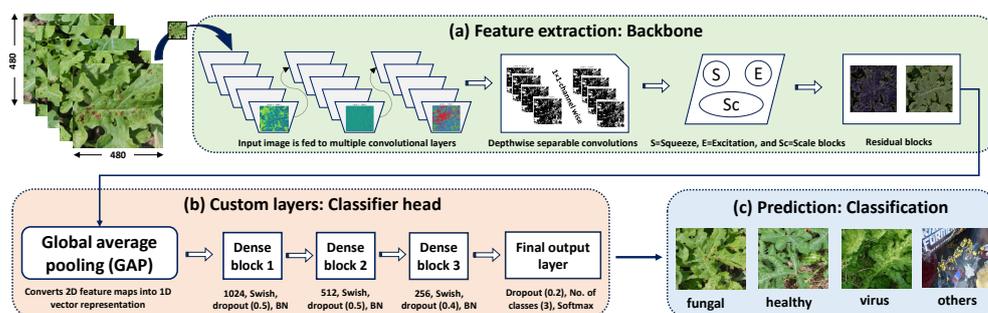


Fig. 4: 一个表示 EfficientNetV2-L 模型的示意图，其中包含各种特征提取块和层。(a) 特征提取步骤，从图像中提取细粒度特征，例如边缘、纹理和结构，(b) 添加的自定义层，用于训练分类器头部，增加了三个密集块和一个最终输出层，用于类的概率映射，(c) 将测试图像预测为其四个相应的类别。

在图 4 a 中，作为特征提取器使用了在 ImageNet 上训练的 EfficientNetV2-L 主干。通过主干进行特征提取时，前几层保持冻结，仅使用较深层提取高层图像信息。例如，低层次特征如边缘和纹理通过 CNN 提取，而深度可分离卷积用于在保留中层特征的同时缩小输入图像样本的尺寸。随后的层为压缩 (S) 和激励 (E) 块，用于在输入图像中提取非常具体的细粒度病变区域。此外，在主干模型之上添加了一个自定义分类头，包含三个具有 Swish 激活、批量归一化和 dropout 正则化的全连接密集层，以增强学习稳定性并防止过拟合 (图 4 a)。自定义分类器首先在基础模型保持

Table 1: 用于训练和测试 EfficientNetV2-L 模型的每类西瓜病害图像概述。

Classes	Total images	Training	Validation	Testing	Split ratio	Purpose
fungals			H0 (only real images)			
healthy	750 (real)	526	112	112 (Kept constant across all the treatments)	70 % :15 % :15 %	Baseline with 100 % real images
virus						
fungals			H1 (only synthetic images)			
healthy	638 (synthetic)	510	128	—	80 % :20 %	Baseline with 100 % synthetic images
virus						
fungals			H2 (1real: 1synthetic)			
healthy	638 (real) + 638 (synthetic)	1,021	255	—	80 % :20 %	Moderate number of synthetic images
virus						
fungals			H2 (1real: 10synthetic)			
healthy	638 (real) + 6,380 (synthetic)	5,614	1,404	—	80 % :20 %	Large number of synthetic images
virus						

∞

冻结的情况下单独训练，以确保仅新添加的层学习与数据特定的疾病特征（图 4 b）。一旦分类器达到稳定准确率，就会在 EfficientNetV2-L 架构的更深层启动微调。此步骤允许在学习病特异性特征方面有更多即兴空间，同时防止可能导致过拟合或模型不稳定的权重大幅变化。所有模型均在调整为 480×480 分辨率的图像上进行训练。使用渐进学习率，在微调期间使用较低的学习率以确保更好的收敛。模型使用 AdamW 优化器进行训练，并使用提前停止和学习率减少回调进一步稳定整体架构，以实现模型对疾病特征的泛化能力。

2.4 评估模型性能的指标

为了评估在数据集所有类别上训练的模型表现，选择了四个关键指标。它们是：精确率、召回率、F1 分数和混淆矩阵。虽然精确率、召回率、F1 分数和加权 F1 分数展示了模型在分类个别类别上的效率，混淆矩阵则显示了被模型正确分类的测试图像数量（方程 1 - 4）。除此之外，为了更好地理解在合成和实地图像的数据集的不同分布下特征的可分性（表 1），还报告了两项测试，t-分布随机邻域嵌入（t-SNE）（Rousseeuw 1987）和统一流形近似与投影（UMAP）（Davies and Bouldin 2009）。t-SNE 是一种非线性降维技术，旨在帮助可视化二维空间中定位的高维大规模数据集及其潜在的簇。与 t-SNE 不同的是，UMAP 则同时强调了二维空间中的局部和全局结构可视化，从而突出类的可分性。尽管基于类的簇可视化非常有帮助，关于这些簇的定量度量，如轮廓系数和戴维森堡丁指数也被报告了。这些分别基于方程 5 & 6。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Weighted F1} = \sum_{i=1}^N w_i \times \text{F1}_i \quad (4)$$

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

$$\text{DBI} = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d_{ij}} \right) \quad (6)$$

3 实验结果

3.1 评估模型性能指标和聚类验证

该模型在 H2-H4 处理的所有类别中均显示出较高的精度、召回率和 F1 得分（表 2）。健康样本的测试图像在几乎所有模型中都以接近 100 % 的准确率被分类，唯有仅使用合成图像时例外（图 5 b）。在仅合成图像条件下，健康和病毒症状常被误分类为真菌症状，显示出模型泛化能力有限。值得注意的是，与仅使用真实图像（H0）条件相比，真菌分类有所改善（正确分类 112 个），但 64 个健康图像和 92 个病毒图像

被误分类为真菌，表明对真菌预测的偏倚。模型泛化能力差的另一个可能原因是被分类为真菌病的不规则褐色病变（由风害或环境应力引起）。包含 1:1 比例的真实和合成图像的处理 H2（图 5 c）在所有类别中表现强劲，与仅使用真实或合成处理相比，精确度得到提高，误分类减少。病毒和真菌类别之间的混淆显著下降，病毒类别实现了超过 99 个的正确预测，仅有 4 个和 10 个分别误分类为真菌和健康。

Table 2: 训练模型在四种测试数据集分布上报告的多个性能指标。

Classes	Precision	Recall	F1-score	Weighted average F1-score
H0 (only real images)				
fungus	0.78	0.25	0.37	0.65
healthy	0.42	1.00	0.59	
virus	0.75	0.21	0.33	
H1 (only synthetic images)				
fungus	0.42	0.98	0.59	0.74
healthy	0.79	0.43	0.56	
virus	1.00	0.09	0.16	
H2 (1real + 1synthetic)				
fungus	0.94	0.89	0.91	0.92
healthy	0.84	0.98	0.91	
virus	0.98	0.88	0.93	
H3 (1real + 10synthetic)				
fungus	0.99	1.00	1.00	1.00
healthy	1.00	1.00	1.00	
virus	1.00	0.99	1.00	
H4 (1real + 10synthetic + unknown)				
fungus	0.97	0.99	0.98	0.99
healthy	1.00	0.99	1.00	
virus	0.99	1.00	1.00	
unknown	1.00	0.97	0.99	

当少量真实图像与大量合成图像结合使用时（H3），所有类别都被近乎完美地分类，包括 114/114 的真菌、113/113 的健康以及 112/113 的病毒预测，仅显示出单个病毒实例被误分类为真菌（图 5 d）。这表明，当模型用真实和合成数据的混合进行训练时，它能够很好地泛化到所有类别。最后，在四分类处理（H4）中，添加一个未知类别，保持了整体模型性能（图 5 e）。模型成功地以高准确度扩展到了额外的类别（未知的为 110/112），并且对其余类别几乎是完美预测。总体而言，结果强调了训练的 EfficientNetV2-L 架构在不同处理方式下的稳健性。此外，至少量的真实图像也显著提高了模型的性能，尤其是在泛化能力方面，使其非常适合在现实世界的疾病分类场景中部署。

在仅使用真实图像的 H0 处理下，模型在不同类别间表现出较差的泛化能力，尤其在真菌和病毒类别上。虽然健康类别的召回率达到完美的 1.00，但其低精确度（0.42）表明有大量假阳性结果，这可能是由于模型过度预测该类别造成的（表 2）。真菌和病毒类别的召回率很低（分别为 0.25 和 0.21），表明模型难以正确识别这些类别。整

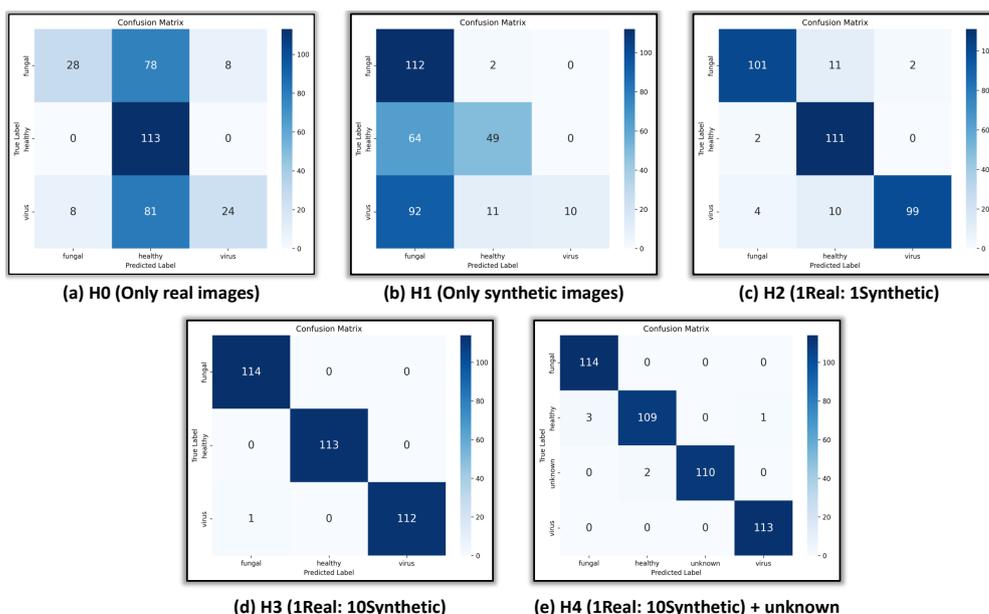


Fig. 5: 混淆矩阵展示了以下处理中真实标签与预测标签的类别数量：(a) 仅使用真实图像的 H1, (b) 仅使用合成图像的 H2, (c) 真实与合成比例为 1:1 的 H2, (d) 真实与合成比例为 1:10 的 H3, 以及 (e) 在 H3 的基础上增加了一个随机 (未知) 类别的 H4。

体加权 F1 分数为 0.65, 显示出模型在仅依赖有限数量的真实图像进行训练时, 因图像多样性不足和视觉模式表征不充分而难以泛化。真菌的召回率保持在 0.98 的高水平, 但精确度下降至 0.42, 表明假阳性率很高。这可能是因为合成数据对与真菌症状相关的特征的过度表现。病毒分类仍然非常差, 召回率为 0.09, F1 分数为 0.16, 这可能是因为合成生成的病毒症状未能充分反映真实环境的变化 (表 2)。健康类别表现中等, 精确度为 0.79, 召回率为 0.43。相比 H0 处理, 总体加权 F1 分数提高到 0.74, 这表明虽然生成合成图像增加了训练量, 但无法替代真实图像或在真实环境条件下引入更好的泛化能力。

使用真实图像与合成图像 1:1 比例训练的模型在性能指标方面表现得更准确。菌类、健康和病毒三类均获得高精确率和召回率 (均在 0.84 以上), F1 评分在 0.91 至 0.93 之间。真实与合成图像的这种平衡使模型能够学习更稳健、可推广的特征 (Bafghi et al. 2025)。相比仅使用真实或合成图像, 组合真实与合成图像的加权 F1 评分提高到了 0.92, 表明这种组合的有效性。使用 1:10 的真实与合成图像比例导致完美的分类, 所有三类的精确率、召回率和 F1 评分均等于 1.00。加权 F1 评分为 1.00, 说明了准确性。最后, 使用 H4 处理且保持与 H3 相同的 1:10 真实与合成比例, 但添加一个额外的未知类别训练的模型, 在处理增加的复杂性方面表现得异常出色, 从而实现了各类别几乎完美的精确率和召回率。F1 评分均高于 0.97, 其中新引入的未知类别高达 0.99。先前在 H0-H2 中显示出一致性的病毒类实现了完美的分类 (表 2)。加权 F1 评分为 0.99, 表明模型不仅适应了扩展的类别结构, 而且几乎没有损失准确性或推广性。

后续的绘图展示了基于 EfficientNetV2-L 模型提取的高维表示进行的 t-SNE 特征嵌入 (图 6 a-e)。每个图显示了高维特征的二维投影, 点按照多类别分类进行色彩编

码。相比于仅使用合成图像的处理，仅使用真实图像的处理（H0）展示了更紧凑的聚类（尤其是健康类别）（图 6 a）。这与最低的轮廓分数和最高的 DBI 分数一致，表明类别内紧凑性差以及类别边界之间有显著重叠（表 3）。仅使用合成数据的情况下，显示的图更加分散，轮廓分数较低（图 6 b & 表 3）。真实与合成数据 1:1 的比例显著改善了聚类，轮廓和 DBI 分数分别为 0.61 和 0.49（图 6 c & 表 3）。这种组合突出了两种图像类型的价值。真实图像捕捉了多样的形态和纹理，呈现了真实的症状表现，而合成图像提供了一致的表示，两者共同改善了类内和类间的分离。使用 1:10 的真实与合成（H3）数据达到了最佳的聚类效果（图 6 d & 表 3）。此外，包含未知类别（H4）并未对聚类产生负面影响（轮廓和 DBI 分数分别为 0.81 和 0.27），这表明模型即使在存在不相关图像的情况下也能学习边界（图 6）。这对于实时疾病分类非常重要，因为未知物体，如杂草、塑料地膜、强风或环境因素引起的棕色病斑，常常在准确分类疾病症状时带来挑战。

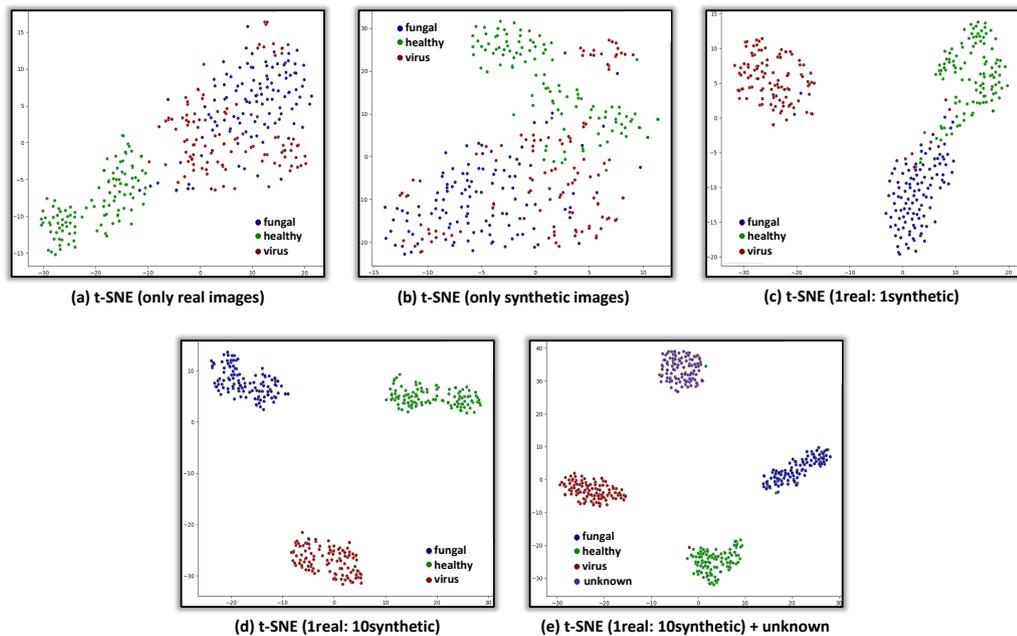


Fig. 6: 二维 t-分布随机邻居嵌入（t-SNE）图展示了模型的测试结果，其中颜色代表类别：（真菌，健康，病毒，和未知）。

在图 7 中，展示了通过 EfficientNetV2-L 模型在所有处理条件下学习到的特征嵌入的 UMAP 投影。与 t-SNE 不同，UMAP 保留了局部和全局特征结构，从而更好地提供模型如何在低维空间中有效地分离和组织类别特征的洞察。当仅使用真实图像时，模型表现出较差的聚类，反映在较低的 Silhouette 和 DBI 分数中（表 3）。同样，仅使用合成图像产生了类似的模式（图 7 b）。当真实与合成图像的比例为 1:1 时，聚类质量显著提高，分别实现了 0.65 和 0.45 的 Silhouette 和 DBI 分数。这种平衡重述了 t-SNE 的发现，即模型能够将学习到的特征推广到测试集。当在处理条件中将真实图像添加到一个大得多的合成集合中时，获得了最佳结果（图 7 d & e, 表 3）。总的来

说，t-SNE 和 UMAP 分析都强化了这样的结论：混合数据集，尤其是在 H3 和 H4 中使用的比例，使得模型能够提取比单独使用真实或合成数据更有意义的特征。尽管仅使用真实图像观察到较低的聚类指标，真实样本的纳入仍然是准确疾病分类的关键。

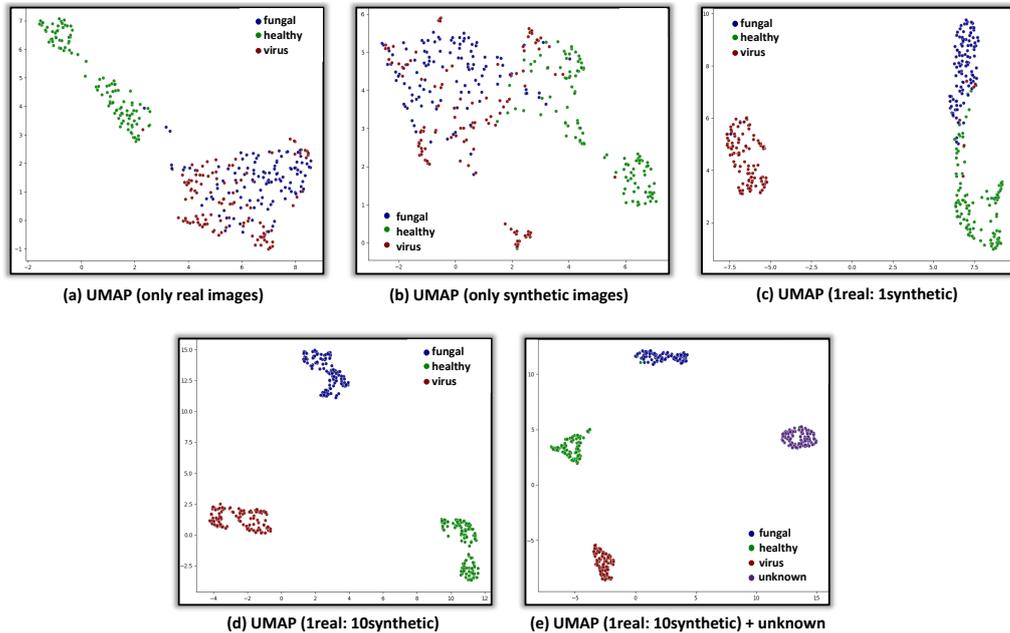


Fig. 7: 二维统一流形近似与投影 (UMAP) 图显示了模型的测试结果，颜色代表类别：（真菌、健康的、病毒和未知）。

4 讨论

4.1 合成图像作为真实图像的替代品

随着当前对于开发涉及生成和训练大规模数据集的制图和定点喷洒技术的需求，使用生成 AI 方法生成的合成图像可能是一种替代方案。然而，这并不一定验证可以替代通过各种传感系统（本研究为 RGB 数码相机）收集的真实图像。真实图像带来了图像内的多种变化，例如不同的土壤背景、动态光照条件、不可预测的天气情况、被遮挡的植物结构和不可预测的变化（灰尘、随风的运动模糊或露水）。此外，真实图像还可能捕捉与传感器相关的失真，例如镜头的瑕疵或快门效应，而生成 AI 模型在生成合成图像时可能无法捕捉到这些。尽管合成图像在视觉清晰度上有所提高，并减少了潜在空间中的类内可分性，但真实图像为模型在真实环境条件下的泛化提供了必要的噪声和表型逼真性。因此，虽然合成图像为大规模数据集增加了价值，但它们不能替代或被用作真实图像的替代品。在这种情况下，结合真实图像和合成图像的混合方法将提高模型在现实世界条件下的性能。

Table 3: t-SNE 和 UMAP 在五种处理分布中凝聚定量指标 (Silhouette 和 Davis-Bouldin 系数)。

Treatments	t-SNE		UMAP	
	Silhouette score ^[1]	DBI ^[2]	Silhouette score ^[1]	DBI ^[2]
H0	0.29	2.28	0.31	1.79
H1	0.16	1.99	0.21	2.04
H2	0.61	0.49	0.65	0.45
H3	0.82	0.26	0.85	0.21
H4	0.81	0.27	0.86	0.21

¹Higher is better.

²Lower is better.

图 8 使用训练好的 EfficientNetV2-L 模型展示了从处理 H0-H4 分类的测试图像。如第 2.1 小节中所讨论的, 仅使用真实图像来测试和评估模型的泛化能力。模型在不同处理中的预测反映了在现实世界预测场景中的优势和挑战。当仅使用真实图像 (H0) 时, 模型在存在阴影和棕色病斑时的泛化能力不佳。例如, 在最后一个测试图像中, 病毒症状表现为因风损伤导致的棕色病斑。由于过拟合问题, 模型未能充分学习病毒特征并将其预测为健康。当仅使用合成图像时 (图. 8), 也可以看到类似的预测。这也凸显了从视觉上相似的健康叶片模式中区分早期或轻微症状的困难。对于真实到合成图像 (1:1), 尽管模型正确分类了大多数图像, 但环境工件如阴影或覆盖物引入了一些不一致性。值得注意的是, 这些类的细微特征被模型正确学习, 并进行了正确的预测。H3 (真实到合成, 1:10) 在所有三个类别中表现出色并准确预测。即使在存在覆盖物或重叠叶片的情况下, 模型也能正确识别大多数测试类别。这再次表明在这种处理中的强大学习能力, 可能是在训练期间高质量真实图像表示的结果 (Singh et al. 2024)。随着不相关类的引入, 模型拒绝分布外图像 (未知类) 并准确预测。总体而言, 尽管合成数据在训练中发挥了关键作用, 但测试预测强调了多样化真实田野图像对于实现鲁棒性的重要性, 特别是在处理噪音、遮挡和现实世界场景中的未知对象时。

4.2 添加真实图像和未知类别时的模型泛化

在合成图像中加入真实图像赋予了自然感, 这是合成图像无法捕捉到的。这包括作物遮挡、运动模糊和空间差异, 比如相机角度和高度。这些是各种“边缘案例”, 帮助模型在测试集上实现了更好的泛化。通过整合少至 638 张在真实场地条件下捕获的真实图像 (真实与合成的比例为 1:10), 模型接触到了有助于其实现更高精度、召回率和 F1 分数的示例。整体加权 F1 分数从 0.65 增加到 1.00 (表 2)。从更实际的角度来看, 添加一个不相关的 (未知的) 类别也帮助模型取得了更好的结果。例如, 未知类别充当了一个干扰器, 模拟了广泛的领域条件, 在该条件下, 模型不仅必须识别疾病模式, 还必须识别出不属于任何作物中心类别 (常见物体) 的事物 (Buda et al. 2018)。此外, 这也在 F1 分数和加权 F1 分数中体现得很明显, 分别为 1.00 和 0.99 (表 2)。

4.3 独特的特征空间

t-SNE 度量倾向于更好地保持局部特征, 重点在于类内可分性。在这项研究中, 使用了两个获得 t-SNE 图的必要参数: 困惑度 (30) 和 “n_components” (2)。为了获得更好的聚类图, 困惑度是一个重要参数, 它倾向于定义有效的最近邻。其值通常

Model predictions across all the treatments

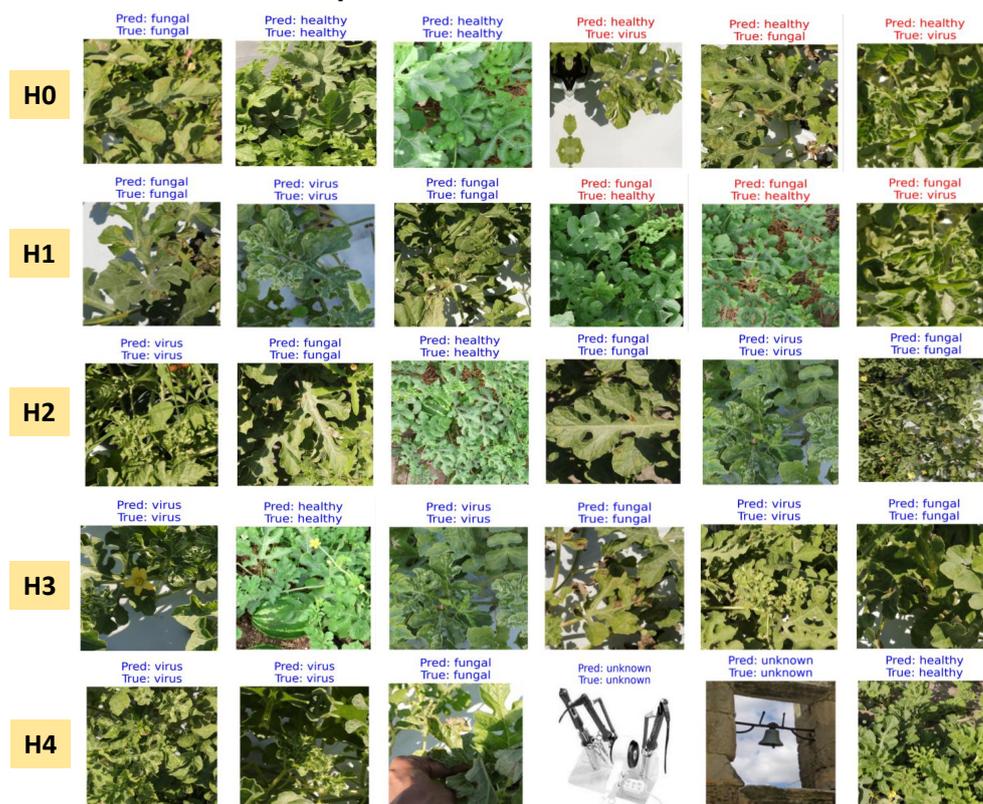


Fig. 8: 带有真实标签的分类测试图像。“Pred”对应于模型在测试集上的预测结果，而“True”表示真实标签的数据集。以下颜色代表预测：蓝色表示正确，红色表示不正确。

在 5-50 之间；然而，设置该值取决于所用数据集的大小。由于这项研究在一定程度上使用了大量的输入样本，因此将困惑度设定为 30 以理解明显的类可分性。另一方面，“n_components”决定了在 2D 还是 3D 空间中的可视化，因此其值设定为 2 用于 2D 聚类可视化。与 t-SNE 不同，UMAP 倾向于保持数据集的局部和全局结构。当真实和合成图像以 1:1 的比例合并时，t-SNE 和 UMAP 都保持了出色的类内外可分性。基于使用 t-SNE 和 UMAP 聚类测试获得的高维度降维聚类表示，很明显，将合成图像与真实图像结合能够让模型学习到更强健的特征，这是实际领域部署这些计算机视觉算法的重要需求。

5 局限性和未来的研究考虑

通过这项实验研究，观察到了几个可能影响整体指标准确性及模型分类性能的局限性。此外，还为感兴趣的研究者添加了未来研究方向的可能性。其中一些局限性和未来研究方向是：

1. 训练和测试数据集样本量的不确定性：尽管有大量数据集用于模型的训练和未见样本的测试，但仍不清楚增加图像数量，无论是在合成训练集还是实际训练集中，是否会提高模型性能。目前的实验设计只提供了一个固定的实际图像子集（H0 中的 750 张），因此此次实际图像对整体模型性能的影响仍不明确。未来的研究必须考虑同时扩大合成和实际数据集，或者以较小的步伐增加实际图像，比如按 1:1, 1:2, 1:3，如此类推直至 1:10 的比例。这应逐步完成，而不是直接从 1:1 跳到 1:10，以更精确地评估他们对模型鲁棒性和现实世界泛化能力的综合影响。
2. 缺乏外部验证和测试：所开发的模型仅在一个实验场所的图像上进行了测试。更广泛的适用性需要进行外部验证，使用从不同地区、季节、管理实践和成像设备独立收集的样本，例如来自具有多种疾病症状的商业农场样本。
3. 合成图像的不现实着色：本研究中使用的合成图像与病叶的着色不一致。例如，在图 2 中，实际图像（右框）是在阳光明媚的条件下拍摄的，叶子的颜色大多为浅绿色。然而，大多数合成图像的颜色是深绿色的。这可能影响了模型的准确性，但在本研究中未进行评估。这可能导致训练集中合成样本和测试集中的真实图像之间出现显著的域转换，从而导致在自然颜色变化下泛化能力下降。
4. 合成图像生成的多种疾病组合：在这项研究中，训练 SD 3.5M 模型的训练样本仅由单一类别的疾病组成，用于生成合成样本。例如，在图 3 中，仅使用炭疽病或霜霉病分类来生成相应的合成样本。然而，在实际田间条件下，情况可能并非如此，叶片表面可能会出现多种疾病症状。在这种情况下，SD 3.5M 模型很难捕获多种疾病症状的特定信息并生成真实的疾病例子。这是基于 GenAI 的文本到图像多模态模型面临的现有挑战之一，其中图像到图像的翻译可能更为合适。

6 结论

本研究提出了一项详尽的评估，测试基于生成式人工智能（GenAI）的合成图像在结合真实图像时对于提高西瓜（*Citrullus lanatus*）疾病分类的有效性。为进行这一分析，评估了五种处理方法。它们分别是：H0（仅真实图像）、H1（仅合成图像）、H2（1:1 真实与合成图像）、H3（1:10 真实与合成图像）和 H4（H3 + 未知类别）。所有处理方法均使用 EfficientNetV2-L 模型以及多种微调和迁移学习方法进行训练。

将少量真实图像与合成图像结合使用似乎能够提高模型的整体性能和泛化能力。这在真实图像与合成图像的比例为 1:1 和 1:10 时尤为明显。不仅指标性能有所提升，还增强了训练的 EfficientNetV2-L 模型在泛化性和特征类别可分性方面的整体表现。基于聚类分析和定量测量，显而易见合成图像无法替代真实图像。换句话说，尽管可以利用各种开放访问的生成 AI 模型大量生成合成图像，但使用各种传感系统获取的真实图像在叶片形态、症状表达的自然变异性和传感器特异噪声方面仍具有重要性。因此，专门在合成数据集上训练的 CNN 模型可能会因这些限制而受到影响，并由于域迁移而在实际田间环境中未能充分分类疾病。

Nitin Rai：数据整理，正式分析，调查，方法学，撰写初稿。 Nathan S. Boyd：资源，项目管理，验证，撰写审阅和编辑。 Gary E. Vallad：资源，项目管理，撰写审阅和编辑。 Arnold W. Schumann：概念化，方法学，监督，撰写审阅和编辑。

7

声明竞争性利益 作者声明在本研究中没有经济或利益冲突。

8

致谢 本研究得到了美国农业部 (USDA) 小企业创新研究 & 技术转让计划 (SBIR/STTR) 项目编号 # 2024-51402-42007 的资助。感谢 Emily Witt 和 Michael Sweat 在田间实验中给予的帮助。

References

- Ampatzidis Y.: UF Ag Engineer Develops Smart-Spray Technology to Help Reduce Fertilizer, Pesticides; 2022. Blog post. <https://blogs.ifas.ufl.edu/news/2022/01/05/uf-ag-engineer-develops-smart-spray-technology-to-help-reduce-fertilizer-pesticides/>.
- Bafghi RA, Harilal N, Monteleoni C, Raissi M. MixDiff: Mixing Natural and Synthetic Images for Robust Self-Supervised Representations. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) IEEE; 2025. p. 7500–7511.
- Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*. 2018;106:249–259.
- Davies DL, Bouldin DW. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*. 2009;(2):224–227.
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition Ieee; 2009. p. 248–255.
- Douarre C, Crispim-Junior CF, Gelibert A, Tougne L, Rousseau D. Novel data augmentation strategies to boost supervised segmentation of plant disease. *Computers and electronics in agriculture*. 2019;165:104967.
- Durham S.: New Sprayer Technology Reduces Pesticide Use; 2016. Research news article. <https://www.ars.usda.gov/news-events/news/research-news/2016/new-sprayer-technology-reduces-pesticide-use/>.
- Gomaa AA, Abd El-Latif YM. Early prediction of plant diseases using CNN and GANs. *International Journal of Advanced Computer Science and Applications*. 2021;12(5).
- Han G, Asiedu DKP, Bennin KE. Plant disease detection with generative adversarial networks. *Heliyon*. 2025;11(7).
- Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S, et al. Lora: Low-rank adaptation of large language models. *ICLR*. 2022;1(2):3.
- Keinath AP, Miller GA.: Watermelon Fungicide Guide for 2022 (Updated); 2022. Land-Grant Press by Clemson Extension, publication LGP 1001; updated from 2019 version. <https://open.clemson.edu/horticulture/22/>.

- Li H, Guo C, Yang Z, Chai J, Shi Y, Liu J, et al. Design of field real-time target spraying system based on improved YOLOv5. *Frontiers in Plant Science*. 2022;13:1072631.
- Lu Y, Chen D, Olaniyi E, Huang Y. Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. *Computers and Electronics in Agriculture*. 2022;200:107208.
- Mossler MA. Watermelon Pest Management Strategic Plan (PMSP): PI-52/PI089, 8/2005. EDIS. 2005;2005(8).
- Oberti R, Marchi M, Tirelli P, Calcante A, Iriti M, Tona E, et al. Selective spraying of grapevines for disease control using a modular agricultural robot. *Biosystems engineering*. 2016;146:203–215.
- Pallottino F, Violino S, Figorilli S, Pane C, Aguzzi J, Colle G, et al. Applications and perspectives of Generative Artificial Intelligence in agriculture. *Computers and Electronics in Agriculture*. 2025;230:109919.
- Rai N, Schumann A, Boyd N. PhytoSynth: Leveraging Multi-modal Generative Model for Crop Disease Data Generation with Novel Benchmarking and Prompt Engineering Approach. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*; 2025. p. 5371–5380.
- Ravikumara BM, Manjunatha L, Naik MK, Sathyanarayana N, Kittan S, Nagaraju DK, et al. Watermelon Mosaic Virus. In: *Compendium of Phytopathogenic Microbes in Agro-Ecology: Vol. 2 Viruses and Viroids*. Springer; 2025. p. 393–414.
- Roberts PD, Vallad G, Zhang S, Dufault N, Paret M. Anthracnose on Cucurbits in Florida. EDIS. 2023;2023(3).
- Robinson C, Trivedi A, Blazes M, Ortiz A, Desbiens J, Gupta S, et al. Deep learning models for COVID-19 chest x-ray classification: Preventing shortcut learning using feature disentanglement. *medRxiv*. 2021;.
- Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*. 1987;20:53–65.
- Ruiz N, Li Y, Jampani V, Pritch Y, Rubinstein M, Aberman K. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2023. p. 22500–22510.
- Shoab M, Sadeghi-Niaraki A, Ali F, Hussain I, Khalid S. Leveraging deep learning for plant disease and pest detection: A comprehensive review and future directions. *Frontiers in Plant Science*. 2025;16:1538163.

- Singh K, Navaratnam T, Holmer J, Schaub-Meyer S, Roth S. Is synthetic data all we need? benchmarking the robustness of models trained with synthetic images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024. p. 2505–2515.
- Sun R, Zhang M, Yang K, Liu J. Data enhancement for plant disease classification using generated lesions. *Applied Sciences*. 2020;10(2):466.
- Venkateswara SM, Padmanabhan J. Deep learning based agricultural pest monitoring and classification. *Scientific Reports*. 2025;15(1):8684.
- Wang X, Wang R, Wu S, Ma X, Du J, Gao J. A Stable Diffusion Based Image Generation Method for Few-Shot Emerging Pest Detection in the Wild. In: 2024 IEEE International Conference on Industrial Technology (ICIT) IEEE; 2024. p. 1–6.
- Weber C, Kramer J.: Most U.S. Watermelon Is Produced in the South, with Florida Leading Output in 2021; 2022. Economic Research Service, U.S. Department of Agriculture, “Charts of Note,” August 3, 2022. <https://ers.usda.gov/data-products/charts-of-note/chart-detail?chartId=104374>.
- Yang Y, Mali P, Arthur L, Molaei F, Atsyo S, Geng J, et al. Advanced technologies for precision tree fruit disease management: A review. *Computers and Electronics in Agriculture*. 2025;229:109704.
- Zeng Q, Ma X, Cheng B, Zhou E, Pang W. Gans-based data augmentation for citrus disease severity detection using deep learning. *IEEE Access*. 2020;8:172882–172891.