Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Cross-Dataset Strawberry Disease Classification from Controlled to African Field Conditions

Isibor Kennedy Ihianle

isibor.ihianle@ntu.ac.uk

Nottingham Trent University https://orcid.org/0000-0001-7445-8573

Pedro Machado

Nottingham Trent University https://orcid.org/0000-0003-1760-3871

Jordan J. Bird

Nottingham Trent University

Salisu W. Yahaya

Nottingham Trent University

Ahmad Lotfi

Nottingham Trent University

Everistus Nwogo

Nottingham Trent University

Isaac Akinwumi

Covenant University

Jonathan Oluranti

Covenant University

Abiodun Adebayo

Covenant University

Abdulwahab Sulaiman Hamza

Ashley Strawberry Farm

Ashley Dimka-Tapgun

Ashley Strawberry Farm

Research Article

Keywords: Strawberry disease, Cross-dataset evaluation, Machine learning, Deep learning, Semi-supervised learning, MobileNetV3-Small, MobileViTv2-0.5, EfficientNet-B0, ConvNeXt-Tiny, Smart agriculture

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10938933/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Cross-Dataset Strawberry Disease Classification from Controlled to African Field Conditions

Isibor Kennedy Ihianle[a,*], Pedro Machado[a], Jordan J. Bird[a], Salisu Yahaya[a], Ahmad Lotfi[a], Everistus Nwogo[a], Isaac Akinwumi[b], Jonathan Oluranti[b], Abiodun Adebayo[b], Abdulwahab Sulaiman Hamza[c] and Ashley Dimka-Tapgun[c]

[a] *Nottingham trent University, Nottingham, NG11 8NS, United Kingdom*
[b] *Covenant University, Ota, Ogun State, Nigeria*
[c] *Ashley's Strawberry Farms, Jos, Plateau State, Nigeria*

ARTICLE INFO

Keywords:
Strawberry disease
Cross-dataset evaluation
Machine learning
Deep learning
Semi-supervised learning
MobileNetV3-Small
MobileViTv2-0.5
EfficientNet-B0
ConvNeXt-Tiny
Smart agriculture

Abstract

Image-based plant-disease classifiers often achieve high accuracy under controlled conditions, but their reliability can deteriorate in real-world field environments. This article investigates cross-dataset strawberry leaf-health classification using controlled PlantVillage images and field images collected through the SmartBerry project in Nigeria. Four deep-learning architectures were evaluated under within-domain, controlled-to-field and pooled-training conditions using healthy-leaf and dark-spot/scorch symptoms. Although all models achieved near-ceiling performance on PlantVillage, direct transfer to SmartBerry reduced macro F1-score to between 0.418 and 0.824, demonstrating a substantial gap between controlled-dataset performance and field reliability. To reduce the amount of locally labelled field data required to address this gap, this paper proposes SmartBerry Domain Adaptation (SmartBerry-DA). This label-efficient adaptation approach combines a model trained on controlled images with a small set of labelled SmartBerry images and additional unlabelled field images. Using MobileViTv2-0.5, SmartBerry-DA achieved a macro F1-score of 0.940 ± 0.015 with only five labelled SmartBerry images per class and 0.988 ± 0.012 with ten, compared with 0.994 ± 0.008 when the full labelled target dataset was used. This shows that strong field performance can be recovered with substantially reduced local annotation. For lightweight offline deployment, 16-bit floating-point representation reduced model size from 4.37 MB to 2.30 MB without changing predictive performance under the evaluated conditions. The findings demonstrate the importance of locally acquired field data and label-efficient adaptation for transferring agricultural AI from benchmark datasets to practical farming environments, while providing a pathway towards lightweight strawberry disease recognition in settings where extensive annotation and continuous connectivity may not be feasible.

1. Introduction

Strawberry (*Fragaria* × *ananassa*) production depends on maintaining healthy foliage throughout vegetative growth and fruit development. Foliar disorders can reduce photosynthetic area, weaken plant vigour, and compromise fruit quality or marketability (Fanourakis, Makraki, Spyrou, Karavidas, Tsaniklidis and Ntatsi, 2025). Strawberry leaf scorch presents visible lesions and progressive tissue damage, but reliable recognition can be difficult when symptoms are early, partially occluded, photographed under uneven illumination, or confused with nutrient stress and other forms of leaf discolouration. Earlier recognition can support timely crop inspection and more targeted management, provided that the recognition system remains reliable under the conditions in which it is used.

*Corresponding author
isibor.ihianle@ntu.ac.uk (I.K. Ihianle); pedro.machado@ntu.ac.uk (P. Machado); jordan.bird@ntu.ac.uk (J.J. Bird); salisu.yahaya@ntu.ac.uk (S. Yahaya); ahmad.lotfi@ntu.ac.uk (A. Lotfi); everistus.nwogo@ntu.ac.uk (E. Nwogo); isaac.akinwumi@covenantuniversity.edu.ng (I. Akinwumi); jonathan.oluranti@covenantuniversity.edu.ng (J. Oluranti); abiodun.adebayo@covenantuniversity.edu.ng (A. Adebayo); ibn.abdulwahabu@gmail.com (A.S. Hamza); ashleysstrawberryfarms@gmail.com (A. Dimka-Tapgun)
ORCID(s):

Deep learning (DL) techniques have become widely used for image-based plant-disease classification, supported by benchmark datasets such as PlantVillage (Hughes, Salathé et al., 2015; Mohanty, Hughes and Salathé, 2016). Although models trained on these datasets often achieve high accuracy within a dataset. PlantVillage images were predominantly acquired under controlled conditions, with centred leaves, relatively uniform backgrounds, and limited acquisition variability. *In the wild* field images may instead contain complex backgrounds, shadows, occlusion, variable viewpoints, compression, and heterogeneous camera characteristics. Consequently, performance on controlled datasets may not reliably indicate performance under real agricultural conditions (Mohanty et al., 2016).

This controlled-to-field problem is particularly relevant to smart agriculture in sub-Saharan Africa, where digital technologies can support crop monitoring and farm-level decision making, but practical deployment may be constrained by connectivity, infrastructure, affordability and access to locally annotated data (ITU, Food and of the United Nations, 2022; Choruma, Dirwai, Mutenje, Mustafa, Chimonyo, Jacobs-Mata and Mabhaudhi, 2024). These constraints vary between farming systems, but motivate agricultural artificial intelligence (AI) approaches that are computationally efficient, capable of offline operation, and adaptable using limited local supervision.

The SmartBerry project provides an empirical setting for investigating these issues within strawberry production in Nigeria. Original SmartBerry field images capture natural variation in illumination, background, viewpoint, image quality and symptom presentation. PlantVillage is therefore used as the controlled source domain and SmartBerry as the Nigerian field target domain. The datasets are first independently evaluated using their original multiclass label spaces. Cross-dataset evaluation then uses symptom-compatible strawberry classes comprising healthy leaves and dark-spot/scorch symptoms. This correspondence is defined at the level of visible symptoms and does not assume pathogen-level equivalence between PlantVillage leaf scorch and SmartBerry leaf dark spot.

Four architectures - MobileNetV3-Small, EfficientNet-B0, MobileViTv2-0.5 and ConvNeXt-Tiny are evaluated to characterise predictive performance and computational efficiency. Direct PlantVillage-to-SmartBerry transfer is then compared with SmartBerry-only and pooled training to quantify the effect of domain shift. To reduce dependence on extensive local annotation, SmartBerry-DA is proposed as a label-efficient semi-supervised domain-adaptation framework combining supervised source and target learning, Correlation Alignment (CORAL) feature alignment, an exponential moving average (EMA) teacher, confidence-filtered pseudo-labels and consistency regularisation. The method is evaluated across restricted SmartBerry annotations with component ablation and cross-backbone sensitivity analysis.

This paper extends earlier SmartBerry-supported work on multi-dataset plant-disease classification (Krishna, Machado, Otuka, Yahaya, Neves dos Santos and Ihianle, 2025; Ihianle, 2024). Unlike our previous work, this paper focuses on original Nigerian strawberry field imagery, controlled-to-field transfer, label-efficient local adaptation and lightweight offline deployment. Accordingly, this paper asks how strongly predictive performance deteriorates when controlled-image models are transferred to SmartBerry field data; how architecture affects predictive behaviour and computational efficiency across domains; whether SmartBerry-DA can recover field performance using limited labelled SmartBerry images and how this changes with annotation budget and backbone; and what accuracy-efficiency trade-offs arise when the selected model is converted for lightweight offline inference and evaluated under representative image degradation. The principal contributions of this paper are:

- a controlled-to-field strawberry classification benchmark combining PlantVillage with original field imagery collected under Nigerian farming conditions, including a symptom-compatible task for direct cross-dataset evaluation;
- an empirical analysis of controlled-to-field domain shift using overall performance, class-specific behaviour and confidence calibration;
- SmartBerry-DA, a label-efficient semi-supervised domain-adaptation approach evaluated across multiple target-label budgets, with component ablation and cross-backbone sensitivity analysis; and
- a lightweight offline deployment evaluation using FP32, FP16 and INT8 representations, considering model size, inference latency, prediction agreement and robustness to simulated field-image degradation.

This article is organised as follows. Section 2 reviews related work on controlled and field plant-disease recognition, strawberry disease classification, domain adaptation, lightweight deployment, and smart agriculture in Nigeria and sub-Saharan Africa. Section 3 presents the proposed SmartBerry framework, including the datasets, SmartBerry-DA, model architectures, deployment pathway, and the experimental protocol. Section 4 reports the benchmarking, cross-dataset generalisation, label-efficient domain adaptation, deployment and robustness experiments, followed by discussion of the findings and limitations. Section 5 concludes the paper.

2. Related Work

This section reviews the literature relevant to the controlled-to-field strawberry classification problem addressed in this paper. It considers controlled and field plant-disease datasets, strawberry disease recognition, domain adaptation and label-efficient learning, lightweight deployment and uncertainty, and the wider relevance of these technologies to smart agriculture in Nigeria and sub-Saharan Africa. The review concludes by positioning this paper within these research areas and identifying the specific gap addressed by SmartBerry-DA.

2.1. Controlled and Field Plant-Disease Recognition

Deep learning methods have achieved strong results in image-based plant-disease classification, with PlantVillage becoming one of the most widely used benchmark datasets (Mohanty et al., 2016; Hughes et al., 2015; Too, Li, Kwao, Njuki, Mosomi and Kibet, 2019). PlantVillage provides more than 50,000 healthy and diseased leaf images across multiple crop–disease categories, but its relatively controlled backgrounds and acquisition conditions differ substantially from routine field imagery. Mohanty et al. (Mohanty et al., 2016) already reported a substantial reduction in performance when PlantVillage-trained models were applied to images acquired outside the original dataset. Subsequent datasets have attempted to capture more realistic acquisition conditions. PlantDoc contains natural-occurring backgrounds and visual clutter (Singh, Jain, Jain, Kayal, Kumawat and Batra, 2020), while the Plant Pathology dataset includes field images acquired under variation in illumination, viewpoint, and image quality (Thapa, 2023). Barbedo (Barbedo, 2026) similarly identified background complexity, symptom variability, image quality and dataset composition as important factors affecting plant-disease recognition. These studies reinforce the distinction between high within-dataset accuracy and generalisation to a separately acquired field domain.

Recent works have explicitly investigated this problem as domain shift. Wu et al. (Wu, Fan, Luo, Choudhury, Tjahjadi and Hu, 2023) evaluated unsupervised domain adaptation from laboratory to field plant-disease imagery and showed that differences in acquisition conditions can substantially affect transfer performance. This paper follows this controlled-to-field perspective but uses PlantVillage and independently collected SmartBerry images from Nigeria as the source and target domains, respectively.

2.2. Strawberry Disease Recognition under Field Conditions

Computer vision in strawberry production has addressed fruit detection, counting, segmentation and harvesting as well as plant-health recognition. Deep learning has also been applied to the recognition of strawberry plant structures, including the classification of trusses and runners (Pomykala, De Lemos, Ihianle, Adama and Machado, 2022). The StrawDI dataset, for example, contains strawberry images acquired under commercial production conditions for instance segmentation (Pérez-Borrero, Marín-Santos, Gegúndez-Arias and Cortés-Ancos, 2020). Although relevant to field vision, these applications do not directly address foliar disease transfer between controlled and field datasets.

Disease-focused studies have demonstrated the feasibility of deep learning for strawberry health assessment. Abbas et al. (Abbas, Liu, Amin, Tariq and Tunio, 2021) evaluated several convolutional architectures for real-time strawberry leaf scorch recognition under field lighting conditions. Hu et al. (Hu, Wang, Du, Hu, Jiao and Xu, 2023) subsequently addressed strawberry disease classification using field and web-derived images and showed that complex backgrounds and visually similar disease symptoms remain important sources of classification difficulty. Other work has investigated disease severity and lesion-oriented recognition, further demonstrating the importance of field-specific visual information in strawberry diagnosis (Abbas et al., 2021; Hu et al., 2023).

These studies establish the relevance of strawberry disease recognition in natural environments, but the present work addresses a different question: whether knowledge learned from a controlled benchmark transfers to original Nigerian field imagery, and how much locally labelled data is required to adapt the classifier when direct transfer is insufficient.

2.3. Domain Adaptation and Label-Efficient Learning

Domain adaptation addresses changes between the distribution used for model training and that encountered in the target environment. CORAL reduces this discrepancy by aligning second-order feature statistics between source and target domains (Sun and Saenko, 2016). Its deep formulation provides a simple non-adversarial mechanism for incorporating feature alignment during network optimisation. When target labels are scarce, semi-supervised learning provides an additional mechanism for exploiting unlabelled target data. Mean Teacher uses an exponential moving average of model weights to provide more stable training targets (Tarvainen and Valpola, 2017). Confidence-based pseudo-labelling and consistency regularisation have likewise been shown to make effective use of unlabelled images; FixMatch, for example, retains high-confidence pseudo-labels generated from weakly transformed images and enforces them under stronger perturbations (Sohn, Berthelot, Carlini, Zhang, Zhang, Raffel, Cubuk, Kurakin and Li, 2020). Similar source-to-field adaptation principles have been applied to plant-disease recognition (Wu et al., 2023).

SmartBerry-DA combines these complementary ideas for the present controlled-to-field problem: supervised source and limited target learning, CORAL alignment, an EMA teacher, confidence-filtered pseudo-labels and teacher - student consistency. Its distinguishing experimental focus is the amount of labelled SmartBerry data required for adaptation and the extent to which the resulting behaviour depends on the selected backbone.

2.4. Lightweight Models, Uncertainty and Deployment

Agricultural vision systems intended for local inference must consider model complexity in addition to predictive performance. MobileNetV3 uses hardware-aware architecture search and efficient convolutional operations for mobile inference (Howard, Sandler, Chu, Chen, Chen, Tan, Wang, Zhu, Pang, Vasudevan et al., 2019); EfficientNet employs compound scaling of network depth, width and resolution (Tan and Le, 2019); and MobileViT combines convolutional processing with lightweight self-attention (Mehta and Rastegari, 2022b,a). ConvNeXt provides a higher-capacity convolutional reference based on modernised architectural design (Liu, Mao, Wu, Feichtenhofer, Darrell and Xie, 2022).

Model confidence is also relevant when performance changes across domains. Deep neural networks can produce poorly calibrated probabilities (Guo, Pleiss, Sun and Weinberger, 2017), while risk–coverage analysis provides a means of examining whether uncertainty estimates meaningfully rank reliable and unreliable predictions (Geifman, Uziel and El-Yaniv, 2018). These considerations are particularly important under domain shift, where confidence behaviour learned from source-domain images cannot be assumed to transfer unchanged.

Deployment efficiency can be improved through reduced-precision inference, although quantisation may introduce a trade-off between computational cost and predictive fidelity (Jacob, Kligys, Chen, Zhu, Tang, Howard, Adam and Kalenichenko, 2018). Accordingly, this paper evaluates FP32, FP16 and INT8 representations using model size, inference latency, prediction agreement and predictive performance rather than assuming that greater numerical compression necessarily produces a superior deployment model.

2.5. Smart Agriculture in Nigeria and Sub-Saharan Africa

The controlled-to-field problem is also relevant to the wider context of digital agriculture in sub-Saharan Africa. Digital technologies can support agricultural monitoring, advisory services and farm-level decision making, but infrastructure, connectivity, affordability and digital capacity remain uneven across the region (ITU et al., 2022; Ayim, Kassahun, Addison and Tekinerdogan, 2022; Choruma et al., 2024). The FAO-ITU assessment of 47 sub-Saharan African countries identifies digital infrastructure and access as important conditions for expanding digital agriculture (ITU et al., 2022), while recent regional evidence similarly highlights connectivity, affordability and digital literacy as barriers to wider use (Choruma et al., 2024).

Nigeria reflects both the opportunity and the constraints associated with this transition. Empirical studies have examined the adoption of e-agriculture and agricultural smartphone applications in Nigeria (Eweoya, Okuboyejo, Odetunmibi and Odusote, 2021; Kolapo and Didunyemi, 2024). Evidence from smallholder farmers in southwest Nigeria shows that exposure to agricultural smartphone applications can influence their uptake, while access to appropriate devices and connectivity remains relevant for their use (Kolapo and Didunyemi, 2024). These findings provide a practical motivation for agricultural AI that minimises dependence on continuous connectivity and large locally annotated datasets.

The SmartBerry setting contributes to this literature through original strawberry imagery acquired within Nigerian agricultural production. The present work does not assume that one Nigerian dataset represents the diversity of African agriculture. Instead, it provides a geographically grounded case study in which controlled-to-field generalisation, local annotation requirements and lightweight offline inference can be evaluated within a sub-Saharan African smart-agriculture context.

2.6. Research Gap and Paper Positioning

Existing research has established strong plant-disease classification on controlled datasets, developed more realistic field datasets, investigated the autonomous recognition of strawberry diseases in natural environments, and proposed methods for domain adaptation and lightweight inference (Mohanty et al., 2016; Singh et al., 2020; Abbas et al., 2021; Hughes et al., 2015; Wu et al., 2023). However, limited evidence jointly examines controlled-to-field strawberry transfer using independently collected African field imagery, restricted target-label budgets, uncertainty behaviour and lightweight model conversion. This paper addresses this intersection using PlantVillage and SmartBerry rather than treating high within-dataset accuracy as sufficient evidence of field generalisation. Table 1 positions this paper relative to conventional controlled-dataset evaluation and earlier SmartBerry-supported work.

3. The Proposed Architecture

The proposed SmartBerry framework addresses the distribution shift between controlled strawberry-leaf images and images acquired under real farming conditions. It combines controlled PlantVillage images with field-derived SmartBerry images to support model benchmarking, cross-dataset evaluation, label-efficient domain adaptation, and lightweight deployment. PlantVillage provides the controlled source domain, while SmartBerry represents the target field domain. The distinction is important because models trained on curated datasets can perform strongly within-domain yet degrade substantially when exposed to changes in illumination, background, image quality, camera position, and other acquisition conditions (Mohanty et al., 2016; Hughes et al., 2015). As illustrated in Fig. 1, images are first resized, normalised, and augmented before being processed by one of four candidate backbones: MobileNetV3-Small, EfficientNet-B0, MobileViTv2-0.5, or ConvNeXt-Tiny. The models are initially benchmarked independently on the PlantVillage and SmartBerry multiclass datasets. Cross-dataset evaluation is then performed using the symptom-compatible classes shared between the two datasets, with PlantVillage acting as the source domain and SmartBerry as the field target domain. The proposed SmartBerry-DA method is subsequently used to adapt the selected lightweight backbone using a small labelled SmartBerry subset together with the remaining unlabelled target images. SmartBerry-DA combines supervised source and target learning with CORAL feature alignment, EMA teacher pseudo-labelling, consistency regularisation, and uncertainty-based pseudo-label filtering. The resulting model is finally evaluated for lightweight offline deployment using model size, inference latency, quantisation behaviour, and robustness to degraded field-image conditions. The final classifier produces a strawberry health prediction together with its associated confidence. Uncertainty estimates are used to assess prediction reliability and to support the identification of cases that may require image recapture or further inspection.

3.1. Datasets

Two image datasets are used in this paper: PlantVillage, representing controlled image-acquisition conditions, and the SmartBerry dataset, representing real strawberry-growing conditions in Nigeria. The datasets differ substantially in acquisition setting, background complexity, illumination, viewpoint and image quality. Table 2 summarises their principal characteristics, while Fig. 2 illustrates representative samples from the SmartBerry field dataset.

Table 1
Positioning of this paper relative to controlled-dataset and earlier SmartBerry-supported plant-disease classification.

Dimension	Typical controlled-dataset study	Earlier SmartBerry-supported study (Krishna et al., 2025)	This paper
Data context	Predominantly curated or controlled imagery	Aggregated multi-crop and web-sourced imagery	Controlled PlantVillage and original Nigerian SmartBerry field imagery
Primary task	Within-dataset disease classification	Multi-crop classification across datasets	Strawberry controlled-to-field classification using symptom-compatible labels
Evaluation	Random or within-dataset hold-out	Within- and cross-dataset comparison	Within-domain, direct controlled-to-field, target-only and pooled evaluation
Adaptation	Standard transfer learning	Not a central objective	Label-efficient SmartBerry-DA using labelled and unlabelled target data
Local annotation	Usually not quantified	Not evaluated	Target budgets of 5, 10, 20 and 50 labelled images per class
Deployment	Often absent or simulated	Not a primary contribution	FP32, FP16 and INT8 conversion with reference CPU latency and robustness evaluation

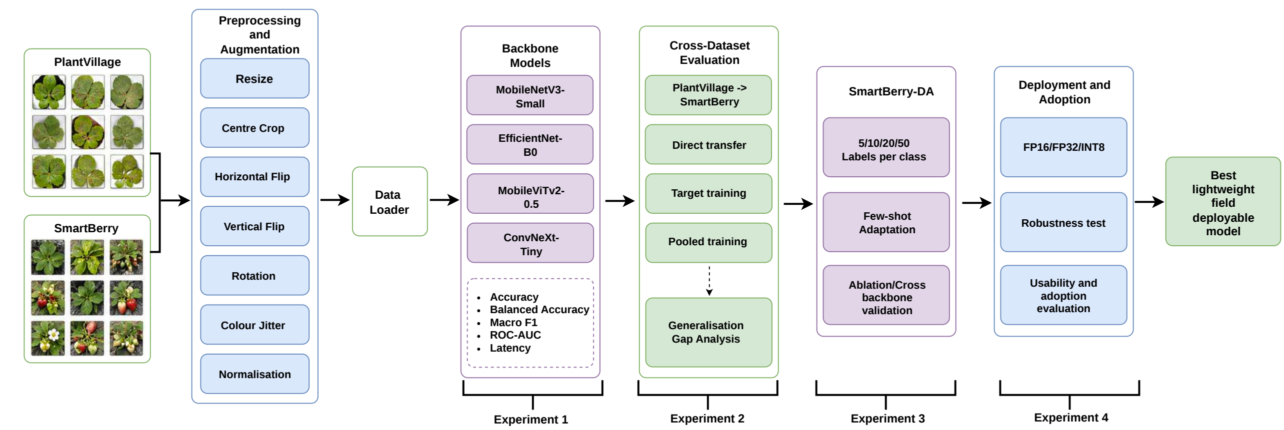
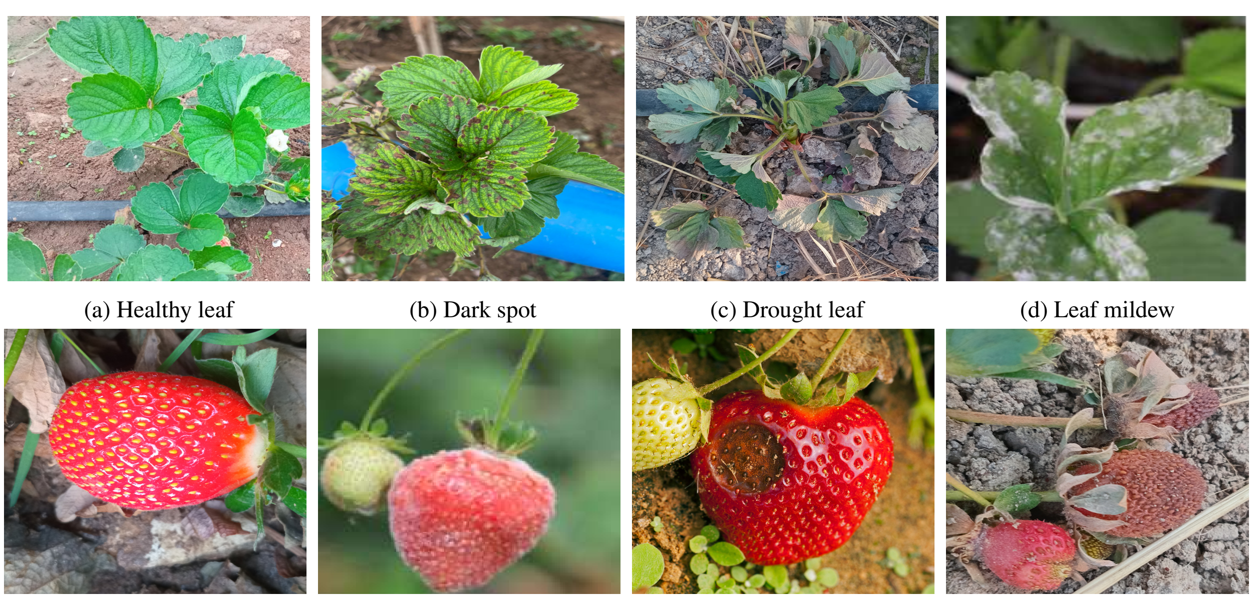


Figure 1: Conceptual overview of the SmartBerry framework, showing the progression from PlantVillage and SmartBerry image preparation through model benchmarking, cross-dataset evaluation, SmartBerry-DA adaptation, and lightweight field deployment.



(a) Healthy leaf (b) Dark spot (c) Drought leaf (d) Leaf mildew

(e) Ripe fruit (f) Fruit mildew (g) Fruit rot (h) Drought fruit

Figure 2: Representative samples from the SmartBerry field dataset, showing selected healthy, diseased and stress-related strawberry leaf and fruit conditions. The examples illustrate the diversity of symptoms, backgrounds, viewpoints and image-acquisition conditions represented in the dataset.

Table 2
Summary of the image datasets used in the paper.

Dataset	Acquisition domain	Number of classes	Strawberry classes of interest
PlantVillage	Controlled	38	Healthy/leaf scorch
SmartBerry	Field	9	Healthy leaves/leaf dark spot

3.1.1. PlantVillage Controlled Dataset

PlantVillage is a publicly available plant-health image dataset containing healthy and diseased leaves from multiple crop species and disease categories (Mohanty et al., 2016; Hughes et al., 2015).[1] The version used in this paper contains 38 classes and is characterised predominantly by individual leaves photographed under relatively consistent illumination and against comparatively uniform backgrounds. Within PlantVillage, the strawberry images comprise two relevant classes: *strawberry healthy* and *strawberry leaf scorch*. These provide the controlled strawberry image domain used in this paper.

3.1.2. SmartBerry Field Dataset

The SmartBerry dataset is an original collection of strawberry images acquired under real agricultural conditions in Nigeria. It contains nine classes representing healthy, diseased and stress-related strawberry conditions. Unlike PlantVillage, the images were captured in natural production environments and therefore contain substantial variation in illumination, background, camera angle, plant position, image quality, symptom severity and partial occlusion. Representative SmartBerry samples are shown in Fig. 2. The examples illustrate the visual diversity of the field dataset across leaf and fruit conditions and the variation encountered under natural image-acquisition conditions. Two SmartBerry classes provide symptom-level correspondence with the PlantVillage strawberry subset: *healthy leaves* and *leaf dark spot*. This correspondence is defined at the level of visible symptoms and does not assume pathogen-level equivalence between PlantVillage leaf scorch and SmartBerry leaf dark spot.

The SmartBerry dataset is publicly available through Zenodo, with a Kaggle mirror provided for convenient access.[2]

The controlled PlantVillage and field-derived SmartBerry datasets therefore represent distinct acquisition domains. Their principal characteristics are summarised in Table 2. For direct comparison between the strawberry domains, the corresponding healthy and dark-spot/scorch symptom classes form the harmonised subset shown in Table 3. This provides a common two-class label space while preserving the distinction between controlled and field image acquisition.

[1] PlantVillage dataset: https://github.com/spMohanty/PlantVillage-Dataset

[2] SmartBerry dataset: Zenodo: https://zenodo.org/records/22056372; Kaggle: https://www.kaggle.com/datasets/isiborihianle/smartberry-strawberry-field-image-dataset

Table 3
Class distribution of the harmonised strawberry subset.

Dataset	Healthy	Dark-spot/scorch symptom	Total
PlantVillage	456	1109	1565
SmartBerry	125	253	378

3.2. SmartBerry-DA Framework

SmartBerry-DA is a label-efficient domain-adaptation framework designed to adapt a classifier trained on controlled PlantVillage images to SmartBerry field conditions using only a small amount of labelled target data. The framework combines supervised learning on labelled source and target images with unlabelled SmartBerry images through an EMA teacher - student architecture.

The student network is optimised using four complementary components: supervised source and target classification, CORAL-based feature alignment, uncertainty-filtered pseudo-label learning, and teacher- student consistency regularisation. The EMA teacher provides predictions for unlabelled SmartBerry images, while only sufficiently confident pseudo-labels are retained for training. CORAL reduces differences between source and target feature distributions, and consistency regularisation encourages stable predictions under image perturbations. The training objective is

$$\mathcal{L}_{\mathrm{DA}} = \lambda_s \mathcal{L}_s + \lambda_t \mathcal{L}_t + r(e) \left(\lambda_c \mathcal{L}_{\mathrm{CORAL}} + \lambda_p \mathcal{L}_{\mathrm{pseudo}} + \lambda_u \mathcal{L}_{\mathrm{cons}} \right), \tag{1}$$

where $\mathcal{L}_s$ and $\mathcal{L}_t$ denote supervised source and target classification losses, respectively; $\mathcal{L}_{\mathrm{CORAL}}$ aligns source and target feature statistics; $\mathcal{L}_{\mathrm{pseudo}}$ uses accepted teacher-generated pseudo-labels; and $\mathcal{L}_{\mathrm{cons}}$ enforces prediction consistency. The term $r(e)$ gradually introduces the adaptation losses during training. This formulation allows the model to exploit the larger controlled source dataset while progressively adapting its representation to field imagery using limited labelled and additional unlabelled SmartBerry data.

3.3. Model Architectures

Four image-classification backbones are considered: MobileNetV3-Small, EfficientNet-B0, MobileViTv2-0.5 and ConvNeXt-Tiny. The architectures were selected to provide complementary accuracy-efficiency characteristics, ranging from models designed explicitly for mobile inference to a higher-capacity convolutional reference. Their respective roles are summarised in Table 4.

MobileNetV3-Small: MobileNetV3-Small is a compact convolutional neural network developed for resource-constrained

Table 4
Image-classification architectures considered in this paper.

Model	Architecture	Role
MobileNetV3-Small	Lightweight CNN	Mobile baseline
EfficientNet-B0	Lightweight CNN	Accuracy–efficiency baseline
MobileViTv2-0.5	Lightweight hybrid	Primary lightweight candidate
ConvNeXt-Tiny	Higher-capacity CNN	Reference model

and mobile inference. Its architecture combines depthwise separable convolution with hardware-aware neural architecture search and lightweight attention mechanisms to reduce computational cost (Howard et al., 2019). It is included as the mobile baseline because its small parameter and memory requirements provide a useful reference for the minimum computational footprint achievable among the evaluated models.

EfficientNet-B0: EfficientNet-B0 is a compact convolutional architecture based on compound scaling of network depth, width and input resolution (Tan and Le, 2019). It provides a stronger-capacity lightweight CNN while retaining substantially lower computational requirements than larger convolutional networks. It is therefore included as an accuracy-efficiency baseline against which the more aggressively compact architectures can be compared.

MobileViTv2-0.5: MobileViTv2 combines convolutional feature extraction with lightweight self-attention, allowing local image structure and longer-range spatial relationships to be represented within a compact architecture (Mehta and Rastegari, 2022b; He, Yuan, Sun, Li, Liu, Ye and Sun, 2026). The 0.5-width configuration was selected as the lightweight hybrid candidate because it provides a different representation strategy from the purely convolutional baselines while maintaining a small model footprint suitable for edge-oriented deployment.

ConvNeXt-Tiny: ConvNeXt-Tiny is a modern convolutional architecture that incorporates design principles associated with contemporary vision models while retaining a fully convolutional structure (Liu et al., 2022). Compared with the three lightweight candidates, it has substantially greater model capacity. It is therefore used as a higher-capacity reference to assess the extent to which predictive performance changes when computational constraints are relaxed, rather than as the primary candidate for low-resource deployment.

The backbone used for the main SmartBerry-DA analysis is determined from the comparative evaluation rather than assumed solely from model size or architecture type. This allows predictive performance and computational efficiency to be considered jointly when selecting the model for subsequent adaptation and deployment.

3.4. Field Deployment Pathway

The deployment pathway targets local inference in agricultural environments where continuous network connectivity cannot be assumed. The selected lightweight SmartBerry-DA model is therefore evaluated for offline operation, model storage requirements, inference latency and robustness to representative field-image degradation. The deployment evaluation considers FP32, FP16 and INT8 model representations together with changes in predictive performance following model conversion. Robustness is assessed under reduced illumination, shadow, blur, JPEG compression, partial occlusion and reduced image resolution. The image-recognition component is evaluated independently of the wider SmartBerry advisory system. This separation allows the reliability and computational characteristics of the classifier to be assessed without confounding them with downstream agronomic advice generation.

3.5. Data Partitioning and Experimental Protocol

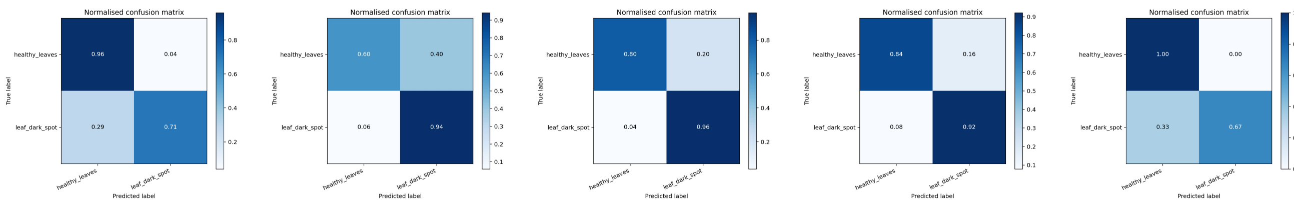
Dataset partitions were defined before model training and retained consistently within each experimental stage. PlantVillage was partitioned using stratified sampling, while SmartBerry partitions were constructed to minimise leakage between closely related field images. Where acquisition metadata permitted, images originating from the same plant or acquisition sequence were kept within the same partition. For cross-dataset evaluation, the same held-out SmartBerry test set was used across comparable training conditions. Test images were excluded from model selection, adaptation and hyperparameter optimisation, ensuring that performance differences reflected the training strategy rather than changes in the evaluation data. Models were initialised from ImageNet-pretrained weights unless otherwise stated, with the classification head adapted to the corresponding multiclass or binary task. Training settings, including input resolution, optimisation schedule, annotation budget and random seeds, are reported with the corresponding experimental procedure. Validation macro F1-score was used for model and hyperparameter selection where a validation set was available.

3.6. Evaluation Metrics

Predictive performance is assessed primarily using macro F1-score and balanced accuracy to account for class imbalance and class-specific performance. Accuracy and ROC–AUC are reported as complementary measures. For cross-domain evaluation, class-specific recall and specificity are additionally examined to identify asymmetric prediction behaviour. Prediction reliability is evaluated using Expected Calibration Error (ECE) (Guo et al., 2017) and, where uncertainty-based ranking is considered, the area under the risk-coverage curve (AURC), following the selective prediction framework (Geifman et al., 2018). For SmartBerry-DA, pseudo-label acceptance rate and accepted-label accuracy are additionally used to assess the reliability of unlabelled target supervision. Computational performance is characterised using parameter count, model size, median and 95th-percentile inference latency, and memory demand. Deployment evaluation additionally considers prediction agreement following model conversion and changes in macro F1-score under quantisation and simulated field-image degradation. Unless otherwise stated, results are reported as mean and standard deviation across repeated random-seed runs.

Table 5
PlantVillage multiclass performance and computational characteristics. Values are mean ± standard deviation across five seeds.

Model	Accuracy	Balanced accuracy	Macro F1	ROC-AUC	Parameters	Size	GPU latency
ConvNeXt-Tiny	0.998 ± 0.001	0.997 ± 0.002	0.997 ± 0.002	1.000 ± 0.000	27.85 M	106.29 MB	6.03 ± 0.09 ms
MobileNetV3-Small	0.996 ± 0.001	0.995 ± 0.001	0.995 ± 0.001	1.000 ± 0.000	1.56 M	6.05 MB	5.77 ± 0.05 ms
EfficientNet-B0	0.993 ± 0.003	0.991 ± 0.004	0.990 ± 0.004	0.999 ± 0.000	4.06 M	15.74 MB	8.49 ± 0.12 ms
MobileViTv2-0.5	0.999 ± 0.001	0.999 ± 0.001	0.998 ± 0.002	1.000 ± 0.000	1.12 M	4.39 MB	8.79 ± 0.12 ms



(a) Seed 11 (b) Seed 22 (c) Seed 33 (d) Seed 44 (e) Seed 55

Figure 3: Row-normalised confusion matrices for ConvNeXt-Tiny under direct PlantVillage-to SmartBerry transfer across five random seeds.

4. Experiments and Results

Four experiments evaluate the proposed framework from model selection to field deployment. Experiment 1 compares candidate architectures under within-dataset conditions; Experiment 2 quantifies controlled-to-field generalisation; Experiment 3 evaluates SmartBerry-DA under limited target annotation; and Experiment 4 assesses lightweight deployment and robustness. Unless otherwise stated, the evaluation protocol and metrics follow Sections 3.5 and 3.6.

4.1. Experiment 1: Lightweight Model Benchmarking

Experiment 1 compares the four candidate architectures independently on the PlantVillage and SmartBerry multiclass datasets. The objective is to establish their predictive performance and computational requirements before cross-dataset evaluation. Model performance is assessed using accuracy, balanced accuracy, macro F1 and ROC-AUC, together with parameter count, model size and inference latency.

Performance on PlantVillage was uniformly high, with macro F1-scores of 0.990 or above for all models with completed multi-seed evaluation (Table 5). ConvNeXt-Tiny achieved a macro F1-score of 0.997 ± 0.002, while MobileNetV3-Small achieved 0.995 ± 0.001 with substantially fewer parameters. The single completed MobileViTv2-0.5 run also achieved a macro F1-score of 0.998, but is not used for multi-seed comparison. The SmartBerry results provided greater separation between architectures (Table 6). ConvNeXt-Tiny achieved the highest macro F1 score (0.873 ± 0.014), followed by EfficientNet-B0 (0.840 ± 0.028), MobileViTv2-0.5 (0.834 ± 0.030) and MobileNetV3-Small (0.823 ± 0.037). ConvNeXt-Tiny therefore provided the strongest predictive performance, but required approximately 28 million parameters and 106 MB of storage. In contrast, MobileViTv2-0.5 required only 1.12 million parameters and 4.37 MB, while MobileNetV3-Small achieved the lowest measured GPU latency. These results show an accuracy-efficiency trade-off: the higher-capacity ConvNeXt-Tiny produced the strongest SmartBerry classification performance, whereas the lightweight architectures reduced model size and parameter count substantially. These results provide the basis for the subsequent cross-dataset evaluation.

4.2. Experiment 2: Cross-Dataset Generalisation

Experiment 2 evaluates how well models trained on controlled PlantVillage images generalise to SmartBerry field images. The analysis uses the harmonised binary classes *healthy leaves* and *leaf dark spot*, with four training-testing conditions: PlantVillage-to-PlantVillage (PV→PV), PlantVillage-to-SmartBerry (PV→SB), SmartBerry-to-SmartBerry (SB→SB), and pooled PlantVillage + SmartBerry training followed by SmartBerry testing (Pooled→SB). The same held-out SmartBerry test set is used for PV→SB, SB→SB and Pooled→SB, and each model-condition combination is evaluated across five random seeds. Macro F1 is the primary metric, supported by balanced accuracy, recall, specificity, ROC-AUC and ECE.

4.2.1. Cross-Dataset Performance

All architectures achieved near-ceiling performance under controlled PV→PV evaluation, with macro F1 ranging from 0.992 ± 0.016 to 1.000 ± 0.000 (Table 7). Performance decreased substantially under direct PV→SB transfer, demonstrating a clear controlled-to-field generalisation gap. ConvNeXt-Tiny retained the strongest direct-transfer performance (0.824 ± 0.058) and the smallest domain gap (0.176). MobileViTv2-0.5 achieved 0.662 ± 0.061, MobileNetV3-Small 0.550 ± 0.123, and EfficientNet-B0 0.418 ± 0.036. Architecture differences under direct transfer were significant ($\chi^2(3) = 12.84$, $p = 0.005$, Kendall's $W = 0.856$).

Access to SmartBerry training data substantially recovered field performance. MobileViTv2-0.5 achieved the highest SB→SB macro F1 (0.994 ± 0.008), followed closely by ConvNeXt-Tiny (0.991 ± 0.014). Pooled training also improved all architectures relative to direct transfer, although

Table 6
SmartBerry multiclass performance and computational characteristics. Values are mean ± standard deviation across five random seeds.

Model	Accuracy	Balanced accuracy	Macro F1	ROC-AUC	Parameters	Size	GPU latency
ConvNeXt-Tiny	0.920 ± 0.011	0.881 ± 0.013	0.873 ± 0.014	0.994 ± 0.002	27.83 M	106.20 MB	6.03 ± 0.08 ms
EfficientNet-B0	0.899 ± 0.022	0.848 ± 0.029	0.840 ± 0.028	0.989 ± 0.003	4.02 M	15.59 MB	8.83 ± 0.10 ms
MobileViTv2-0.5	0.891 ± 0.013	0.844 ± 0.026	0.834 ± 0.030	0.985 ± 0.009	1.12 M	4.37 MB	9.28 ± 0.07 ms
MobileNetV3-Small	0.871 ± 0.038	0.842 ± 0.039	0.823 ± 0.037	0.979 ± 0.009	1.53 M	5.94 MB	5.78 ± 0.12 ms

Table 7
Macro F1 across the four cross-dataset evaluation conditions. Domain gap is PV→PV minus PV→SB, while target and pooled gains are measured relative to PV→SB. Values are mean ± standard deviation across five seeds.

Model	PV→PV	PV→SB	SB→SB	Pooled→SB	Domain gap	Target gain	Pooled gain
ConvNeXt-Tiny	1.000 ± 0.000	0.824 ± 0.058	0.991 ± 0.014	0.986 ± 0.032	0.176	0.167	0.162
MobileViTv2-0.5	0.998 ± 0.002	0.662 ± 0.061	0.994 ± 0.008	0.948 ± 0.016	0.336	0.332	0.286
EfficientNet-B0	0.999 ± 0.002	0.418 ± 0.036	0.905 ± 0.066	0.960 ± 0.037	0.582	0.487	0.542
MobileNetV3-Small	0.992 ± 0.016	0.550 ± 0.123	0.878 ± 0.053	0.921 ± 0.029	0.441	0.328	0.371

Table 8
Class-specific performance following direct PlantVillage-to-SmartBerry transfer. Values are mean ± standard deviation across five seeds.

Model	Balanced accuracy	Dark-spot recall	Healthy-leaf specificity
ConvNeXt-Tiny	0.840 ± 0.045	0.839 ± 0.141	0.840 ± 0.158
MobileViTv2-0.5	0.654 ± 0.050	0.957 ± 0.016	0.352 ± 0.104
MobileNetV3-Small	0.571 ± 0.087	0.918 ± 0.067	0.224 ± 0.166
EfficientNet-B0	0.508 ± 0.018	1.000 ± 0.000	0.016 ± 0.036

Table 9
ECE on SmartBerry before and after temperature scaling using the PlantVillage validation set. Values are mean ± standard deviation across five seeds.

Model	Uncalibrated ECE	Source-calibrated ECE	Change
ConvNeXt-Tiny	0.129 ± 0.027	0.160 ± 0.056	+0.031
MobileViTv2-0.5	0.120 ± 0.029	0.240 ± 0.026	+0.120
MobileNetV3-Small	0.231 ± 0.068	0.298 ± 0.070	+0.067
EfficientNet-B0	0.210 ± 0.047	0.326 ± 0.007	+0.116

it was not consistently best to SmartBerry-only training. In particular, MobileViTv2-0.5 decreased from 0.994 ± 0.008 under SB→SB to 0.948 ± 0.016 under pooled training.

4.2.2. *Class-Specific and Calibration Behaviour*

Direct transfer produced strongly asymmetric class behaviour for several lightweight models. As shown in Table 8 and Fig. 5, EfficientNet-B0 achieved perfect dark-spot recall (1.000 ± 0.000) but only 0.016 ± 0.036 healthy-leaf specificity, indicating substantial overprediction of the dark-spot class. MobileViTv2-0.5 and MobileNetV3-Small showed the same tendency to a lesser extent, whereas ConvNeXt-Tiny maintained the most balanced recall-specificity profile. The corresponding seed-level confusion matrices for ConvNeXt-Tiny are shown in Fig. 3, illustrating the variability under direct controlled-to-field transfer.

Confidence estimates were also affected by domain shift. Temperature scaling fitted using only the PlantVillage validation set increased SmartBerry ECE for all four architectures (Table 9), indicating that calibration learned under controlled conditions did not transfer reliably to the field domain (Guo et al., 2017). The five-seed paired comparisons showed consistent practical improvements from direct transfer to both SmartBerry-only and pooled training, although the exact two-sided Wilcoxon tests returned $p = 0.0625$, reflecting the limited resolution of the test at $n = 5$. Image-level McNemar tests provided complementary evidence, with SmartBerry-only and pooled training outperforming direct transfer in 19/20 and 18/20 model-seed comparisons, respectively. The results from Experiment 2 show that strong controlled-dataset performance does not guarantee reliable field generalisation. Access to local SmartBerry data substantially reduces this gap, while the class-specific and calibration results show that domain shift affects both prediction balance and confidence. These findings motivate the label-efficient adaptation evaluated in Experiment 3.

4.3. Experiment 3: Label-Efficient Domain Adaptation

Experiment 3 evaluates how much labelled SmartBerry data is required to adapt a PlantVillage-trained model to field conditions. The harmonised binary task from Experiment 2 is retained, comprising *healthy leaf* and *leaf dark-spot* symptoms. MobileViTv2-0.5 is used for the primary analysis because it provided the strongest lightweight target-domain performance in the preceding experiment. The number of labelled SmartBerry images is restricted to $k \in \{5, 10, 20, 50\}$ per class. Each run therefore uses $2k$ labelled target images, while the remaining SmartBerry training images are treated as unlabelled target data. This provides 292, 282, 262 and 202 unlabelled images for $k = 5, 10, 20$ and 50, respectively. All available labelled target images within each budget are used for training; no target-domain validation subset is reserved. The same held-out SmartBerry test set of 76 images is used throughout and is excluded from model selection, threshold tuning and calibration.

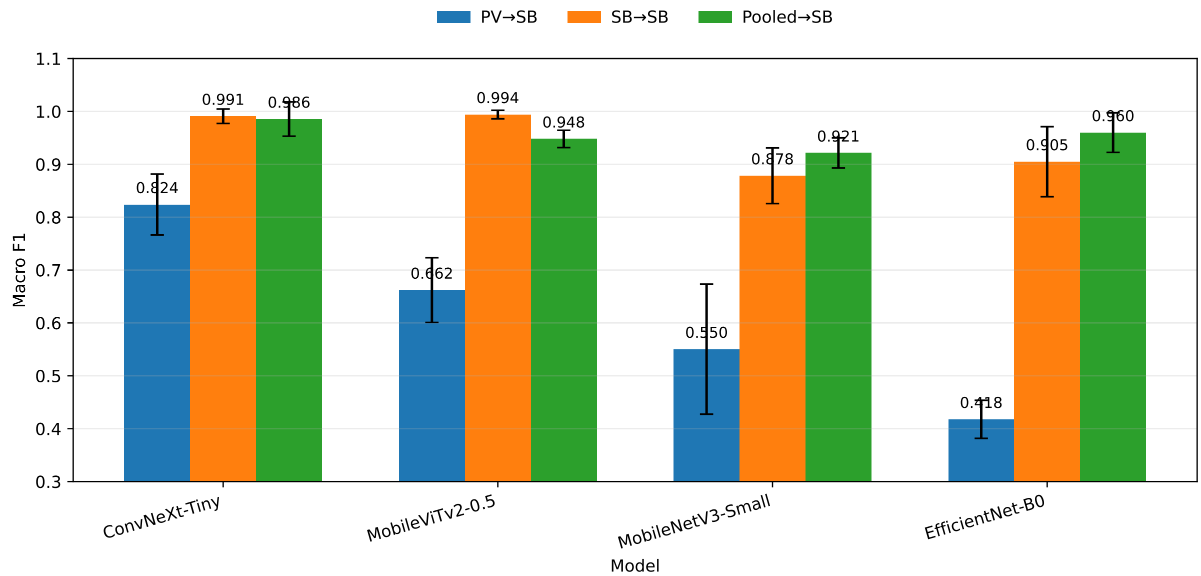


Figure 4: Macro F1 on the held-out SmartBerry test set under direct controlled-to-field transfer, SmartBerry-only training and pooled training.

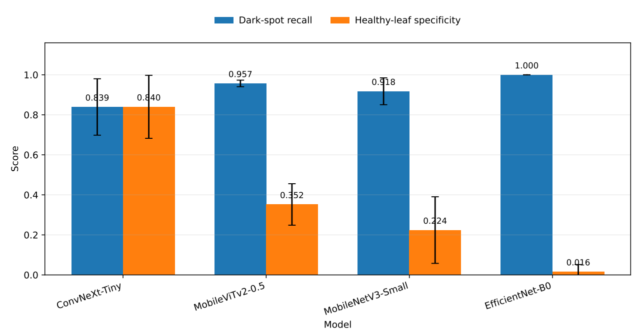


Figure 5: Dark-spot recall and healthy-leaf specificity following direct PlantVillage-to-SmartBerry transfer.

Three adaptation strategies are compared at each annotation budget: *target fine-tuning*, where the PlantVillage-trained model is fine-tuned using only the labelled SmartBerry subset; *labelled pooling*, where PlantVillage and labelled SmartBerry images are jointly used for supervised training; and *SmartBerry-DA*, which additionally exploits the remaining unlabelled SmartBerry images. The direct-transfer and full-target target-only results from Experiment 2 are retained as reference points.

SmartBerry-DA follows the teacher-student formulation described in Section 3.2. The EMA teacher uses a decay of 0.999, while class-aware confidence filtering is applied to pseudo-labels, with the global confidence threshold scheduled from 0.95 to 0.80 during training. The adaptation losses are introduced through a ten-epoch sigmoid ramp-up. All trainable methods use AdamW with a learning rate of 10^{-4}, weight decay of 10^{-4}, batch size 16, label smoothing of 0.05, cosine learning-rate decay and gradient clipping at 1.0. Models are trained for 25 epochs. The primary MobileViTv2-0.5 analysis uses five random seeds $\{11, 22, 33, 44, 55\}$, with the random state reset for each independent model-method-budget run. Macro F1-score is the primary outcome, with balanced accuracy, ECE and AURC used as complementary measures.

4.3.1. Label-Efficiency Results

Table 10 summarises the primary MobileViTv2-0.5 results. Direct PlantVillage-to-SmartBerry transfer achieved a macro F1-score of 0.662 ± 0.061. With only five labelled SmartBerry images per class, SmartBerry-DA increased macro F1 to 0.940 ± 0.015, compared with 0.795 ± 0.038 for target fine-tuning and 0.915 ± 0.017 for labelled pooling. At $k = 10$, SmartBerry-DA achieved 0.988 ± 0.012, compared with the full-target target-only reference of 0.994 ± 0.008. Performance increased to 0.991 ± 0.013 at $k = 20$. At $k = 50$, differences between the methods narrowed, with labelled pooling, SmartBerry-DA and target fine-tuning achieving 1.000 ± 0.000, 0.997 ± 0.007 and 0.994 ± 0.008, respectively.

The overall method effect was significant at $k = 5$ ($\chi^2_F = 8.40$, $p = 0.015$, $W = 0.840$), $k = 10$ ($\chi^2_F = 8.40$, $p = 0.015$, $W = 0.840$), and $k = 20$ ($\chi^2_F = 8.32$, $p = 0.016$, $W = 0.832$), but not at $k = 50$ ($\chi^2_F = 3.71$, $p = 0.156$, $W = 0.371$). SmartBerry-DA exceeded target fine-tuning in all five seeds at $k = 5$, 10 and 20. However, the corresponding exact two-sided Wilcoxon tests returned $p = 0.0625$, reflecting the limited resolution of the test with five paired runs. The performance trend across annotation budgets is shown in Fig. 6.

The results show that the benefit of SmartBerry-DA is greatest under limited target supervision. Performance approaches the full-target reference with 10-20 labelled images

Table 10
Label-efficiency results using MobileViTv2-0.5. Values are mean ± standard deviation across five seeds. The best result within each restricted-label budget is shown in bold.

Method	$k = 5$	$k = 10$	$k = 20$	$k = 50$
Target fine-tuning	0.795 ± 0.038	0.857 ± 0.044	0.908 ± 0.030	0.994 ± 0.008
Labelled pooling	0.915 ± 0.017	0.973 ± 0.016	0.979 ± 0.020	$\mathbf{1.000 \pm 0.000}$
SmartBerry-DA	$\mathbf{0.940 \pm 0.015}$	$\mathbf{0.988 \pm 0.012}$	$\mathbf{0.991 \pm 0.013}$	0.997 ± 0.007

Table 11
SmartBerry-DA uncertainty and pseudo-label diagnostics. Values are mean ± standard deviation across five seeds. Lower ECE and AURC are better.

k	ECE	AURC	Acceptance rate	Accepted-label accuracy
5	0.0879 ± 0.0160	0.00378 ± 0.00257	0.841 ± 0.026	0.981 ± 0.012
10	0.0957 ± 0.0098	0.00058 ± 0.00085	0.870 ± 0.030	0.994 ± 0.004
20	0.0827 ± 0.0041	0.00032 ± 0.00047	0.894 ± 0.022	0.995 ± 0.003
50	0.0663 ± 0.0039	0.00014 ± 0.00032	0.907 ± 0.032	0.992 ± 0.009

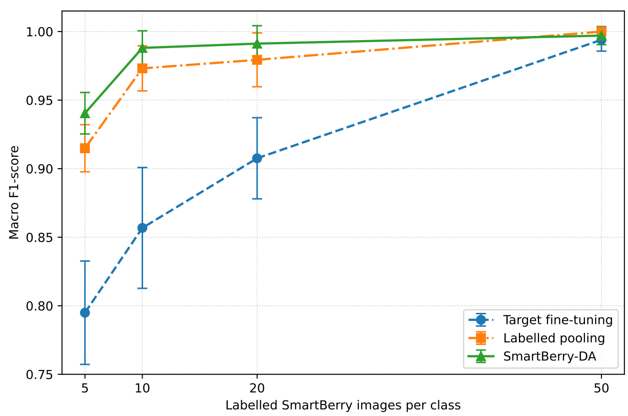


Figure 6: Label-efficient adaptation performance of MobileViTv2-0.5 on the held-out SmartBerry test set.

per class, while differences between adaptation strategies diminish as the annotation budget increases.

4.3.2. Uncertainty and Pseudo-Label Behaviour

Table 11 summarises the uncertainty and pseudo-label diagnostics for SmartBerry-DA. The pseudo-label acceptance rate increased from 0.841 ± 0.026 at $k = 5$ to 0.907 ± 0.032 at $k = 50$, while accepted-label accuracy remained high, ranging from 0.981 to 0.995. AURC decreased from 0.00378 at $k = 5$ to 0.00014 at $k = 50$, indicating improved uncertainty-based ranking as the amount of labelled target data increased. These results indicate that confidence filtering retained a substantial proportion of the unlabelled target images while maintaining high pseudo-label accuracy.

4.3.3. Cross-Backbone Sensitivity

A supplementary analysis examined whether the observed adaptation behaviour extended to the other model architectures. ConvNeXt-Tiny, EfficientNet-B0 and MobileNetV3-Small were evaluated at $k \in \{5, 20, 50\}$. Four complete seeds, $\{22, 33, 44, 55\}$, were available and the results are therefore interpreted descriptively. As shown in Table 12, SmartBerry-DA performed very well with ConvNeXt-Tiny, achieving 0.982 ± 0.022 at $k = 5$ and 1.000 ± 0.000 at both $k = 20$ and $k = 50$. In contrast, EfficientNet-B0 and MobileNetV3-Small showed weaker and more variable SmartBerry-DA performance, with supervised pooling or fine-tuning generally producing higher macro F1-scores.

The cross-backbone results indicate that SmartBerry-DA is not architecture-invariant. The strongest behaviour is observed with MobileViTv2-0.5 and ConvNeXt-Tiny, while its effectiveness is less consistent with EfficientNet-B0 and MobileNetV3-Small.

4.3.4. Component Ablation

Component ablation was performed using MobileViTv2-0.5 at $k = 20$ over five runs. The complete SmartBerry-DA configuration was compared with variants excluding CORAL, pseudo-label cross-entropy, teacher–student consistency or uncertainty filtering. Table 13 shows that removing uncertainty filtering produced the largest deterioration, reducing macro F1 from 0.994 ± 0.008 to 0.946 ± 0.047 and increasing AURC from 0.00032 to 0.00916. Removing pseudo-label cross-entropy or consistency produced smaller reductions in macro F1, while removing CORAL resulted in the smallest mean change.

The overall ablation effect was significant ($\chi^2_F = 9.76$, $p = 0.0447$, $W = 0.488$). The complete method exceeded the variant without uncertainty filtering in all five runs, although the exact paired Wilcoxon comparison returned $p = 0.0625$. Under the evaluated $k = 20$ setting, uncertainty filtering therefore had the largest measurable influence on SmartBerry-DA performance, while the remaining components provided smaller complementary contributions.

Results from Experiment 3 show that the principal benefit of SmartBerry-DA occurs when labelled target data are scarce. With only 10 labelled SmartBerry images per class, mean macro F1 reached 0.988 ± 0.012, close to the full-target target-only reference of 0.994 ± 0.008. The diminishing differences at larger annotation budgets indicate that the value of the approach lies primarily in label-efficient adaptation rather than improving performance when extensive target supervision is already available.

Table 12
Supplementary cross-backbone macro F1-score for the three additional architectures evaluated alongside the primary MobileViTv2-0.5 analysis. Values are mean ± standard deviation across four complete seeds (22, 33, 44 and 55).

Backbone	Method	$k = 5$	$k = 20$	$k = 50$
ConvNeXt-Tiny	Target fine-tuning	0.947 ± 0.027	0.996 ± 0.008	0.996 ± 0.008
	Labelled pooling	0.969 ± 0.034	1.000 ± 0.000	1.000 ± 0.000
	SmartBerry-DA	$\mathbf{0.982 \pm 0.022}$	$\mathbf{1.000 \pm 0.000}$	$\mathbf{1.000 \pm 0.000}$
EfficientNet-B0	Target fine-tuning	0.625 ± 0.099	0.766 ± 0.110	0.915 ± 0.020
	Labelled pooling	$\mathbf{0.872 \pm 0.052}$	$\mathbf{0.929 \pm 0.025}$	$\mathbf{0.971 \pm 0.012}$
	SmartBerry-DA	0.656 ± 0.135	0.671 ± 0.249	0.782 ± 0.193
MobileNetV3-Small	Target fine-tuning	0.671 ± 0.142	$\mathbf{0.893 \pm 0.058}$	$\mathbf{0.925 \pm 0.034}$
	Labelled pooling	$\mathbf{0.748 \pm 0.051}$	0.854 ± 0.055	0.911 ± 0.053
	SmartBerry-DA	0.414 ± 0.020	0.569 ± 0.134	0.543 ± 0.212

Table 13
SmartBerry-DA component ablation using MobileViTv2-0.5 and 20 labelled SmartBerry images per class. Values are mean ± standard deviation across five runs. ΔF1 is relative to the complete method.

Configuration	Macro F1	ΔF1	ECE	AURC
Complete SmartBerry-DA	$\mathbf{0.994 \pm 0.008}$	—	0.0802 ± 0.0019	0.00032 ± 0.00044
Without CORAL	0.991 ± 0.013	−0.0029	0.0811 ± 0.0063	0.00062 ± 0.00120
Without pseudo-label CE	0.988 ± 0.012	−0.0060	0.1074 ± 0.0098	0.00039 ± 0.00040
Without consistency	0.988 ± 0.012	−0.0059	$\mathbf{0.0696 \pm 0.0046}$	0.00050 ± 0.00075
Without uncertainty filtering	0.946 ± 0.047	−0.0479	0.0836 ± 0.0101	0.00916 ± 0.01114

4.4. Experiment 4: Lightweight Deployment and Robustness

Experiment 4 evaluates the deployment characteristics of the selected MobileViTv2-0.5 SmartBerry-DA model. The model trained with 20 labelled SmartBerry images per class was selected, and the seed-11 checkpoint was used as a pre-specified deployment instance to avoid post-hoc selection based on test performance. The checkpoint was exported from PyTorch to Open Neural Network Exchange (ONNX) in FP32, FP16 and static Quantize-Dequantize (QDQ) INT8 formats. INT8 post-training quantisation used 128 SmartBerry training images for calibration; no test images were used during model conversion. All exports were evaluated using ONNX Runtime 1.28.0 with the CPU execution provider in an x86-64 reference environment. Batch-one latency was measured over 200 repetitions following 20 warm-up runs. Predictive performance was evaluated on the held-out SmartBerry test set of 76 images at 224×224 resolution. All inference was performed locally without an active network connection.

4.4.1. *Deployment Performance*

Table 14 summarises the deployment results. FP16 reduced model size from 4.37 MB to 2.30 MB, a 47.2% reduction, while preserving all FP32 predictions and a macro F1-score of 1.000. Median latency was effectively unchanged at 13.03 ms compared with 13.14 ms for FP32. Static INT8 quantisation reduced model size further to 1.57 MB but substantially degraded predictive performance. Macro F1 decreased to 0.590, prediction agreement with FP32 fell to 59.2%, and median latency increased to 30.65 ms. FP16 therefore provided the most favourable size-performance trade-off among the evaluated representations.

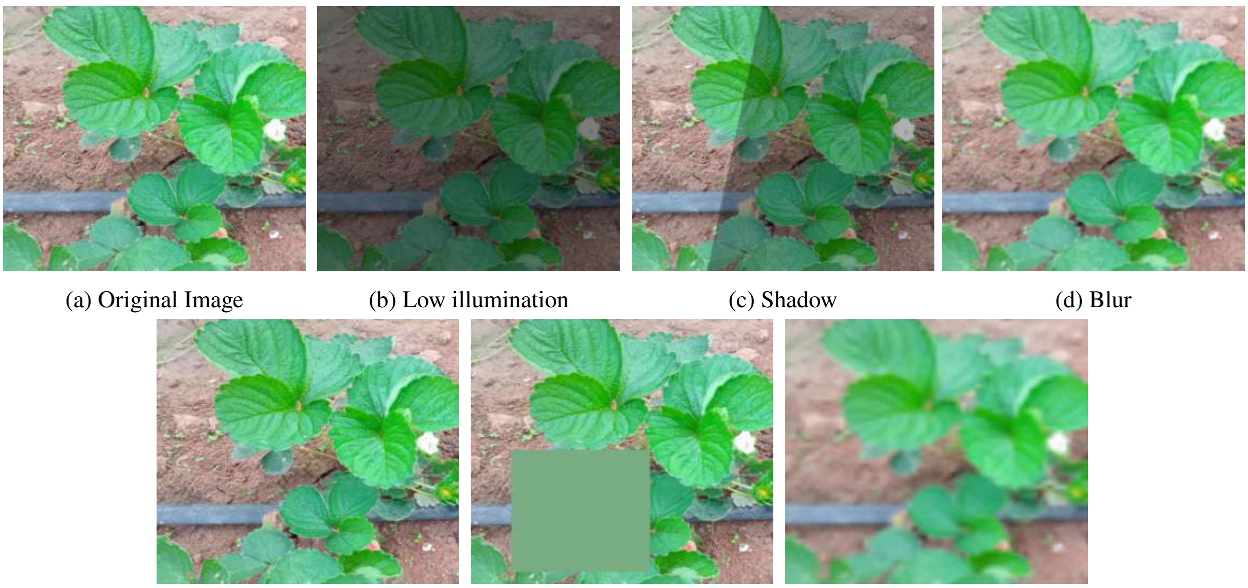
The reported latency values represent a reference portability benchmark rather than final smartphone or embedded-device performance. Measurements should therefore not be interpreted as direct estimates of latency on Android devices or Raspberry Pi hardware.

4.4.2. *Robustness under Field-Image Degradation*

Robustness was assessed by applying six deterministic degradations to every image in the held-out SmartBerry test set. These comprised low illumination (brightness factor 0.50), artificial shadow (opacity 0.55), Gaussian blur (radius 2.0), JPEG compression (quality 25), partial occlusion covering 20% of the image area, and downsampling to 64×64 pixels before resizing to the model input resolution. Representative examples of these transformations are shown in Fig. 7. A fixed transformation seed of 2026 was used, and no model retraining or parameter adjustment was performed. FP32 and FP16 produced identical macro F1-scores under all evaluated conditions. Both retained a macro F1-score of 1.000 under low illumination, shadow, blur and JPEG compression. Partial occlusion reduced macro F1 to 0.970, while reduced resolution produced the largest reduction, to 0.926. These results show that conversion to FP16 preserved the predictive behaviour of the FP32 model under the evaluated degradations. INT8 was substantially less stable. The macro F1-score decreased from 0.590 on the original images to 0.311 under low illumination and 0.460 under reduced resolution. The increase to 0.658 under partial occlusion should not be interpreted as improved robustness, since the clean INT8 model was already substantially degraded. The deployment evaluation identifies FP16 as the most suitable of the three evaluated representations, reducing the model storage by approximately half without altering the classification performance or robustness relative to FP32. However,

Table 14
Reference CPU deployment performance of the selected MobileViTv2-0.5 SmartBerry-DA checkpoint. Latency was measured at batch size one over 200 repetitions following 20 warm-up runs.

Precision	Model size (MB)	Median latency (ms)	95th percentile (ms)	FP32 agreement	Macro F1
FP32	4.37	13.14	13.37	100.0%	1.000
FP16	2.30	13.03	13.36	100.0%	1.000
INT8	1.57	30.65	31.43	59.2%	0.590



(a) Original Image (b) Low illumination (c) Shadow (d) Blur
(e) JPEG compression (f) Partial occlusion (g) Reduced resolution

Figure 7: Representative SmartBerry image under the field-image transformations used in the robustness evaluation: original image, low illumination, artificial shadow, Gaussian blur, JPEG compression, partial occlusion and reduced resolution. The same pre-specified transformations were independently applied to all images in the held-out test set.

Table 15
Macro F1-score under simulated field-image degradation.

Condition	FP32	FP16	INT8
Original images	1.000	1.000	0.590
Low illumination	1.000	1.000	0.311
Shadow	1.000	1.000	0.487
Blur	1.000	1.000	0.562
JPEG compression	1.000	1.000	0.562
Partial occlusion	0.970	0.970	0.658
Reduced resolution	0.926	0.926	0.460

robustness results should be interpreted as controlled stress tests at fixed degradation levels rather than as a complete representation of the variability encountered during routine field deployment.

4.5. Discussion

The findings are discussed in relation to controlled-to-field generalisation, the value of limited local supervision, the contribution and scope of SmartBerry-DA, and lightweight deployment. Particular attention is given to the conditions under which the proposed approach is effective and to the limitations affecting broader generalisation.

4.5.1. Controlled-to-Field Generalisation and the Value of Local Data

The experiments show a clear gap between controlled-dataset performance and field generalisation. All architectures achieved near-ceiling performance on PlantVillage, yet direct transfer to SmartBerry reduced macro F1 to 0.824 ± 0.058 for ConvNeXt-Tiny, 0.662 ± 0.061 for MobileViTv2-0.5, 0.550 ± 0.123 for MobileNetV3-Small, and 0.418 ± 0.036 for EfficientNet-B0. This supports previous concerns that performance obtained on controlled plant-disease datasets may not reflect behaviour under natural acquisition conditions (Mohanty et al., 2016; Hughes et al., 2015). The class-specific results further showed that domain shift affected prediction balance, not only aggregate accuracy. Several lightweight models retained high dark-spot recall while exhibiting poor healthy-leaf specificity, indicating systematic overprediction of the diseased class. Source-domain temperature scaling also increased SmartBerry ECE across all architectures, showing that confidence calibration learned under controlled conditions did not transfer reliably to the field domain (Guo et al., 2017). Access to local SmartBerry data substantially reduced this gap. In particular, SmartBerry-DA achieved a macro F1-score of 0.940 ± 0.015 with five labelled images per class and 0.988 ± 0.012 with ten, compared with the full-target reference of 0.994±0.008. The main benefit of local adaptation therefore occurred under limited annotation, while differences between methods reduced as more labelled target data became available.

4.5.2. Contribution and Scope of SmartBerry-DA

SmartBerry-DA contributes a label-efficient adaptation strategy rather than a new backbone architecture. It combines limited labelled target supervision with unlabelled SmartBerry images through EMA teacher-student learning, confidence-filtered pseudo-labels, consistency regularisation, and CORAL-based feature alignment. The ablation study identified uncertainty filtering as the most influential component under the evaluated $k = 20$ setting. Removing it reduced macro F1 from 0.994 ± 0.008 to 0.946 ± 0.047 and substantially increased AURC. Removing pseudo-label cross-entropy, consistency or CORAL produced smaller changes. The results therefore support selective use of reliable pseudo-labels as an important element of adaptation under limited target supervision. SmartBerry-DA was not uniformly effective across architectures. Strong results were obtained with MobileViTv2-0.5 and ConvNeXt-Tiny, whereas EfficientNet-B0 and MobileNetV3-Small showed weaker and more variable adaptation behaviour. The method should therefore be interpreted as backbone-sensitive rather than architecture-invariant.

4.5.3. Lightweight Deployment under Field-Oriented Conditions

The deployment evaluation showed that FP16 provided the most favourable representation among those tested. For the selected MobileViTv2-0.5 SmartBerry-DA checkpoint, FP16 reduced model size from 4.37 MB to 2.30 MB while preserving all FP32 predictions and a macro F1-score of 1.000. Reference CPU latency was also effectively unchanged, at 13.03 ms for FP16 compared with 13.14 ms for FP32. Static INT8 quantisation reduced storage further to 1.57 MB but substantially degraded both accuracy and prediction agreement and increased latency in the evaluated environment. More compression therefore did not translate into a better deployment configuration. FP16 also reproduced FP32 performance under all simulated image degradations. Performance remained unchanged under the evaluated low-illumination, shadow, blur and JPEG conditions, while partial occlusion and reduced resolution decreased macro F1 to 0.970 and 0.926, respectively. These results support FP16 as the preferred representation among those evaluated, while also identifying resolution loss and occlusion as relevant failure conditions. The measurements demonstrate offline inference and model portability, but they were obtained on a hosted x86-64 CPU rather than final farm hardware. Consequently, they should not be interpreted as direct evidence of smartphone or Raspberry Pi latency. Nevertheless, the small model footprint and offline execution provide a practical basis for further evaluation on resource-limited devices and in settings where continuous connectivity cannot be assumed.

4.5.4. Limitations

Although this paper reported several good results, some limitations remain. First, the cross-dataset analysis is restricted to the symptom-compatible healthy and dark-spot/scorch classes and should not be generalised to other strawberry diseases without further evaluation. Second, the held-out SmartBerry test set contained 76 images, so near-ceiling scores correspond to a small number of classification errors. Third, the SmartBerry-DA sensitivity analysis used four complete seeds for the additional backbones and showed clear architecture-dependent behaviour. Fourth, robustness was evaluated using fixed simulated degradations and a single pre-specified deployment checkpoint; these tests cannot reproduce the full range of variability encountered under routine field conditions. Furthermore, the deployment latency was measured on a reference x86-64 CPU rather than the intended end-user mobile or embedded hardware.

5. Conclusion

This paper investigated controlled-to-field strawberry disease classification, using PlantVillage as the controlled-source domain and original SmartBerry imagery collected under Nigerian farming conditions as the field target domain. Although the evaluated architectures achieved near-ceiling performance on PlantVillage, direct transfer to SmartBerry reduced macro F1-score to between 0.418 and 0.824, demonstrating that strong within-dataset performance does not necessarily translate to reliable field generalisation. Access to locally collected SmartBerry data substantially reduced this gap, confirming the importance of target-domain information when deploying plant disease classifiers under acquisition conditions that differ from those represented during source training.

To reduce the amount of local annotation required, SmartBerry-DA combined limited labelled target supervision with unlabelled SmartBerry images through CORAL, an EMA teacher, confidence-filtered pseudo-labels and consistency regularisation. Using MobileViTv2-0.5, SmartBerry-DA achieved macro F1-scores of 0.940±0.015, 0.988±0.012 and 0.991±0.013 with only 5, 10 and 20 labelled SmartBerry images per class, respectively, compared with the full-target reference of 0.994±0.008. The ablation analysis showed that uncertainty filtering had the largest measurable influence under the evaluated setting. However, the cross-backbone analysis also showed that the effectiveness of SmartBerry-DA is architecture-dependent, with stronger behaviour observed for MobileViTv2-0.5 and ConvNeXt-Tiny than for EfficientNet-B0 and MobileNetV3-Small. The deployment evaluation further showed that 16-bit FP16 representation reduced the selected MobileViTv2-0.5 model from 4.37 MB to 2.30 MB while preserving the predictions and predictive performance of the 32-bit floating-point (FP32) model under the evaluated conditions. In contrast, static INT8 post-training quantisation reduced storage further but substantially degraded predictive performance. These results identify FP16 as the most favourable representation among those evaluated for subsequent lightweight offline deployment.

The findings demonstrate the importance of field-derived data, limited local supervision and uncertainty-aware adaptation when transferring plant-disease recognition models from controlled datasets to real agricultural environments. Within the SmartBerry setting, the results provide a technically evaluated pathway towards label-efficient and lightweight strawberry disease recognition under Nigerian field conditions. Future work should extend the evaluation across additional farms, seasons, cultivars, acquisition devices, and symptom categories, and benchmark the selected model directly on target mobile and embedded hardware. Further work should also investigate the integration of SmartBerry environmental sensing data with image-based recognition to support multimodal crop-health assessment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or specify any relevant commercial relationship with the SmartBerry partners.

Ethics statement

NA

Data availability

The PlantVillage dataset is publicly available from its cited source. The SmartBerry dataset is publicly available through Zenodo, with a Kaggle mirror provided for convenient access.SmartBerry dataset: Zenodo: https://zenodo.org/records/22056372; Kaggle: https://www.kaggle.com/datasets/isiborihianle/smartberry-strawberry-field-image-dataset.

Funding

This research was supported by the SmartBerry Project (Project Number: 10071867) and the Department of Computer Science, Nottingham Trent University.

Acknowledgements

The authors thank the Ashley Strawberry Farm staff who supported image collection, annotation and field evaluation.

Bibliography

Abbas, I., Liu, J., Amin, M., Tariq, A., Tunio, M.H., 2021. Strawberry fungal leaf scorch disease identification in real-time strawberry field using deep learning architectures. Plants 10, 2643.

Ayim, C., Kassahun, A., Addison, C., Tekinerdogan, B., 2022. Adoption of ict innovations in the agriculture sector in africa: a review of the literature. Agriculture & Food Security 11, 22.

Barbedo, J.G.A., 2026. Transferability and robustness in proximal and uav crop imaging. Agronomy 16, 364.

Choruma, D.J., Dirwai, T.L., Mutenje, M.J., Mustafa, M., Chimonyo, V.G.P., Jacobs-Mata, I., Mabhaudhi, T., 2024. Digitalisation in agriculture: A scoping review of technologies in practice, challenges, and opportunities for smallholder farmers in sub-saharan africa. Journal of agriculture and food research 18, 101286.

Eweoya, I., Okuboyejo, S.R., Odetunmibi, O.A., Odusote, B.O., 2021. An empirical investigation of acceptance, adoption and the use of e-agriculture in nigeria. Heliyon 7.

Fanourakis, D., Makraki, T., Spyrou, G.P., Karavidas, I., Tsaniklidis, G., Ntatsi, G., 2025. Environmental drivers of fruit quality and shelf life in greenhouse vegetables: Species-specific insights. Agronomy 16, 48.

Geifman, Y., Uziel, G., El-Yaniv, R., 2018. Bias-reduced uncertainty estimation for deep neural classifiers. arXiv preprint arXiv:1805.08206 .

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks, in: International conference on machine learning, PMLR. pp. 1321–1330.

He, Y., Yuan, Z., Sun, W., Li, Y., Liu, Y., Ye, Y., Sun, L., 2026. Visionmor: Scaling vision transformer via patch-level mixture-of-recursions, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4699–4707.

Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., 2019. Searching for mobilenetv3, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314–1324.

Hu, X., Wang, R., Du, J., Hu, Y., Jiao, L., Xu, T., 2023. Class-attention-based lesion proposal convolutional neural network for strawberry diseases identification. Frontiers in Plant Science 14, 1091600.

Hughes, D., Salathé, M., et al., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 .

Ihianle, I.K., 2024. Web sourced dataset for plant disease detection. URL: https://doi.org/10.5281/zenodo.14051480, doi:10.5281/zenodo.14051480.

ITU, Food, of the United Nations, A.O., 2022. Status of digital agriculture in 47 sub-Saharan African countries. Food & Agriculture Organization of the United Nations.

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., Kalenichenko, D., 2018. Quantization and training of neural networks for efficient integer-arithmetic-only inference, in: 2018 IEEE/CVF conference on computer vision and pattern recognition, IEEE. pp. 2704–2713.

Kolapo, A., Didunyemi, A.J., 2024. Effects of exposure on adoption of agricultural smartphone apps among smallholder farmers in southwest, nigeria: implications on farm-level-efficiency. Agriculture & Food Security 13, 31.

Krishna, M.S., Machado, P., Otuka, R.I., Yahaya, S.W., Neves dos Santos, F., Ihianle, I.K., 2025. Plant leaf disease detection using deep learning: A multi-dataset approach. J 8, 4.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s, in: 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR), IEEE. pp. 11966–11976.

Mehta, S., Rastegari, M., 2022a. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer, in: International Conference on Learning Representations.

Mehta, S., Rastegari, M., 2022b. Separable self-attention for mobile vision transformers. arXiv preprint arXiv:2206.02680 .

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7, 215232.

Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M.E., Cortés-Ancos, E., 2020. A fast and accurate deep learning method for strawberry instance segmentation. Computers and Electronics in Agriculture 178, 105736.

Pomykala, J., De Lemos, F., Ihianle, I.K., Adama, D.A., Machado, P., 2022. Deep learning approach for classifying trusses and runners of strawberries, in: UK Workshop on Computational Intelligence, Springer. pp. 427–435.

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N., 2020. Plantdoc: A dataset for visual plant disease detection, in: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 249–253.

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C., Cubuk, E.D., Kurakin, A., Li, C.L., 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems 33, 596–608.

Sun, B., Saenko, K., 2016. Deep coral: Correlation alignment for deep domain adaptation, in: European conference on computer vision, Springer. pp. 443–450.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PmLR. pp. 6105–6114.

Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems 30.

Thapa, K., 2023. Characterization of physical and biochemical traits in wheat and corn plants using high throughput image analysis .

Too, E.C., Li, Y., Kwao, P., Njuki, S., Mosomi, M.E., Kibet, J., 2019. Deep pruned nets for efficient image-based plants disease classification. Journal of Intelligent & Fuzzy Systems 37, 4003–4019.

Wu, X., Fan, X., Luo, P., Choudhury, S.D., Tjahjadi, T., Hu, C., 2023. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. Plant Phenomics 5, 0038.