Extending HSV-Based Color Selection for Non-Invasive Plant Disease Detection Using UAV Visible-Light Imagery

Toni Kusnandar[1]*, Heru Nugroho[2], Erna Hikmawati[3], Aisah Nur Endah Sari[4], Fitri Rizqiati[5]
Informatics Engineering, STMIK Mardira Indonesia, Bandung, Indonesia[1, 4, 5]
Faculty of Applied Sciences, Telkom University, Bandung, Indonesia[2, 3]

***Abstract*—Per-pixel HSV color selection methods from prior UAV-based research do not generalize across imaging sessions. This study evaluates plot-level aggregated RGB and HSV statistics of vegetation pixels (separated using the Excess Green index with Otsu thresholding) with a Random Forest, on 732 UAV RGB plot images from a multi-year common bean trial in Colombia (2022–2024), classifying non-inoculated control plots versus Common Bacterial Blight-inoculated plots; labels represent the experimental treatment at plot level, not disease severity. Under plot-grouped cross-validation, HSV features achieved 66.9% accuracy versus 59.1% for RGB features (majority-class baseline: 59.0%; plot-level permutation test, $p < 0.001$), with Saturation and Value contributing most. However, under leave-one-session-out validation, accuracy fell to 52.2% (HSV) and 50.1% (RGB), and per-session normalization recovered at most 60.5% without a significant HSV–RGB difference. Within single sessions, spatially blocked validation reduced HSV accuracy from 70.7% to 63.3%, below the within-session majority baseline (65.5%). HSV features are therefore more informative than RGB features under pooled validation, but the results do not demonstrate disease-specific detection that transfers across sessions or field positions; explicit radiometric calibration and lesion-level labels are needed.**

***Keywords*—UAV; HSV color selection; excess green; random forest; radiometric calibration; common bean**

I. Introduction

Unmanned Aerial Vehicles (UAVs) have become the main data acquisition platform in precision agriculture for plant disease detection and monitoring, as they are able to efficiently provide high-resolution imagery on a large field scale [1]. A systematic review of the literature on UAVs for plant disease detection shows a significant surge in the number of publications in recent years, dominated by deep learning approaches [2] [3] [4], with multispectral and RGB sensors as the most commonly used sensors. Various reviews also confirm that artificial intelligence, especially deep learning, has revolutionized plant disease detection to a large extent [5], although its limitations including the need for large training data and computing resources are also widely discussed.

A literature search on the Scopus database with the query TITLE-ABS-KEY (("UAV" OR "unmanned aerial vehicle" OR "drone") AND ("plant disease" OR "crop disease") AND "detection"), limited to publication years 2019–2026 and searched on 13-August-2026, shows a sharp trend of publication growth in the last five to six years (Fig. 1): of the 101 relevant publications identified (2019-2026), 94 publications (93.1%) were published in the 2021-2026 range, with the number of publications increasing from an average of 5-7 publications per year in the 2019-2021 period to 27 publications per year in 2025-2026. This trend confirms that the topic of UAV-based plant disease detection is a very active and growing research area. Interestingly, of the 94 publications for the 2021-2026 period, almost half (40 of 94 publications) explicitly address deep learning approaches (CNN, transformer, YOLO), a sharp increase compared to the period before 2021 (2 of 7 publications). This shift emphasizes the relevance of the contribution of this research which actually takes the opposite direction back to classic machine learning approaches based on aggregated color features as a direct response to the generalization limitations found empirically in pixel segmentation approaches, without requiring the complexity and big data requirements that accompany deep learning.

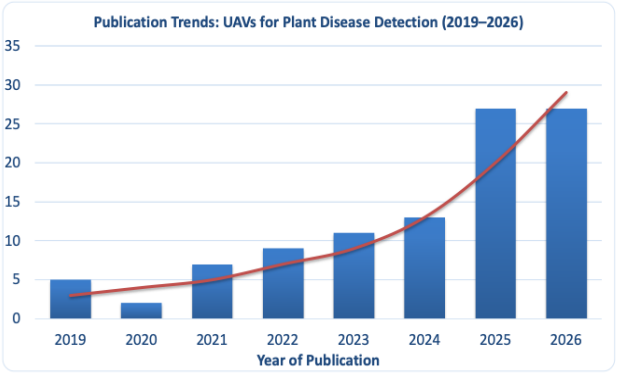

Fig. 1. Trend in the number of annual publications related to UAVs for plant disease detection (Scopus database, 2019-2026, n=101).

However, UAV image acquisition across shooting sessions (different dates, times, lighting conditions, and sensor calibrations) poses its own challenges: inter-session radiometric variations can result in significant differences in pixel color even if the object being photographed does not physically change. Pourazar et al. [6] explicitly shows that radiometric calibration is a necessity, not simply optional, for multispectral aerial image-based plant disease detection applications. Without

*Corresponding author.

This work is supported by the Directorate of Research, Technology, and Community Service, Ministry of Higher Education and Technology of the Republic of Indonesia (Grant numbers: 283/C3/DT.05.00/PL-BARU/2026, 1700/LL4/PG/2026 and 151/LPPM/A/STMIKMI/2026).

proper calibration, inter-session variations can obscure the actual signal to be detected (in this case, disease symptoms).

Kusnandar, in his research, proposed a color selection method in the HSV (Hue-Saturation-Value) color space based on a visible light camera on a UAV, which calculates the HSV reception range of a reference sample using a statistical approach of mean ± k·std to separate vegetation from the background on a per-pixel basis [7]. This method was then developed into a UAV visible-light camera-based vegetation index as a cheap alternative to multispectral NDVI [8], refined with white balance correction to maintain color consistency under lighting variations [9], and further refined through improving the stability of color values resulting from RGB to HSV conversion [10]. The basic premise of this method that the HSV color space has good discriminating power for vegetation has been quantitatively validated in a companion study comparing RGB and HSV features for plant disease classification using machine learning, finding a statistically significant superiority of HSV.

This research initially aimed to expand previous research methods directly: from vegetation-vs-background segmentation to healthy-vs-disease segmentation, using a dual-reference approach (two HSV reference samples, one for healthy tissue and one for diseased tissue) applied per-pixel. Various literature reviews confirm that UAVs are widely used for general disease monitoring [11] [12] [13], and pure RGB color-based approaches (without multispectral sensors) have been applied for example for Cercospora leaf spot detection using UAV RGB images [14]. However, testing on additional images from different acquisition sessions revealed that per-pixel dual-reference approaches using either one or many pooled reference patches do not generalize across images from different shooting sessions, as color variations due to inter-session calibration/exposure differences far exceed color variations due to disease status itself. These findings are consistent with the warnings of Pourazar et al. [6] regarding the importance of radiometric calibration in aerial imagery.

In response to these limitations, various classical machine learning-based plant disease detection studies (not deep learning) show that aggregated color features calculated as summary statistics (mean, std) over an area/plot, rather than raw pixel values are much more robust to variations in acquisition conditions [15] [16] [17]. A similar approach has been applied to machine learning-based UAV disease detection for bananas [18] and various reviews confirm that non-invasive (non-destructive) detection remains the main research goal in this area [19], with classical machine learning offering lightweight computing advantages over deep learning [20] [21].

Based on this description, this research revises the extension direction of the Kusnandar method: from dual-reference per-pixel segmentation to plot-level aggregated color feature-based classification, which maintains the core premise of the method proposed in previous research (the advantage of the HSV color space) but changes the mechanism to aggregated features intended to reduce sensitivity to radiometric variations between UAV sessions. Research questions answered: do plot-level aggregated HSV features remain superior to RGB on real field UAV imagery with a high variety of acquisition conditions, and how can the Kusnandar native vegetation segmentation method be integrated as a pre-processing stage that supports rather than replaces these feature-based approaches? This research contributes by: (1) empirically documenting the limited generalizability of per-pixel HSV segmentation approaches to multi-session UAVs, a practical finding rarely reported explicitly in the literature; (2) proposing and validating an alternative architecture that remains rooted in the HSV Kusnandar premise and is evaluated under plot-grouped, session-wise, and spatially blocked validation; and (3) evaluating this approach on 732 real UAV images (312 plots, seven sessions) from a multi-year common bean trial [22] [23], including an analysis of where its advantage over RGB features does and does not hold.

II. Related Work

This section summarizes research relevant to UAV-based plant disease detection, radiometric calibration challenges, and aggregated color-feature-based methods, grouped thematically.

A. Overview of UAVs for Plant Disease Detection

Kouadio et al. [1] conducted a systematic review of UAV applications for plant disease detection and monitoring, identifying a dominant trend toward multispectral sensors and deep learning. Neupane and Baysal-Gurel [11] and Abbas et al. [13] confirmed similar findings, with the latter highlighting UAVs as a pathway to smart agriculture. Chin et al. [12] specifically reviewed the use of drones in precision agriculture. These four reviews consistently noted that the majority of research leans toward expensive sensors (multispectral/hyperspectral) and deep learning models requiring large training datasets—a gap that motivates the classical machine learning approach based on color features used in this study.

B. Radiometric Calibration Challenges in UAV Imagery

Pourazar et al. [6] made a significant methodological contribution directly relevant to the empirical findings of this study: radiometric calibration of multispectral aerial imagery is not merely a best practice but a necessity for reliable plant disease detection, as variations in acquisition conditions (lighting, sun angle, sensor calibration) can produce significant reflectance differences without actual changes in plant conditions. While that study focused on multispectral sensors, the present research provides empirical evidence that similar challenges—potentially exacerbated by the absence of built-in reflectance calibration—also apply to visible-light RGB imagery; this contribution addresses that knowledge gap within the context of low-cost visible-light camera sensors.

C. RGB/HSV Color-Based Disease Detection Without Multispectral Imaging

Görlich et al. [14] demonstrated the detection of Cercospora leaf spot using standard UAV RGB imagery—without multi- or hyperspectral sensors—thereby showing the feasibility of a visible-light-only approach for specific disease detection. Zhang et al. [18] compared supervised and unsupervised methods for detecting Fusarium wilt in bananas using multispectral UAV imagery; while they still relied on multispectral sensors, their work provided an important methodological benchmark. Together, these studies established a precedent that color-based

disease detection is a viable direction; however, neither explicitly tested generalization across acquisition sessions—a gap addressed by the present study.

D. Classical Feature-Based Machine Learning for Disease Detection

Ganesh Babu and Chellaswamy [15] applied machine learning to detect various disease stages in pumpkin plants, while Ouhami et al. [16] provided a comprehensive review of combining computer vision, IoT, and data fusion for machine learning-based disease detection. Vasavi et al. [17] reviewed both classical and deep learning methods for detecting diseases based on visual symptoms, whereas Kohut et al. [20] and Ahmed & Yadav [21] demonstrated the application of classical machine learning models for plant disease detection. Collectively, these five studies indicate that feature-based approaches (as opposed to raw pixel-based or end-to-end deep learning methods) remain a relevant and practical choice—particularly for researchers with limited computational resources—aligning with the methodological direction proposed in this study.

E. Non-Invasive Detection and Baselines

Wang et al. [19] conducted an extensive review of non-destructive (non-invasive) technologies for plant disease detection, underscoring the relevance of the non-invasive objective central to this study and the baseline established by Kusnandar. In his dissertation, Kusnandar [7] proposed a statistical HSV color selection method—based on reference samples—to separate vegetation from the background in visible-light UAV imagery; this method was subsequently developed into a vegetation index [8], refined through white balance correction [9], and improved regarding color conversion stability [10]. These four studies serve as the primary methodological baseline that is expanded upon in the present research, incorporating significant revisions to the detection mechanism (utilizing aggregated features rather than per-pixel thresholds) based on empirical evidence regarding observed limitations in generalizability.

F. Alternative Sensors and Wider Application Context

A number of studies have compared multispectral and hyperspectral sensors for UAV-based plant disease detection, providing important context regarding the cost-accuracy trade-off underlying the choice of visible-light sensors in these studies. Kuswidiyanto et al. [24] reviewed airborne hyperspectral image-based deep learning, while Zhang et al. [25] and Bohnenkamp et al. [26] respectively detected yellow rust on wheat using high-resolution UAV hyperspectral imagery and a comparison of ground versus UAV canopy scales. Heim et al. [27] developed a specific spectral disease index for myrtle rust, Ualiyeva et al. [28] used VNIR hyperspectral signals for wheat disease classification, Wang et al. [29] integrated aerial and ground canopy reflectance spectroscopy for rice blast disease detection, Radócz et al. [30] used UAV multispectral vegetation index for hybrid corn disease identification, and Scutelnic et al. [31] developed a VIS-NIR multispectral camera with a thermal sensor for agricultural precision non-invasive analysis. All of these studies demonstrate high accuracy but consistently rely on sensors costing much more than the standard RGB visible-light cameras that are the focus of this research, underscoring the relevance of the proposed methodological contributions.

In line with the preference for low-cost sensors, several studies have explicitly explored low-cost cameras and edge computing devices for drone-based plant disease detection. Martínez-Heredia et al. [32] demonstrated the feasibility of detecting ocher spot on almonds using a very low-cost camera mounted on a drone, Hemalatha and Sangeetha [33] identified the pest using low-cost image acquisition, Joice et al. [34] reviewed Raspberry Pi applications for precision agriculture, and Annadata et al. [35] optimized TinyML models on edge devices for precision agriculture drones. These studies strengthen the argument that low-cost and computationally lightweight solutions including the aggregated feature approach proposed by this research have significant practical value for wide adoption, especially in developing countries.

The wider context of application is also addressed by recent systematic reviews: Ibrahim et al. [36] integrated AI and IoT for plant disease detection and treatment, El Sakka et al. [37] reviewed CNN applications on multimodal data of smart agriculture, Yilmaz et al. [38] reviewed computer vision for plant disease detection, Shawon et al. [39] reviewed federated learning approaches in precision agriculture, Ahmed et al. [40] reviewed the integration of UAVs and deep learning with real field challenges (using the example of cassava plants), Ma et al. [41] systematically reviewed AI-based pest and disease detection, Nyakuri et al. [42] reviewed IoT-based deep learning algorithms for pest and disease detection, and Manoj et al. [43] applied AI-based drone technology for early disease detection in large agricultural areas. Wiesner-Hanks et al. [44] demonstrated millimeter-level disease detection from aerial photos using deep learning and crowdsourced data, while Zhao et al. [45] evaluated the influence of spatial resolution on the accuracy of plant disease detection a methodological consideration that is relevant to the decision of this study to work at the plot-level instead of the pixel-level.

As shown in Table I, the common thread across the thematic groups is that the limitations of radiometric generalization that have been explicitly warned about for multispectral sensors by Pourazar et al. There has not been much empirical documentation for low-cost visible-light camera-based HSV approaches on UAVs. This research fills this gap with direct empirical evidence and proposes a revised architecture (aggregated features) that improves over RGB features under pooled validation but does not yet transfer across sessions.

TABLE I. COMPARISON OF RELATED RESEARCH

No	Research	Method	Sensor/ Platform	Advantages	Limitations
1	Kouadio et al. (2023)	Systematic review of UAV	Multispectral/RGB	Comprehensive research roadmap	Does not propose methods; ref [1]
2	Pourazar et al. (2019)	Radiometric calibration	Multispectral	Evidence of the importance of calibration	Multispectral focus, not RGB; ref [6]

No	Research	Method	Sensor/ Platform	Advantages	Limitations
3	Görlich et al. (2021)	Pure RGB Cercospora detection	RGB UAV	Without multispectral sensor	No cross-session testing; ref [14]
4	Ganesh Babu & Chellaswamy (2022)	ML pumpkin disease stages	Non-UAV	Effective aggregated features	Small-scale, non-UAV; ref [15]
5	Ouhami et al. (2021)	Overview of CV, IoT, and data fusion	Diverse	Broad scope	Review, not a new study; ref [16]
6	Kohut et al. (2025)	Classic machine learning for disease detection	Non-UAV	Computationally lightweight	Non-UAV; ref [20]
7	Wang et al. (2025)	Overview of non-destructive detection	Diverse	Reaffirming the non-invasive objective	General overview; ref [19]
8	Kusnandar (2023, ITB Dissertation)	Per-pixel HSV color selection	RGB UAV visible-light	A strong methodological foundation	Susceptible to inter-session radiometric variation (finding of this study); ref [7]
9	Kusnandar et al. (2021, 2023, 2024)	Vegetation index, white balance, HSV stability	RGB UAV visible-light	Step-by-step development of the basic method	Not yet tested on specific disease detection tasks; refs. [8, 9, 10]
10	This research	Plot-level aggregated RGB/HSV features + Random Forest	RGB UAV visible-light	Higher accuracy than RGB under plot-grouped validation (312 plots, 7 sessions); no cross-session transfer shown.	Plot-level treatment label (not severity); 1 species; labels confounded with field block

III. METHOD

A. Datasets

This study uses the UAV RGB Common Bean Disease Trial Dataset, containing individual plot images extracted from UAV imagery (DJI Matrice 300 with DJI Zenmuse L1 camera) in evaluation trials for bean disease (common bean, Phaseolus vulgaris L.) in Valle del Cauca, Colombia, in three acquisition seasons (2022, 2023, 2024) covering seven different shooting sessions dates. Treatment labels follow the official dataset documentation: the 2022 season has an image range of 1-60 for Control (non-inoculated/healthy), 61-96 for Common Bacterial Blight (CBB/sick) inoculation, and 97-132 for Angular Leaf Spot (ALS) inoculation. The last group was excluded from this study because ALS inoculation failed to establish disease symptoms in that season according to the dataset documentation. The 2023 and 2024 seasons have a range of 1-36 Control and 37-108 CBB (two inoculation blocks combined into one Diseased class).

After exclusion of ALS and filtering of images that could not be parsed, 732 images with valid labels were obtained: 300 images of the Healthy (Control) class and 432 images of the Diseased (CBB) class. All 732 images had their features successfully extracted (no images were skipped due to a lack of vegetation pixels detected). Throughout this paper, "Healthy" and "Diseased" denote the experimental treatment label at plot level (non-inoculated control vs. CBB-inoculated plots), not measured disease severity. Images of the same plot acquired in different sessions share one label; the 732 images therefore represent 312 unique plots over seven sessions.

B. Initial Evaluation: Limitations of Per-Pixel Segmentation

As a first step, the per-pixel HSV dual-reference method (a direct extension of Kusnandar's baseline [7], using two reference samples one healthy tissue, one diseased tissue to calculate two acceptability HSV ranges) was tested on image pairs from the same or different acquisition sessions. Initial results on a single pair of images (one Control, one CBB from the same session) looked promising, but testing on additional images from different acquisition sessions revealed a generalization failure: ground/shadow pixels in a given session, due to differences in camera exposure or color grading, fell in the same HSV range as the "sick" reference learned from another session, resulting in massive misclassifications. Fig. 2 illustrates this failure on a Control image (supposed to show very low levels of "sickness"), where the per-pixel dual-reference method instead misclassifies areas of ground and dark shadows as sick patches (red overlay). The failure was documented on illustrative images (Fig. 2) and was not quantified across the full dataset.

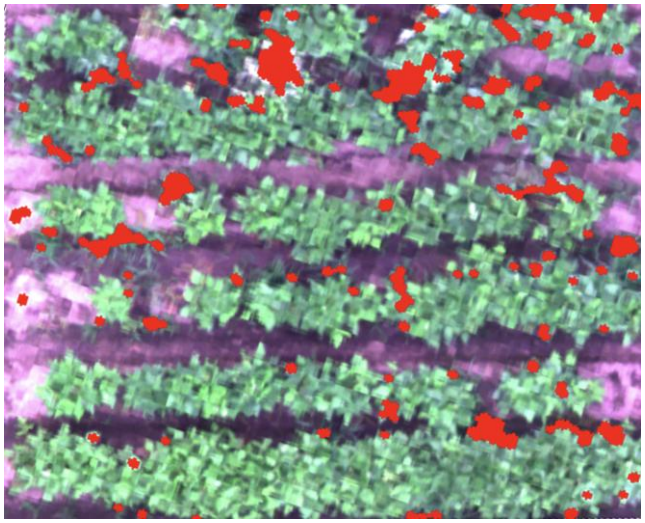

Fig. 2. Illustration of generalization failure of per-pixel dual-reference HSV segmentation on Control plot images: ground and shadow pixels (not diseased tissue) are incorrectly classified as diseased areas (red overlay), due to variations in color calibration between UAV acquisition sessions.

These findings are consistent with the warnings of Pourazar et al. [6] regarding the importance of radiometric calibration, and encouraging a revision of the method architecture from per-pixel segmentation towards plot-level aggregated feature-based classification, is explained in the following section.

C. Masking Vegetation (Operationalization of the Original Goal of the Kusnandar Method)

Prior to color feature extraction, vegetation pixels are separated from ground/background using the Excess Green (ExG) index with automatic Otsu thresholding, operationalizing the original intent of the Kusnandar color selection method (vegetation-vs-background separation) as a pre-processing stage that supports, rather than replaces, feature-based approaches:

$$ExG = 2G - R - B \quad (1)$$

The Otsu threshold is selected automatically per-image from the histogram of ExG values, thereby remaining adaptive to inter-session lighting variations without requiring a fixed threshold value that is susceptible to the same generalization problems described in Section III.B. Fig. 4 shows an example of the resulting vegetation mask.

D. Plot-Level Color Feature Extraction

Each plot image, after white balance correction following Kusnandar et al. [9] and vegetation masking, is summarized into eight aggregated features calculated only from vegetation pixels: R_mean, G_mean, B_mean (RGB channel), H_mean, S_mean, V_mean (HSV channel), GLI_mean (Green Leaf Index), and ExGR_mean (Excess Green minus Excess Red). This feature scheme deliberately imitates the feature scheme validated in a companion study on color-based plant disease classification, so that the two studies can be directly compared methodologically.

$$\mathrm{GLI} = \frac{(2\mathrm{G}-\mathrm{R}-\mathrm{B})}{(2\mathrm{G}+\mathrm{R}+\mathrm{B})} \quad (2)$$

$$ExGR = (2G - R - B) - (1.4R - G) \quad (3)$$

E. Feature Subset and Classifier Design

As in the companion study, two subsets of features are compared head-to-head with an equivalent number of variables (5 features each): RGB Features (R_mean, G_mean, B_mean, GLI_mean, ExGR_mean) and HSV Features (H_mean, S_mean, V_mean, ExGR_mean, GLI_mean), sharing two identical complementary features. The Random Forest used scikit-learn defaults (criterion = gini, max_depth = None, min_samples_split = 2, max_features = "sqrt", bootstrap = True) except n_estimators = 100 and random_state = 42; no hyperparameter tuning was performed. Software versions: Python 3.13.16, scikit-learn 1.6.1, OpenCV 4.14.0, SHAP 0.54.1. Models were evaluated with repeated stratified group k-fold cross-validation (k = 5, 5 repetitions, seeds 42–46) in which all images of a plot were kept in the same fold, and with leave-one-session-out and leave-one-season-out validation. Within each session, a random 5-fold and a spatially blocked 5-fold validation were also run; in the spatial variant, test folds were contiguous ranges of plot numbers within each class and two neighboring plot numbers were excluded from training (assuming that adjacent plot numbers are adjacent in the field).

F. Random Forest Model Formalism

Formally, each plot image is represented as a feature vector $x \in \mathbb{R}^5$ (five aggregated features per subset, Sections III.D-III.E) with a class label $y \in \{\text{Healthy, Diseased}\}$. Random Forest builds an ensemble of T = 100 decision trees $\{h_1(x), h_2(x), ..., h_T(x)\}$, each trained on independent bootstrap samples of the training data (bagging) with a random subset of features at each node split. The final prediction is determined by majority voting:

$$\hat{y} = \mathrm{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\} \quad (4)$$

Each h_t tree is constructed by selecting the split at each node that minimizes the Gini impurity at the resulting child nodes:

$$Gini(D) = 1 - \sum_i p_i^2 \quad (5)$$

where, D is the set of samples at a node and p_i is the proportion of class i samples at that node. The selected split is the split that maximizes the reduction in impurities (information gain):

$$\Delta Gini = Gini(D) - \left(\frac{n_L}{n}\right) Gini(D_L) - \left(\frac{n_R}{n}\right) Gini(D_R) \quad (6)$$

where, D_L and D_R are the left and right child subsets of the split results, respectively, and n_L, n_R, n are the number of samples in each subset. Random Forest was used because it handles correlated features and small samples well. The combination of averaging over many trees reduces estimation variance without significantly increasing bias, making it more resistant to noise in aggregated color features originating from variations in UAV acquisition conditions.

G. Statistical Significance Test

The 25 accuracy values from repeated cross-validation share most of their training data, so they are not independent, and a standard paired t-test on them underestimates the variance and overstates significance [46], [47]. The standard paired statistic and Cohen's d, used in the companion study, are shown below for reference only:

$$t = \frac{\bar{d}}{\left(\frac{s_d}{\sqrt{d}}\right)} \quad (7)$$

$$\mathrm{d} = \frac{\bar{\mathrm{d}}}{\mathrm{s_d}} \quad (8)$$

where, $\bar{d}$ is the mean HSV–RGB difference in accuracy across folds, s_d is its standard deviation, and n = 25 is the number of folds (5 folds × 5 repetitions).

Inference is based instead on the corrected resampled t-test of Nadeau and Bengio [47], which replaces 1/n in the variance term with 1/n + n_test/n_train: t_c = $\bar{d}$ / √((1/n + n_test/n_train)·s_d²), where n_test/n_train = 1/4 for 5-fold cross-validation, compared with a t distribution with n − 1 = 24 degrees of freedom. Cohen's d is reported only as a descriptive effect size. For the independent evaluations (LOSO and cross-season), McNemar's test [52] compares the paired errors of the two classifiers on the same held-out images, and 95% confidence intervals for accuracy are obtained by bootstrap resampling of held-out images (2,000 resamples), with resampling done by plot to respect the grouping. Post-hoc power analysis is not reported, because power computed from the observed effect size adds no information beyond the p-value.

IV. Results

No image was excluded for lack of vegetation pixels (an example mask is shown in Fig. 4); the ExG-Otsu mask was not, however, validated against manual annotation. The majority-

class baseline is 59.0% (432/732). Under plot-grouped cross-validation, HSV features achieved a mean fold accuracy of 66.9% (SD 2.4) and RGB features 59.1% (SD 3.5) (Fig. 3). On pooled out-of-fold predictions the accuracies were 67.2% and 58.3%; the difference of 8.9 points had a 95% plot-level bootstrap interval of 5.1 to 12.5 points (permutation test $p < 0.001$; corrected resampled t-test on fold accuracies $p = 0.001$). HSV reached a sensitivity of 76.2% and a specificity of 54.3% for the Diseased class; RGB reached 70.4% and 41.0%. HSV-only (H, S, V) and RGB-only (R, G, B) subsets gave 66.4% and 57.5%, so the two shared indices do not explain the gap. A vegetation-fraction feature alone gave 53.3%, below the baseline. Table II summarizes accuracy and balanced accuracy under plot-grouped, leave-one-session-out, and leave-one-season-out validation.

TABLE II. CLASSIFICATION ACCURACY OF RGB AND HSV FEATURE SUBSETS UNDER DIFFERENT VALIDATION SCHEMES

Validation	Features	Accuracy (%)	Balanced accuracy (%)
Plot-grouped 5x5 CV (mean of folds)	RGB5	59.1	55.7 (one repetition)
Plot-grouped 5x5 CV (mean of folds)	HSV5	66.9	65.2 (one repetition)
Leave-one-session-out (pooled)	RGB5	50.1	48.9
Leave-one-session-out (pooled)	HSV5	52.2	50.6
Leave-one-season-out (pooled)	RGB5	46.9	45.2
Leave-one-season-out (pooled)	HSV5	51.9	49.9
Majority-class baseline	none	59.0	50.0

RGB5 = R_mean, G_mean, B_mean, GLI_mean, ExGR_mean; HSV5 = H_mean, S_mean, V_mean, ExGR_mean, GLI_mean. Plot-grouped CV: stratified group 5-fold with plot as group, 5 repetitions; accuracy is the mean over 25 folds, and balanced accuracy was computed from the pooled out-of-fold predictions of one repetition. Leave-one-session-out and leave-one-season-out: accuracy and balanced accuracy computed from pooled predictions over all held-out sessions (7) or seasons (3). The majority-class baseline predicts the Diseased class for all images (432/732). With only three seasons, leave-one-season-out results were not tested statistically.

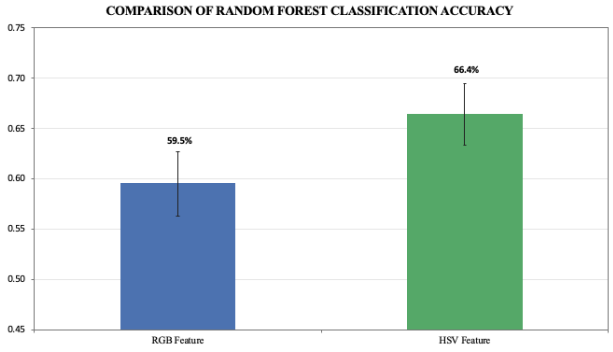

Fig. 3. Comparison of Random Forest classification accuracy, RGB Features vs HSV Features, on 732 common bean UAV plot images (standard deviation from repeated stratified 5-fold cross-validation, 5 repetitions).

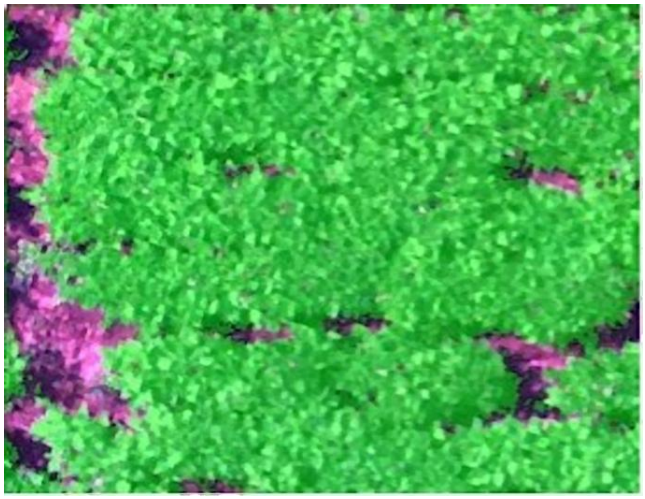

Fig. 4. Example of vegetation (green) versus background/soil masking with the ExG index and Otsu thresholding, the pre-processing stage before feature extraction; the figure is illustrative and does not indicate disease per pixel.

The absolute accuracy on both feature subsets (59.1%–66.9%) is moderate compared to the accuracy achieved on the curated feature dataset in the companion study (89-94%). This is consistent with the nature of the data: treatment labels are plot-level (all plots are labeled the same based on inoculation treatment), not leaf- or individual-pixel-level, so natural variations in symptom severity within a single plot (some areas of vegetation in a "Diseased" plot may not yet show obvious visual symptoms, and vice versa) adds noise to the labels that limits the maximum accuracy any classifier can achieve—a limitation on the ground truth of the data, not on the method itself. Plot-level labels and the confounding of treatment with field block (see Section V) are further limits on what these accuracies can show.

A. Cross-Season Generalization

Training on the 2022 and 2023 seasons (516 images) and testing on 2024 (216 images: 72 Healthy, 144 Diseased; majority baseline 66.7%) gave 61.1% accuracy for HSV (macro-F1 0.563) and 50.9% for RGB (0.498). Both are below the majority baseline, so this test does not demonstrate transferable discrimination. Leave-one-session-out validation gave 52.2% (HSV) and 50.1% (RGB), below the 59.0% baseline, with no significant difference (difference 2.0 points, 95% CI −2.6 to 6.5). Normalizing features within each session without labels (centering, z-score, or median/IQR scaling) gave 55.1%–60.5% for HSV and 54.1%–58.9% for RGB, again without a significant HSV–RGB difference; the highest value comes from the best of four variants and should be considered optimistic. Within single sessions, HSV features outperformed RGB features in all seven sessions under random 5-fold validation (mean 70.7% vs. 65.7%; mean session baseline 65.5%), with the strongest signal in the 2022 sessions (75.7%–82.9% vs. a 62.5% baseline). When test folds were spatially blocked, accuracy fell to 63.3% for HSV and 62.0% for RGB, and HSV exceeded the session baseline in two of seven sessions. Part of the within-session signal therefore reflects similarity between neighboring plots rather than treatment-specific color.

B. Feature Ablation: Marginal Contribution of Each Feature

To understand the source of HSV superiority in more detail, a feature ablation study was carried out: each feature was tested univariately (alone) and in a leave-one-out manner (subset of 5 features minus one). Fig. 5 shows the decrease in accuracy (compared to the full 5-feature baseline) when individual features are removed.

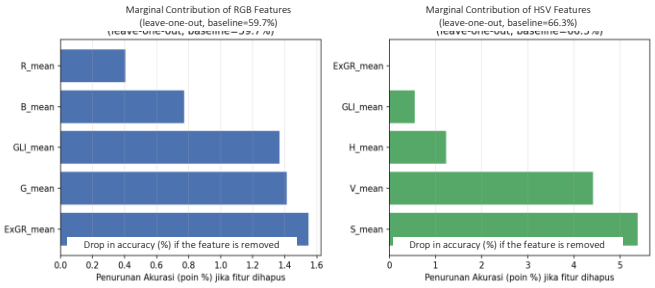

Fig. 5. Marginal contribution of each feature (leave-one-out) to classification accuracy, for a subset of RGB Features (left) and HSV Features (right).

Under plot-grouped cross-validation (25 folds), removing S_mean reduced HSV accuracy by 5.8 points (corrected t-test $p < 0.001$) and removing V_mean by 4.6 points ($p = 0.004$); removing H_mean changed accuracy by 1.1 points, and removing GLI_mean or ExGR_mean by less than 1 point (not significant). No single RGB feature produced a significant change. The leave-one-out bars in Fig. 5 were obtained with random cross-validation and show the same ordering. Saturation and Value are therefore the main contributors to the HSV subset, although the possible reasons (canopy density, greenness) were not tested directly.

C. Statistical Reporting

Confidence intervals and significance tests are reported at the beginning of this section; the bootstrap intervals from fold accuracies and the post-hoc power analysis reported in the earlier version were removed because folds of repeated cross-validation are not independent.

D. Plot-Level Explainable AI (SHAP) Analysis

SHAP values for the plot-level Random Forest model on both feature subsets were computed on a random 70/30 split and are reported descriptively, following the same methodology as the companion study.

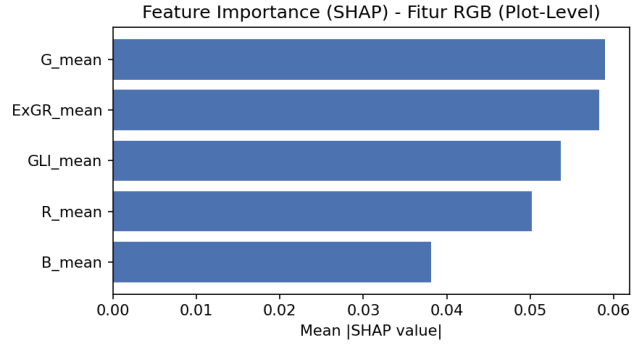

Fig. 6. Random Forest model feature importance (SHAP) plot for the RGB feature subset.

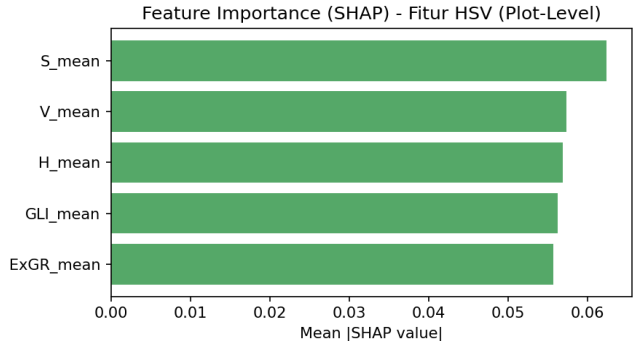

Fig. 7. Random Forest model feature importance (SHAP) plot for the HSV feature subset.

Interestingly, the SHAP patterns in Fig. 6 and 7 show a more balanced distribution compared to the leave-one-out patterns in Fig. 5 for both RGB (SHAP value range: 0.038–0.059) and HSV (range: 0.055–0.062); all five features in each subset contribute at relatively similar magnitudes, without the sharp dominance observed in the leave-one-out ablation. In the HSV subset, S_mean remains the highest SHAP contributor—consistent with the ablation findings—and V_mean also remains high; however, H_mean, GLI_mean, and ExGR_mean—whose contributions were near zero in the leave-one-out ablation—exhibit SHAP values comparable to those of S_mean and V_mean.

This difference in patterns is not a contradiction but rather a methodological consequence well-documented in interpretability literature: leave-one-out measures a feature's marginal contribution solely in the context of all other features being present (meaning a feature with redundant information appears to contribute little because others can "compensate" for its absence), whereas SHAP calculates the average marginal contribution across all possible feature subset combinations (including subsets that exclude other redundant features), thereby assigning fairer credit to features that possess overlapping information with others yet remain valuable on their own. The convergence of both methods on the finding that S_mean and V_mean are the top contributors to HSV features—across both ablation and SHAP analyses— is consistent with the grouped ablation above. Different patterns among other supporting features serve as a reminder that "feature importance" is not a monolithic concept but depends on the specific definition and measurement method employed.

V. Discussion

This study addresses the research questions posed in Section I in two parts. First, regarding the limitations of per-pixel segmentation: empirical evidence in Section III.B and Fig. 2 confirms that directly extending Kusnandar's method [7]— in its per-pixel dual-reference extension to disease detection—fails to generalize across UAV acquisition sessions. This aligns with the warnings from Pourazar et al. [6] regarding the need for radiometric calibration (previously documented for multispectral sensors) and demonstrates that these issues also apply—potentially even more severely, given the lack of built-in reflectance calibration—to the low-cost, visible-light RGB imagery used in Kusnandar's method.

Second, plot-level aggregated HSV features were more informative than RGB features when sessions were pooled, with Saturation and Value as the main contributors. However, this advantage did not transfer across sessions (leave-one-session-out 52.2% vs. 50.1%), and within sessions it shrank when test folds were spatially blocked (HSV 63.3% vs. 70.7% under random folds; mean within-session baseline 65.5%). Treatments occupy fixed field blocks, so treatment, position, and canopy structure cannot be separated with these data. We therefore do not claim that these features detect disease; they separate treatment plots better than RGB features under pooled validation. Differences in signal strength between sessions, which was strongest in 2022, could not be related to disease stage because inoculation dates were not available.

These findings complement the classical machine learning literature on plant disease detection [15] [16] [17] [20] [21] and specifically extend the precedent set by Görlich et al. [14]—who utilized pure RGB for UAV-based disease detection— by adding session-wise and spatially blocked validation, which that earlier study did not report. The ExG and Otsu-based vegetation masking method (Section III.C) also demonstrates that the original objective of Kusnandar's method (vegetation-versus-background segmentation) remains relevant and valuable; it serves not as a disease detection method in itself, but rather as an essential preprocessing stage for feature-based pipelines.

Several limitations should be noted. First, treatment labels are plot-level, not lesion-level; treatments occupy fixed field blocks, so treatment, position, and canopy structure cannot be separated, and label noise limits the achievable accuracy. Second, the study covers one species (common bean) and one disease (CBB, with ALS excluded because inoculation failed); validation on rice, maize, and chilli is future work. Third, the failure of per-pixel segmentation was documented on illustrative images rather than quantified across the dataset; the ExG-Otsu mask was not validated against manual annotation, and hue was averaged arithmetically. Fourth, inoculation dates and field layout were not available, so differences between sessions and spatial effects could not be analyzed further; the spatially blocked validation assumes that adjacent plot numbers are adjacent in the field. Fifth, no texture-based or deep learning baseline was evaluated. Model outputs should not be used for treatment decisions without field verification; at the observed accuracy, they are at most a screening indicator that requires verification by field inspection.

VI. Conclusion

This research documents that a per-pixel extension of the color selection method of Kusnandar [7] to disease classification does not generalize across UAV acquisition sessions (Fig. 2), and evaluates plot-level aggregated color features with a Random Forest on 732 field images (312 plots, seven sessions) as an alternative. Plot-level HSV features outperformed RGB features under plot-grouped validation (66.9% vs. 59.1%; majority baseline 59.0%), mainly through Saturation and Value, but this advantage did not transfer across acquisition sessions (52.2% vs. 50.1%) and decreased under spatially blocked validation within sessions. The ExG-Otsu masking stage served as a pre-processing step that supports rather than replaces feature-based approaches.

Practically, these findings provide an important warning for researchers and practitioners developing color-based disease detection systems on low-cost visible-light UAVs: static threshold-based per-pixel segmentation risks failure in multi-session or multi-date applications in the field, while aggregated feature approaches combined with classical machine learning performed better than per-pixel thresholding in our tests, although they did not yet transfer across sessions without the need for multispectral sensors or computationally expensive deep learning particularly relevant for resource-limited contexts. However, several limitations need to be honestly acknowledged: the treatment labels used are plot-level, not lesion-level, so variations in symptom severity within one plot add noise to the labels and also limit the absolute accuracy (66.4%) that can be achieved; validation is also still limited to one plant species (common bean) and one disease type (Common Bacterial Blight), with Angular Leaf Spot excluded due to unsuccessful inoculation in the 2022 season.

These limitations simultaneously open clear directions for further development: leaf-level or lesion-level annotation has the potential to increase the maximum achievable accuracy; validation on Indonesia's national strategic food crops such as rice, corn and chillies will test generalization across species; the combination of texture features (GLCM) with color features, as well as cross-session explicit radiometric calibration that goes beyond just per-image white balance, could potentially be the next step in methodological refinement. A companion paper, “Improve Accuracy for Plant Disease Classification using HSV Color Based”, reports a similar HSV-over-RGB comparison on a more curated soybean dataset and adds hyperparameter robustness and SHAP analyses that could be adapted to the field UAV context in future work; however, the two studies differ in data and task and are not directly comparable.

References

[1] L. Kouadio, M. El Jarroudi, Z. Belabess, S. E. Laasli, M. Z. K. Roni, I. D. I. Amine, et al., "A Review on UAV-Based Applications for Plant Disease Detection and Monitoring," Remote Sensing, vol. 15, no. 17, 2023, doi: 10.3390/rs15174273.

[2] T. B. Shahi, C. Y. Xu, A. Neupane, and W. Guo, "Recent Advances in Crop Disease Detection Using UAV and Deep Learning Techniques," Remote Sensing, vol. 15, no. 9, 2023, doi: 10.3390/rs15092450.

[3] D. Radočaj, P. Radočaj, I. Plaščak, and M. Jurišić, "Evolution of Deep Learning Approaches in UAV-Based Crop Leaf Disease Detection: A Web of Science Review," Applied Sciences (Switzerland), vol. 15, no. 19, 2025, doi: 10.3390/app151910778.

[4] A. Tripathi, U. Chourasia, P. Dixit, and V. Chang, "A survey: Plant disease detection using deep learning," International Journal of Distributed Systems and Technologies, vol. 12, no. 3, 2021, doi: 10.4018/IJDST.2021070101.

[5] A. Jafar, N. Bibi, R. A. Naqvi, A. Sadeghi-Niaraki, and D. Jeong, "Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations," Frontiers in Plant Science, vol. 15, 2024, doi: 10.3389/fpls.2024.1356260.

[6] H. Pourazar, F. Samadzadegan, and F. Dadrass Javan, "Aerial multispectral imagery for plant disease detection: radiometric calibration necessity assessment," European Journal of Remote Sensing, vol. 52, no. sup3, pp. 17-31, 2019, doi: 10.1080/22797254.2019.1642143.

[7] T. Kusnandar, "Metode Seleksi Warna dalam Ruang Warna HSV Berbasis Kamera Cahaya Tampak pada Unmanned Aerial Vehicle," Disertasi Doktor, Institut Teknologi Bandung, 2023.

[8] T. Kusnandar, J. Santoso, and K. Surendro, "Camera-Based Vegetation Index from Unmanned Aerial Vehicles," in Proc. 6th Int. Conf.

Sustainable Information Engineering and Technology (SIET '21), Malang, Indonesia, 2021, pp. 296-301, doi: 10.1145/3479645.3479661.

[9] T. Kusnandar, J. Santoso, and K. Surendro, "The Effect of White Balance for Color Constancy on Visible Light," in Proc. 2023 12th Int. Conf. Software and Computer Applications (ICSCA '23), Kuantan, Malaysia, 2023, pp. 144-149, doi: 10.1145/3587828.3587850.

[10] T. Kusnandar, J. Santoso, and K. Surendro, "Enhancing Color Selection in HSV Color Space," Ingenierie des Systemes d'Information, vol. 29, no. 4, pp. 1483-1491, 2024, doi: 10.18280/isi.290421.

[11] K. Neupane and F. Baysal-Gurel, "Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review," Remote Sensing, vol. 13, no. 19, 2021, doi: 10.3390/rs13193841.

[12] R. Chin, C. Catal, and A. Kassahun, "Plant disease detection using drones in precision agriculture," Precision Agriculture, vol. 24, no. 5, pp. 1663-1682, 2023, doi: 10.1007/s11119-023-10014-y.

[13] A. Abbas, Z. Zhang, H. Zheng, M. M. Alami, A. F. Alrefaei, Q. Abbas, et al., "Drones in Plant Disease Assessment, Efficient Monitoring, and Detection: A Way Forward to Smart Agriculture," Agronomy, vol. 13, no. 6, 2023, doi: 10.3390/agronomy13061524.

[14] F. Görlich, E. Marks, A. K. Mahlein, K. König, P. Lottes, and C. Stachniss, "Uav-based classification of cercospora leaf spot using rgb images," Drones, vol. 5, no. 2, 2021, doi: 10.3390/drones5020034.

[15] R. Ganesh Babu and C. Chellaswamy, "Different stages of disease detection in squash plant based on machine learning," Journal of Biosciences, vol. 47, no. 1, 2022, doi: 10.1007/s12038-021-00241-8.

[16] M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research," Remote Sensing, vol. 13, no. 13, 2021, doi: 10.3390/rs13132486.

[17] P. Vasavi, A. Punitha, and T. Venkat Narayana Rao, "Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: A review," International Journal of Electrical and Computer Engineering, vol. 12, no. 2, pp. 2079-2086, 2022, doi: 10.11591/ijece.v12i2.pp2079-2086.

[18] S. Zhang, X. Li, Y. Ba, X. Lyu, M. Zhang, and M. Li, "Banana Fusarium Wilt Disease Detection by Supervised and Unsupervised Methods from UAV-Based Multispectral Imagery," Remote Sensing, vol. 14, no. 5, 2022, doi: 10.3390/rs14051231.

[19] Y. Wang, J. Sun, Z. Wu, Y. Jia, and C. Dai, "Application of Non-Destructive Technology in Plant Disease Detection: Review," Agriculture (Switzerland), vol. 15, no. 15, 2025, doi: 10.3390/agriculture15151670.

[20] N. Kohut, O. Basystiuk, N. Shakhovska, and N. Melnykova, "Detecting Plant Diseases Using Machine Learning Models," Sustainability (Switzerland), vol. 17, no. 1, 2025, doi: 10.3390/su17010132.

[21] I. Ahmed and P. K. Yadav, "Plant disease detection using machine learning approaches," Expert Systems, vol. 40, no. 5, 2023, doi: 10.1111/exsy.13136.

[22] W. Albattah, A. Javed, M. Nawaz, M. Masood, and S. Albahli, "Artificial Intelligence-Based Drone System for Multiclass Plant Disease Detection Using an Improved Efficient Convolutional Neural Network," Frontiers in Plant Science, vol. 13, 2022, doi: 10.3389/fpls.2022.808380.

[23] I. Anam, N. Arafat, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture," Smart Agricultural Technology, vol. 9, 2024, doi: 10.1016/j.atech.2024.100647.

[24] L. W. Kuswidiyanto, H. H. Noh, and X. Han, "Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review," Remote Sensing, vol. 14, no. 23, 2022, doi: 10.3390/rs14236031.

[25] X. Zhang, L. Han, Y. Dong, Y. Shi, W. Huang, L. Han, et al., "A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images," Remote Sensing, vol. 11, no. 13, 2019, doi: 10.3390/rs11131554.

[26] D. Bohnenkamp, J. Behmann, and A. K. Mahlein, "In-field detection of Yellow Rust in Wheat on the Ground Canopy and UAV Scale," Remote Sensing, vol. 11, no. 21, 2019, doi: 10.3390/rs11212495.

[27] R. H. J. Heim, I. J. Wright, A. P. Allen, I. Geedicke, and J. Oldeland, "Developing a spectral disease index for myrtle rust (Austropuccinia psidii)," Plant Pathology, vol. 68, no. 4, pp. 738-745, 2019, doi: 10.1111/ppa.12996.

[28] R. M. Ualiyeva, M. M. Kaverina, A. V. Osipova, Y. B. Kairbayev, S. B. Zhangazin, N. N. Iksat, et al., "VNIR Hyperspectral Signatures for Early Detection and Machine-Learning Classification of Wheat Diseases," Plants, vol. 14, no. 23, 2025, doi: 10.3390/plants14233644.

[29] A. Wang, Z. Song, Y. Xie, J. Hu, L. Zhang, and Q. Zhu, "Detection of Rice Leaf SPAD and Blast Disease Using Integrated Aerial and Ground Multiscale Canopy Reflectance Spectroscopy," Agriculture (Switzerland), vol. 14, no. 9, 2024, doi: 10.3390/agriculture14091471.

[30] L. Radócz, C. Juhász, A. Tamás, Illés Á., and P. Ragán, "Multispectral UAV-Based Disease Identification Using Vegetation Indices for Maize Hybrids," Agriculture (Switzerland), vol. 14, no. 11, 2024, doi: 10.3390/agriculture14112002.

[31] D. Scutelnic, R. Muradore, and C. Daffara, "A multispectral camera in the VIS–NIR equipped with thermal imaging and environmental sensors for non invasive analysis in precision agriculture," HardwareX, vol. 20, 2024, doi: 10.1016/j.ohx.2024.e00596.

[32] J. M. Martínez-Heredia, A. I. Gálvez, F. Colodro, J. L. Mora-Jiménez, and O. E. Sassi, "Feasibility Study of Detection of Ochre Spot on Almonds Aimed at Very Low-Cost Cameras Onboard a Drone," Drones, vol. 7, no. 3, 2023, doi: 10.3390/drones7030186.

[33] S. Hemalatha and M. Sangeetha, "Identifying pests in precision agriculture using low-cost image data acquisition," Brazilian Journal of Biology, vol. 84, 2024, doi: 10.1590/1519-6984.281671.

[34] A. Joice, T. Tufaique, H. Tazeen, C. Igathinathane, Z. Zhang, C. Whippo, et al., "Applications of Raspberry Pi for Precision Agriculture—A Systematic Review," Agriculture (Switzerland), vol. 15, no. 3, 2025, doi: 10.3390/agriculture15030227.

[35] Y. S. H. Annadata, A. Thazhathethil, V. Moganarengam, and T. Nikoubin, "Enhancing Drone-Based Precision Agriculture: Performance Optimization of TinyML Models on Edge Devices and Adaptive Path Planning," SN Computer Science, vol. 6, no. 2, 2025, doi: 10.1007/s42979-024-03641-3.

[36] A. S. Ibrahim, S. Mohsen, I. M. Selim, R. Alroobaea, M. Alsafyani, A. M. Baqasah, et al., "AI-IoT based smart agriculture pivot for plant diseases detection and treatment," Scientific Reports, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-98454-6.

[37] M. El Sakka, M. Ivanovici, L. Chaari, and J. Mothe, "A Review of CNN Applications in Smart Agriculture Using Multimodal Data," Sensors, vol. 25, no. 2, 2025, doi: 10.3390/s25020472.

[38] E. Yilmaz, S. C. Bocekci, C. Safak, and K. Yildiz, "Advancements in smart agriculture: A systematic literature review on state-of-the-art plant disease detection with computer vision," IET Computer Vision, vol. 19, no. 1, 2025, doi: 10.1049/cvi2.70004.

[39] S. M. Shawon, I. A. Adan, M. Moniruzzaman, O. H. Sinan, M. J. Abed, M. Al Qwaid, et al., "Federated Learning Approaches in Precision Agriculture: A Systematic Review of Current Trends, Challenges, and Opportunities," IEEE Access, vol. 14, pp. 30513-30538, 2026, doi: 10.1109/ACCESS.2026.3667548.

[40] W. A. Ahmed, O. A. Abiola, D. Yang, S. F. Olatoyinbo, and G. Jing, "Integrating UAVs and Deep Learning for Plant Disease Detection: A Review of Techniques, Datasets, and Field Challenges with Examples from Cassava," Horticulturae, vol. 12, no. 1, 2026, doi: 10.3390/horticulturae12010087.

[41] Z. Ma, C. Wang, X. Wang, and X. Chen, "Advances in Artificial Intelligence-Enabled Crop Pest and Disease Detection: A Systematic Review," Agriculture (Switzerland), vol. 16, no. 12, 2026, doi: 10.3390/agriculture16121262.

[42] J. P. Nyakuri, C. Nkundineza, O. Gatera, and K. Nkurikiyeyezu, "State-of-the-Art Deep Learning Algorithms for Internet of Things-Based Detection of Crop Pests and Diseases: A Comprehensive Review," IEEE Access, vol. 12, pp. 169824-169849, 2024, doi: 10.1109/ACCESS.2024.3455244.

[43] H. M. Manoj, D. L. Shanthi, B. N. Lakshmi, K. J. Archana, E. V. N. Jyothi, and K. Archana, "AI-driven drone technology and computer vision for early detection of crop disease in large agricultural areas," Scientific Reports, vol. 16, no. 1, 2026, doi: 10.1038/s41598-025-32384-1.

[44] T. Wiesner-Hanks, H. Wu, E. Stewart, C. DeChant, N. Kaczmar, H. Lipson, et al., "Millimeter-Level Plant Disease Detection From Aerial Photographs via Deep Learning and Crowdsourced Data," Frontiers in Plant Science, vol. 10, 2019, doi: 10.3389/fpls.2019.01550.

[45] H. Zhao, Y. Yang, C. Yang, R. Song, and W. Guo, "Evaluation of spatial resolution on crop disease detection based on multiscale images and category variance ratio," Computers and Electronics in Agriculture, vol. 207, 2023, doi: 10.1016/j.compag.2023.107743.

[46] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," Neural Computation, vol. 10, no. 7, pp. 1895-1923, 1998, doi: 10.1162/089976698300017197.

[47] C. Nadeau and Y. Bengio, "Inference for the generalization error," Machine Learning, vol. 52, no. 3, pp. 239-281, 2003, doi: 10.1023/A:1024068626366.