International Journal of Technology

http://ijtech.eng.ui.ac.id

Research Article

Hyperparameter-Tuned Convolutional Neural Network Models for Unmanned Aerial Vehicle-Assisted Plant Pathology Monitoring

Jiaju Wang[1], Uzair Iqbal[1*], Nurzati Iwani Othman[2], Hamza Ibrahim[3]**

[1]Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

[2]School of Technology, Asia Pacific University of Technology and Innovation, 57000 Kuala Lumpur, Malaysia

[3]IT-Convergence Engineering, Kumoh National Institute of Technology, Gumi, 39177, South Korea

[*]Corresponding author: uzairiqbal@um.edu.my

[**]Corresponding author: nurzati.othman@apu.edu.my

Abstract: This study addresses the challenges of hyperparameter optimization in Convolutional Neural Networks (CNNs) for crop health classification under simulated UAV-based monitoring conditions using a publicly available multimodal dataset of laboratory-captured leaf images and environmental sensor readings. While CNNs have proven effective in agricultural image analysis, their performance is highly sensitive to hyperparameter configurations such as learning rate, batch size, and network architecture, particularly in multimodal, end-to-end training scenarios. Traditional methods, such as grid search and random search, are computationally inefficient for high-dimensional agricultural data. To overcome these limitations, this study implements advanced optimization techniques, including Bayesian optimization and Particle Swarm Optimization (PSO), to enhance CNN performance in terms of classification accuracy, training convergence, and robustness to environmental variability. The experimental results show that optimized CNNs significantly outperform both unoptimized baselines and traditional machine learning models. The proposed optimization framework contributes to scalable and adaptive CNN deployment in precision agriculture, supporting more reliable and efficient crop health monitoring in UAV-integrated systems.

Keywords: Bayesian optimization; Convolutional neural networks; Hyperparameter optimization; Particle swarm optimization; Precision agriculture; UAV-based crop monitoring

1. Introduction

1.1 Research Background

Precision agriculture uses technology to increase yield while reducing the environmental impact. UAVs with high-resolution sensors help monitor crop health and collect environmental data quickly and at scale, enabling frequent coverage that is difficult with manual field inspection (Abdullahi & Sheriff, 2022).

CNNs are effective for analyzing crop images (drone or lab) for crop classification, disease detection, stress detection, and yield prediction because they automatically learn useful visual patterns (Ajayi et al., 2024). However, CNN performance strongly depends on hyperparameters (e.g., learning rate, batch size, depth, and filter size) that must be manually tuned. Grid and random search are often too slow and costly, especially for multimodal data that combine images with sensor readings. Because agricultural data vary across crop types, growth stages, environments, and sensor setups, careful hyperparameter tuning is essential for robust performance across conditions (Mythili & Rangaraj, 2021). UAV imagery, CNNs, and environmental sensors support automated crop health monitoring (Taji et al., 2024), but success still depends on efficient hyperparameter optimization. Grid and random search do not scale well for large,

https://doi.org/10.14716/ijtech.v17i4.8333
Received January 2026; Revised February 2026; Accepted May 2026

variable multimodal datasets and may not achieve the robustness, adaptability, and efficiency required for real deployment (Naulia et al., 2024). Although Bayesian optimization and metaheuristic methods work well in other domains, they are less used for CNN tuning in UAV-based agriculture (Ibrahim et al., 2025). This leaves a clear gap: scalable optimization methods that reliably support robust multimodal crop monitoring across diverse conditions are lacking.

This study investigates how systematic hyperparameter optimization improves CNN performance, robustness, and adaptability for UAV-based crop health classification. It compares Bayesian optimization with Particle Swarm Optimization (PSO) and examines how each method affects stability, convergence, and generalization, as well as which hyperparameters most influence performance and efficiency.

We propose an optimized CNN framework and benchmark it against conventional tuning approaches and recent deep learning models to quantify improvements in accuracy, training efficiency, and robustness under UAV-simulated conditions. The experiments use a public multimodal dataset of laboratory leaf images and environmental sensor readings as a proxy for UAV inputs, allowing controlled evaluation without hardware deployment. Overall, this study shows that Bayesian and metaheuristic optimization can tune CNNs more effectively than grid and random search, thereby supporting more efficient and practical precision agriculture solutions.

1.2 Related Works

In recent years, CNN-based techniques have yielded excellent outcomes in UAV-assisted crop health monitoring. Nevertheless, a significant and chronic limitation of current work on UAV-CNNs is that they are tested in controlled and single-modality environments that are not representative of the real conditions of field work. In particular, most of the existing literature maximizes the performance on clean benchmark data that is balanced and pays little attention to the fact that model performance will become less optimal as it is exposed to sensor noise due to UAV vibration or altitude variations, uneven illumination of flight paths, and high ratio of classes that occur naturally when diseased crop samples are inherently rare in populations compared to their healthy counterparts. These do not constitute peripheral issues; they are the main failure modes that have been witnessed as the trained CNN models in laboratory environments are put to the real UAV environments. Moreover, although there is an increasing interest in the field of hyperparameter optimization, grid search and random search are the most prevalent methods to conduct studies in the field of agriculture when utilizing CNN, which are expensive to compute and do not scale well when using heterogeneous data, e.g. RGB images and spectral or environmental sensor readings. More adaptive algorithms, especially Bayesian optimization and metaheuristic algorithms, which can search the high-dimensional multimodal search space, are barely explored in this field. This systematic review of these gaps creates a case of a CNN optimization framework that would be explicitly designed with deployment-relevant constraints in mind and benchmark performance.

Plant pathology, a field that aims to diagnose and control plant diseases, is also embracing AI to achieve faster and more precise results (Ahuja & Zaheer, 2024). AI-supported tools enable better and more timely information detection than traditional methods, which tend to be manual and subject to human error (Anand, 2025). Stemming from the possibility of significant loss of the economy due to the disease of plants by viruses, fungi, and bacteria, fast and precise diagnosis must be performed to minimize damage to crops and to secure future yields (Jafar et al., 2024). AI has also been used to define plant science and precision agriculture, helping to intervene earlier and be more proactive in managing diseases and pests (Gupta et al., 2024). This is significant because diseases are only detected when the symptoms are too severe, which poses the threat of extensive infestation (Ali et al., 2024).

Conventionally, the detection of plant diseases has been dependent on the manual inspection of experts, which is time-consuming and challenging to scale, particularly in remote locations (Alhammad et al., 2025). Conversely, AI image processing and machine learning methods can be automated, saving time and expert resources to laboriously detect patients and render accurate diagnosis (Raiaan et al., 2024). In general, such a transition toward AI-based diagnostics offers more effective and accurate identification, categorization, and treatment of plant diseases than the traditional practice that requires many resources.

1.2.1 Gaps in the CNN for UAV Monitoring

Hyperparameter optimization is key to improving the performance of Convolutional Neural Networks (CNN) for crop health monitoring under simulated UAV conditions. Many studies focus on enhancing model accuracy but overlook important factors, such as environmental conditions, which significantly affect model behavior. (Ahmed et al., 2023) used Satin Bowerbird Optimization (SBO) with Bayesian optimization to improve CNN models for food crop classification using unmanned aerial vehicle (UAV) images. Although their method worked better than traditional techniques, they did not explore how this approach handles the differences in agricultural data, such as varying crop types and sensor configurations.

Teixeira et al., 2023 reviewed machine learning and deep learning techniques for UAV-based agriculture. They found a gap in the use of metaheuristic optimization techniques like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) for CNN. These methods have been used successfully in other fields but are not yet widely used in agriculture. This study highlights that Bayesian optimization needs to be adapted for agricultural data, which can be complex due to changes in environmental conditions and sensor types. (Corceiro et al., 2023) applied Bayesian optimization to CNN models for crop classification with unmanned aerial vehicle images. They focused on tuning hyperparameters, such as the learning rate and batch size. Their method showed potential; however, it struggled with the variability in agricultural data. They also noted that when working with large datasets, traditional methods, such as grid search and random search, are slow and inefficient. Their study highlights the need for more scalable and effective optimization methods to improve the performance of CNN in agriculture.

Mohanty et al., 2016 demonstrated the high accuracy of CNN models, such as AlexNet and GoogLeNet, in classifying plant diseases based on leaf images, achieving up to 99.35% accuracy on a curated dataset. However, their work did not focus on hyperparameter optimization, as key parameters such as learning rate, batch size, and dropout rate were not systematically tuned. More critically, their model performance dropped significantly (to ˜31% accuracy) when tested on real-world images acquired from diverse environments. This highlights a major limitation in the generalizability of CNNs trained under controlled conditions, especially when applied to UAV-based agricultural monitoring where sensor types, lighting, and environmental noise vary considerably. The absence of adaptive or environment-aware hyperparameter tuning in their pipeline exposes a gap that limits the robustness and scalability of CNN-based crop monitoring in real UAV deployments.

1.2.2 Optimization Methods for Enhancing CNN Robustness and Training Performance

Much recent work reports strong accuracy, but few studies explain how optimization affects training efficiency and convergence, which is important for computationally heavy CNNs. (El-Kenawy et al., 2022) proposed a hybrid metaheuristic (sine-cosine + gray wolf) for weed detection in UAV wheat images. They achieved high accuracy and F-scores using AlexNet features, but they did not analyze training efficiency or convergence behavior.

(PP et al., 2023) used Harris Hawks Optimization (HHO) to tune DenseNet-121 and DenseNet-201 for UAV-based weed detection. While performance improved and some convergence curves were shown, the paper provided limited discussion of training efficiency (e.g., time/epochs to convergence) or learning stability. (Ahmed et al., 2023) introduced SBODL-FCC, combining Satin Bowerbird Optimization (SBO) and Bayesian Optimization to tune a ConvLSTM model for UAV-based food crop classification, reporting over 97% accuracy. However, their evaluation focused mainly on accuracy and did not provide detailed evidence on efficiency or convergence dynamics.

Mohiuddin and Yousuf, 2025 proposed a single-stream CNN for multimodal plant disease detection by combining RGB, thermal, and hyperspectral data into a 4-channel input and using a modified VGG-16 for edge deployment. Although the model achieved 90% accuracy and handled fusion well, training used fixed hyperparameters and did not compare convergence speed or training efficiency. In addition, the study did not assess how Bayesian or metaheuristic optimization could improve training dynamics or reduce computational cost under complex multimodal conditions.

Beyond efficiency, recent studies highlight the need for robust CNNs under variable agricultural conditions. (Maghdid et al., 2024) discussed CNNs for large IoT-driven agricultural datasets and stressed adaptability but did not examine how Bayesian optimization could be tailored for highly variable datasets or across sectors. (Gülmez, 2024) reviewed CNN-based potato disease detection and emphasized the importance of hyperparameter optimization; however, he did not address how tuning approaches could be transferred to other crops or scaled to large monitoring settings. (Ngugi et al., 2024) noted that tuning can improve CNN accuracy for crop disease detection, but they did not study generalization across environments or integration with real-time acquisition systems.

Fuentes et al., 2017 developed a Faster R-CNN framework for detecting multiple tomato diseases and pests under challenging field conditions by training on diverse real-world images, improving robustness. However, the training parameters were fixed, and we did not explore whether adaptive hyperparameter optimization (e.g., Bayesian optimization) could further improve adaptability across environments or disease classes. This gap limits insight into how the framework would generalize under unseen conditions or changing imaging settings.

1.2.3 Evaluating Optimized Convolutional Neural Network vs. Traditional Methods in Agriculture

To evaluate the performance and efficiency of the optimized CNN compared with traditional methods, several studies compared CNN with classical machine learning models in agricultural tasks. (El Sakka et al., 2025) examined the use of CNN in smart agriculture for tasks such as weed detection, disease detection, crop classification, and yield prediction. They compared CNN with traditional methods like Support Vector Machines (SVMs) and decision trees. The study showed that the CNN worked better, especially with high-dimensional and time-based data from UAVs and satellites. However, it did not discuss how CNN performs with different agricultural data or environmental conditions, which is a gap in understanding how well CNN works across various agricultural settings.

Mohyuddin et al., 2024 reviewed machine learning methods in precision farming, focusing on tasks such as disease detection, pest control, and irrigation. The review showed that when optimized, CNN was better than traditional methods in terms of accuracy and efficiency, especially for image-based tasks. However, the study pointed out that CNN can be costly in terms of computation, and traditional methods, such as decision trees or random forests, could be improved to compete in large-scale farming. (Zualkernan et al., 2023) studied machine learning for precision agriculture using unmanned aerial vehicle imagery. They compared CNN with traditional methods such as SVMs and k-nearest neighbors (KNN) for tasks such as crop classification and pest detection. The results showed that the CNN was better, especially for image segmentation. The study also noted that CNN was better for real-time processing and high-resolution images but did not explore how CNN performs in different agricultural environments and with large datasets.

Wang et al., 2017 proposed a deep learning-based framework to estimate the severity of plant diseases from leaf images using a regression convolutional neural network (CNN) model trained on a labeled dataset with severity scores. This study compared CNN-based severity prediction with classical machine learning methods, including SVM and KNN, and found that CNN significantly outperformed traditional approaches in terms of accuracy and generalization.

2. Methods

2.1 Research Design

This study uses a quantitative experimental design to optimize CNN hyperparameters for crop health monitoring in simulated unmanned aerial vehicle conditions. Bayesian optimization and metaheuristic methods are applied to tune CNN settings to improve model performance, training efficiency, and adaptability in agricultural image analysis.

The CNN processes laboratory-captured leaf images representing UAV-style viewpoints for crop health classification. Hyperparameter optimization is important because agricultural data, including changes in environmental sensor readings and input conditions, vary widely. Since grid search and random search are often too slow for large datasets, Bayesian optimization and metaheuristic algorithms are used in this study to improve not only the classification results but also the training process. In addition to accuracy and generalization, the design emphasizes training efficiency and convergence behavior to support practical use in simulated UAV monitoring.

This approach combines image and sensor data to produce an optimized CNN that remains robust across different agricultural conditions. It considers variations in sensor measurements and input patterns that affect model adaptability and accuracy to support real-world applicability.

Finally, the optimized CNN models are benchmarked against a baseline (unoptimized) CNN and traditional machine learning classifiers such as SVM and random forest. To represent UAV-based monitoring conditions, all models use the same public multimodal dataset of leaf images and synchronized environmental sensor readings. The evaluation focuses on the predictive accuracy, training efficiency, convergence speed, and feasibility of quantifying the benefits of Bayesian Optimization and Particle Swarm Optimization. Figure 1 summarizes the workflow from data preparation and model design to optimization (Bayesian and PSO), training/validation, and benchmarking across the optimized CNN, baseline CNN, and traditional classifiers to develop scalable and robust models for UAV-based crop health monitoring.

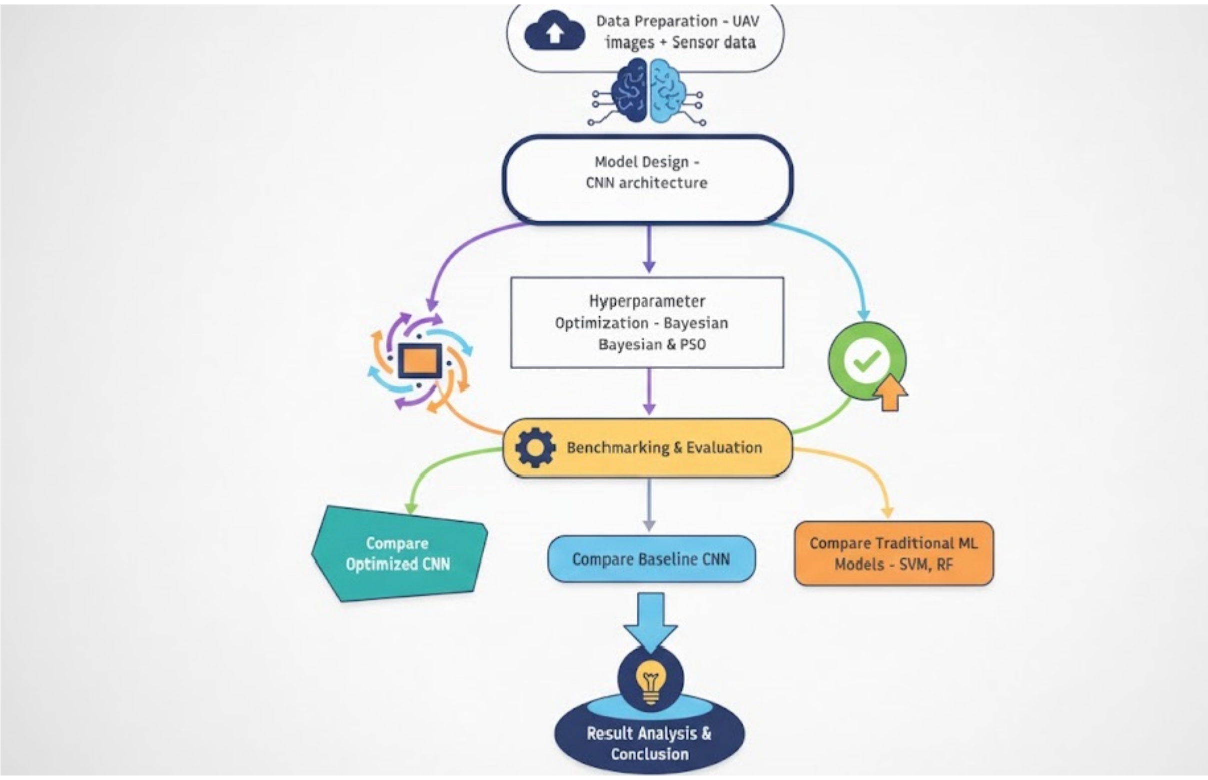

Figure 1 Overall research design and workflow

2.2 Data Collection and Integration

Data collection is critical for the development and validation of the Convolutional Neural Network (CNN) model and its hyperparameter optimization. In this research, a comprehensive series of datasets will be collected to ensure the effectiveness of the CNN model for crop health monitoring under simulated UAV conditions. The data collection process is summarized below.

2.2.1 Data Sources

A publicly available multimodal dataset combining plant leaf images with corresponding environmental sensor readings was adopted to achieve the research objectives.

Subham Divakar's Multimodal Plant Disease Dataset: This dataset contains over 6,000 paired plant leaf image records and corresponding environmental sensor data collected under controlled laboratory conditions. The dataset includes multiple plant species and disease types, along with sensor features such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. Each image was labeled as either healthy or unhealthy, forming a binary classification task. The strength of this dataset lies in its naturally aligned multimodal structure—each image sample is directly associated with sensor readings, allowing for early fusion in an end-to-end CNN framework. This eliminates the need for artificial data merging and ensures a consistent training environment. The diversity of crop types and environmental conditions makes it ideal for studying CNN performance under realistic precision agriculture scenarios.

2.2.2 Data Preprocessing

A structured data preprocessing pipeline was implemented to ensure the reliability of subsequent analyses and simulate realistic conditions encountered in UAV-based agricultural monitoring. The procedures were designed to improve data integrity, enhance model robustness, and reflect the environmental variability that is commonly observed in field scenarios.

1. Data Cleaning: Duplicate and incomplete records were removed, and consistency across image paths, sensor features, and labels was checked. Because the dataset did not provide explicit ground-truth labels, we assigned binary labels (healthy vs. unhealthy) by extracting directory names from the image file paths. This ensured that the data were suitable for supervised multimodal learning.

2. Simulated Sensor Variability: Field sensors often contain noise due to calibration drift and environmental fluctuations. To simulate this, we added Gaussian noise to each sensor feature, scaled by its original standard deviation, to introduce realistic variation while maintaining the overall distribution.

3. Image–Sensor Desynchronization Simulation: UAV monitoring can produce timing or location mismatches between images and sensor readings. To emulate this, 15% of the image paths were randomly reassigned while the sensor values remained unchanged. This created controlled misalignment to test model robustness under imperfect multimodal pairing.

4. Feature Normalization and Input Standardization: To ensure consistent scaling and stable training, all sensor features were standardized to zero mean and unit variance. All images were resized to 128×128 pixels to maintain a uniform input format compatible with CNN architectures.

5. Visual Data Augmentation: We applied augmentation during training to mimic real UAV imaging variation, including random rotation, brightness/contrast/saturation changes, Gaussian blur, perspective distortion, and occasional grayscale conversion.

These transformations represent practical effects, such as viewpoint changes, motion blur, and uneven lighting, which improve generalization.

6. Global Sample Shuffling: Finally, the dataset was shuffled and re-indexed to avoid ordering bias and support balanced splitting into training and validation sets.

7. Optimization Trial Budget: All strategies were evaluated under a budget-equivalent setting to ensure fair comparison across optimization methods. For Bayesian optimization, the search budget was chosen to reflect the point at which performance typically begins to stabilize in a moderately constrained hyperparameter space, allowing the surrogate model to identify promising regions without excessive sampling. For Particle Swarm Optimization (PSO), the budget was matched to a comparable population-by-generation setting so that the resulting search effort remained consistent across methods. Equivalent computational budgets were also assigned to grid search and random search baselines to avoid bias caused by unequal numbers of evaluations. This design ensures that any observed performance differences are attributable to the optimization strategy's effectiveness rather than to the search volume alone. From a deployment perspective, this trial budget primarily affects offline tuning time before deployment, whereas the selected hyperparameters influence the efficiency and stability of real-time on-edge inference during UAV operation.

8. Search Space Boundary: A fix to these boundaries prior to optimization means that a solution chosen by the optimizer is not only statistically performant but also physically implementable. Therefore, the search will not find optimal solutions in the case of unlimited resources but are not practically implementable in the desired UAV setting.

2.2.3 Dataset Integration Strategy

This study employed a pre-aligned multimodal dataset that integrates leaf imagery with corresponding environmental sensor measurements to support end-to-end multimodal modeling. The Multimodal Plant Disease Dataset by Subham Divakar was selected due to its coherent structure: each sample comprises a high-resolution plant image and an associated set of environmental variables, including nutrient concentrations, temperature, humidity, pH, and rainfall. Although the dataset does not explicitly include ground-truth labels, binary class labels indicating plant health status (healthy vs. unhealthy) were programmatically derived based on the directory structure embedded in image file paths. This labeling process was uniformly applied across the dataset and ensured consistency between visual features, environmental context, and classification outcomes. The dataset serves as a controlled proxy for UAV-based agricultural monitoring systems. The integration of sensor and image data within a single acquisition framework closely reflects the types of inputs collected by autonomous aerial platforms equipped with imaging and sensing devices. As such, it provides a reliable setting for evaluating CNN performance under multimodal input conditions.

2.3 Framework Development and Workflow

The framework combines advanced machine learning techniques with simulated UAV-based agricultural inputs to optimize CNNs for crop health monitoring. The framework aims to improve the performance, adaptability, and robustness of CNN across varying environmental and input conditions by using hyperparameter optimization methods such as Bayesian optimization and metaheuristic algorithms. This framework is implemented using an end-to-end multimodal CNN architecture, where image and sensor features are jointly learned during training.

2.3.1 Framework Structure

The framework is organized into the following three main phases:

1. Multimodal Data Preparation: A structured dataset containing image and environmental sensor inputs is organized for supervised learning. The binary class labels are programmatically assigned based on the image file path metadata. The preprocessing procedures introduce realistic noise and variation to simulate UAV-based monitoring conditions.

2. Training and Optimization: A multimodal CNN is trained using the prepared dataset. The hyperparameters are tuned using Bayesian optimization and metaheuristic algorithms to enhance the classification performance across varying input conditions.

3. Benchmarking and Comparative Evaluation: The optimized CNN models are benchmarked against an unoptimized CNN baseline and traditional machine learning classifiers (e.g., SVM and random forest). All models are trained and validated on the same dataset to ensure a fair comparison. The evaluation focuses on predictive accuracy, convergence speed, training efficiency, and robustness.

4. Result Analysis and Interpretation: The experimental results were analyzed to assess the effectiveness of different optimization techniques and model configurations. Performance metrics, such as accuracy, F1-score, and confusion matrices, are used to interpret the benefits of the proposed optimization strategies.

2.3.2 Key Components

1. Multimodal Input Configuration: A unified dataset comprising spatially aligned plant leaf images and corresponding environmental sensor measurements is used to construct the input pipeline. Binary classification labels denoting plant health status (i.e., healthy or unhealthy) are derived programmatically from structured image file paths. This configuration facilitates joint representation learning from visual and contextual data modalities in an early-fusion paradigm.

2. Multimodal Supervised Learning: A CNN is trained on the integrated image–sensor pairs to perform binary classification under simulated UAV conditions. To emulate environmental variability, the learning process incorporates dynamic image augmentation and standardized sensor scaling. A weighted loss function is applied to address inherent class imbalance, with class weights determined by the inverse frequency distribution of the target labels. This ensures equitable learning dynamics and mitigates bias toward the majority class.

3. Hyperparameter Optimization: Advanced hyperparameter optimization strategies are employed to enhance model generalizability and training efficiency. These include Bayesian optimization and population-based metaheuristic algorithms, which systematically explore the parameter space targeting learning rate, batch size, and network depth, among other variables. This component improves convergence stability and allows the model to adapt to input-level perturbations introduced during preprocessing.

4. Comparative Model Benchmarking: To contextualize the performance of the optimized CNN, traditional machine learning algorithms such as support vector machines (SVMs) and random forests, are implemented as baseline comparators. These models are evaluated on the same multimodal dataset using standard classification metrics, thereby enabling a comparative assessment of representational capacity and predictive accuracy across modeling paradigms.

2.3.3 Workflow Overview

1. Dataset Utilization and Preprocessing: The primary input source is a publicly available multimodal dataset comprising paired plant leaf images and environmental sensor readings. Binary classification labels (healthy vs. unhealthy) are programmatically derived from the path metadata of the image file. Image resizing, sensor value normalization, and augmentation techniques (e.g., rotation, brightness variation, and noise perturbation) are used to simulate UAV-related data variability and ensure standardized input across modalities.

2. Model Training and Optimization: A multimodal CNN is trained on the prepared dataset to perform binary classification. The model is optimized using Bayesian and metaheuristic algorithms to refine hyperparameter configurations, improve convergence dynamics, and enhance classification performance under variable input conditions.

3. Benchmarking and Evaluation: The performance of the optimized CNN models is benchmarked against a baseline CNN and traditional machine learning classifiers such as SVM and Random Forest. All models are evaluated on a consistent data split using standard metrics, including accuracy, F1-score, and confusion matrix. This comparative analysis demonstrates the benefits of hyperparameter optimization and informs model selection for UAV-based crop health monitoring applications.

2.4 Metrics for the Framework Evaluation

Evaluation metrics are essential for assessing the performance and effectiveness of the proposed framework for CNNs in crop health monitoring under simulated UAV conditions. These metrics ensure that the framework aligns with the research objectives by providing accurate predictions, effectively integrating laboratory-captured plant leaf images with simulated environmental sensor data, and providing actionable insights for agricultural decision-making. This section outlines the key evaluation metrics adopted in this research, encompassing both quantitative and qualitative assessments.

2.4.1 Predictive accuracy (quantitative analysis)

1. Accuracy: Ratio of correctly predicted instances to the total number of instances. This metric will evaluate the overall performance of the CNN model in classifying crop health conditions and assessing plant health status using jointly learned features from end-to-end fused image and sensor inputs under simulated UAV monitoring.

2. Precision, recall, and F1 score: These metrics are critical for evaluating the model in scenarios where class imbalance may exist (e.g., rare diseases or healthy samples). Precision reflects the accuracy of positive predictions, while recall measures the model's ability to detect all relevant cases. The F1 score combines both scores into a single performance metric. These metrics are especially important in assessing the generalizability of the model across varying environmental and physiological inputs.

3. Confusion Matrix: The confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives. This visualization offers insight into the CNN's ability to distinguish between different crop health categories, supporting model debugging and error analysis within the multimodal simulation setup.

2.4.2 Performance of Hyperparameter Optimization (Quantitative and Qualitative Analysis)

This section evaluates the impact of hyperparameter optimization techniques on CNN performance:

1. Improvement in Model Performance: The performance of CNN models before and after hyperparameter optimization (e.g., Bayesian optimization) is compared. Improvements in key metrics such as accuracy, precision, recall, and F1 score will be assessed.

2. Optimization Efficiency and Convergence: This component evaluates the optimization process's convergence behavior. The smoothness and stability of the training loss curves will be analyzed to determine the model convergence reliability. The primary focus of this study is to achieve consistent and efficient optimization performance under controlled experimental conditions.

3. Generalizability: The robustness of the optimized model is evaluated based on its ability to maintain stable performance under controlled input variability. This includes perturbations introduced through image augmentation techniques (e.g., rotation, blur, color jitter) and sensor noise injection. The stability across such conditions is interpreted as indicative of the generalization capacity within the simulated UAV-based monitoring environments.

3. Results

3.1 Data Preprocessing

This study used a publicly available multimodal dataset consisting of 6,427 samples, each comprising a laboratory-captured plant leaf image and a corresponding set of environmental sensor readings. The dataset is designed for binary classification, with each sample labeled as either unhealthy (label = 0) or healthy (label = 1), based on the image file path structure metadata.

A lightweight cleaning step was applied to the metadata to ensure basic integrity before modeling. Rows with invalid or missing image paths were removed, thereby preventing data loading failures and preserving image–sensor pairing consistency.

Each sample includes seven environmental sensor features: nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. These features represent key variables associated with crop growth and susceptibility to disease. The class distribution is moderately imbalanced, with 4,175 samples labeled as unhealthy ($\approx 64.96\%$) and 2,252 as healthy ($\approx 35.04\%$). A class-weighted loss function was employed during model training to mitigate this imbalance. Sensor features were standardized using z-score normalization implemented with the StandardScaler of Scikit-learn, ensuring that all features have zero mean and unit variance. In addition, small zero-mean Gaussian perturbations were added to each sensor channel with a standard deviation equal to the empirical standard deviation of that feature to ensure stress-test robustness under sensor uncertainty; negative values were clipped at zero when necessary. This simple perturbation mimics fluctuations encountered by a UAV system in field measurements without changing the label semantics.

Unlike sensor data, image augmentation was not applied offline but was incorporated dynamically during training using a PyTorch transform pipeline. Runtime augmentation includes resizing, random rotation, color jitter, Gaussian blur, random prospective distortion, and grayscale conversion. These augmentations were designed to simulate UAV-induced image variability and improve model generalization under real-world conditions. A controlled desynchronization was introduced to emulate the imperfect cross-modal synchronization encountered in field deployments. Fifteen percent of instances were sampled without replacement, and visual modality identifiers were randomly permuted within this subset while the corresponding sensor vectors remained unchanged. A fixed pseudo-random seed ensured replicability. This intervention preserves the marginal distributions and class labels while deliberately disrupting the one-to-one correspondence for the selected pairs, enabling a robustness assessment of the multimodal pipeline under moderate misalignment. This dataset serves as the foundation for all modeling and evaluation tasks in this study, supporting the benchmarking of CNN models under consistent input conditions and simulated UAV-based crop monitoring scenarios.

3.2 Exploratory Data Analysis (EDA)

3.2.1 Label Distribution Analysis

Several exploratory analyses were conducted on the preprocessed dataset to better understand the structure and characteristics of the input data. The class label distribution is moderately imbalanced, with 4,175 samples labeled as unhealthy (label = 0) and 2,252 samples labeled as healthy (label = 1) (Figure 2). This imbalance implies a potential bias toward the majority class during training, which was mitigated through a class-weighted cross-entropy in subsequent model training.

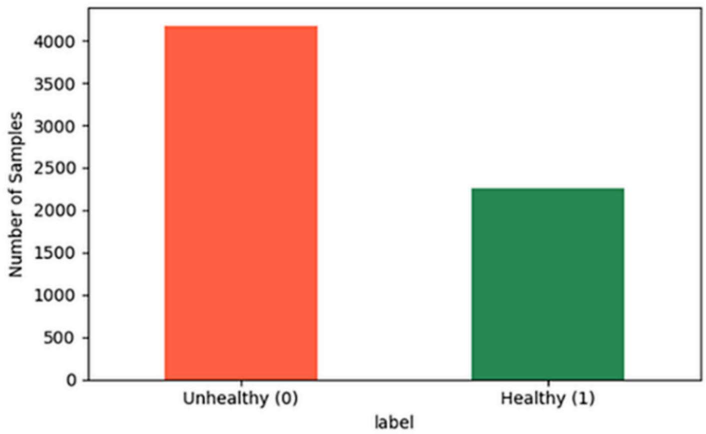

Figure 2 Class label distribution health monitoring status

3.2.2 Sensor Feature Distribution

To evaluate the statistical behavior of environmental sensor inputs, Figure 3 shows the distribution of each normalized feature. All sensor variables were standardized using z-score normalization, and their distributions reflect the inherent variability in agricultural measurements.

Among the features, temperature, humidity, and pH follow approximately normal distributions centered near zero, suggesting a balanced spread of values around the mean. In contrast, nutrient-related variables (N, P, K) and rainfall display noticeable right-skewness, with frequency concentrated on the lower value range. These skewed distributions indicate that many crop samples were associated with low nutrient levels and rainfall conditions.

These distributional patterns inform training expectations and motivate network standardization and regularization choices.

3.2.3 Correlation of Sensor Feature

Figure 4 shows the pairwise Pearson correlation coefficients among the seven standardized environmental sensor features. Most variable pairs demonstrate weak correlations, with values near zero, indicating low redundancy among the input features. Notably, phosphorus (P) and potassium (K) showed a moderate positive correlation of 0.43, suggesting a tendency for these two nutrient indicators to co-vary in certain samples. All other features exhibit negligible linear relationships, including temperature, humidity, and rainfall, which appear largely independent of one another and the nutrient group. This low inter-feature correlation supports the assumption that each sensor channel contributes unique information, thereby justifying their combined use in a multimodal learning framework.

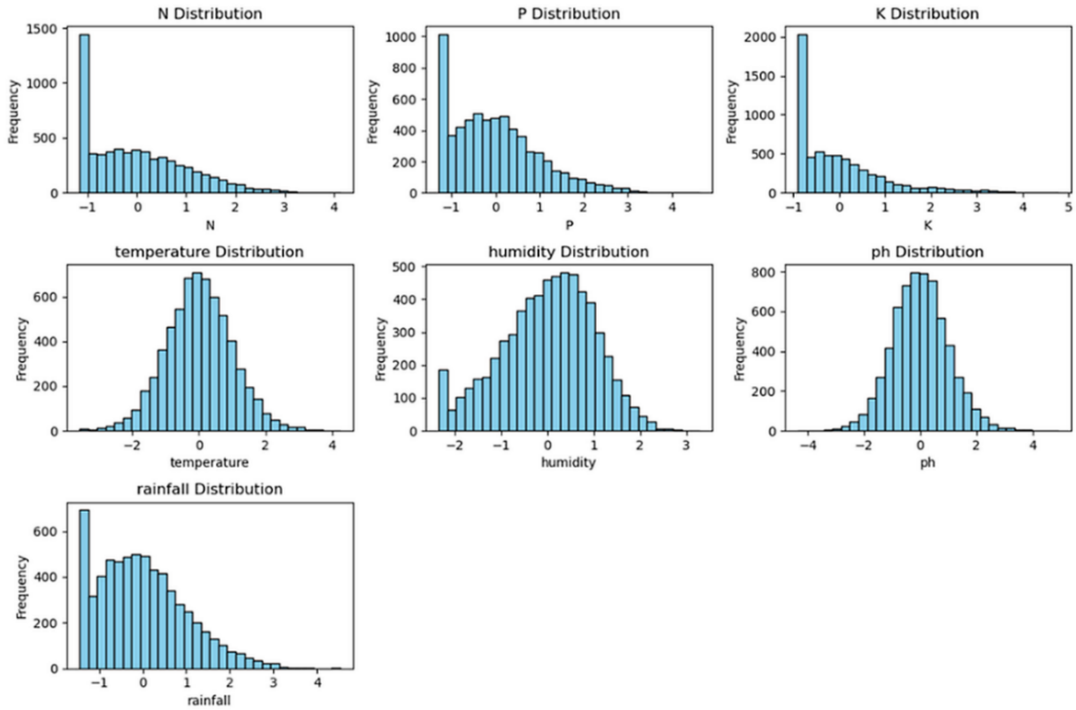

Figure 3 Distribution of the normalized Sensor features status

3.2.4 Feature–Label Correlation

The Pearson correlation coefficients between each environmental sensor feature and the binary classification label are presented in Figure 5. The analysis reveals generally weak correlations, which is expected given the complexity of plant health as a multifactorial outcome. Among the features, nitrogen (N) shows a weak positive correlation with the label (health = 1), suggesting that higher nitrogen levels are associated with healthier samples. In contrast, potassium (K), phosphorus (P), and humidity have weak negative correlations, indicating that elevated levels of these features may be slightly associated with unhealthy plants. Temperature, pH, and rainfall exhibited near-zero correlations. These low but interpretable associations help justify the inclusion of all sensors features in the model, as each contributes subtle but potentially complementary health prediction signals.

3.2.5 Visualization of Image Samples

Figure 6 illustrates representative image samples from the dataset of both classes. The top row shows samples labeled as unhealthy (label = 0), which typically exhibit leaf spots, discoloration, wilting, or surface damage. In contrast, the bottom row contains healthy samples (label = 1), characterized by uniform green coloration, intact structure, and the absence of visible disease indicators.

These visual differences provide an intuitive understanding of the classification task and highlight the variability in lighting, background, and leaf orientation. To improve model robustness against such variations, image augmentations—such as rotation, brightness adjustment, and blurring—were applied dynamically during training, as described in Sec. 4.1 and Sec. 4.3.1.

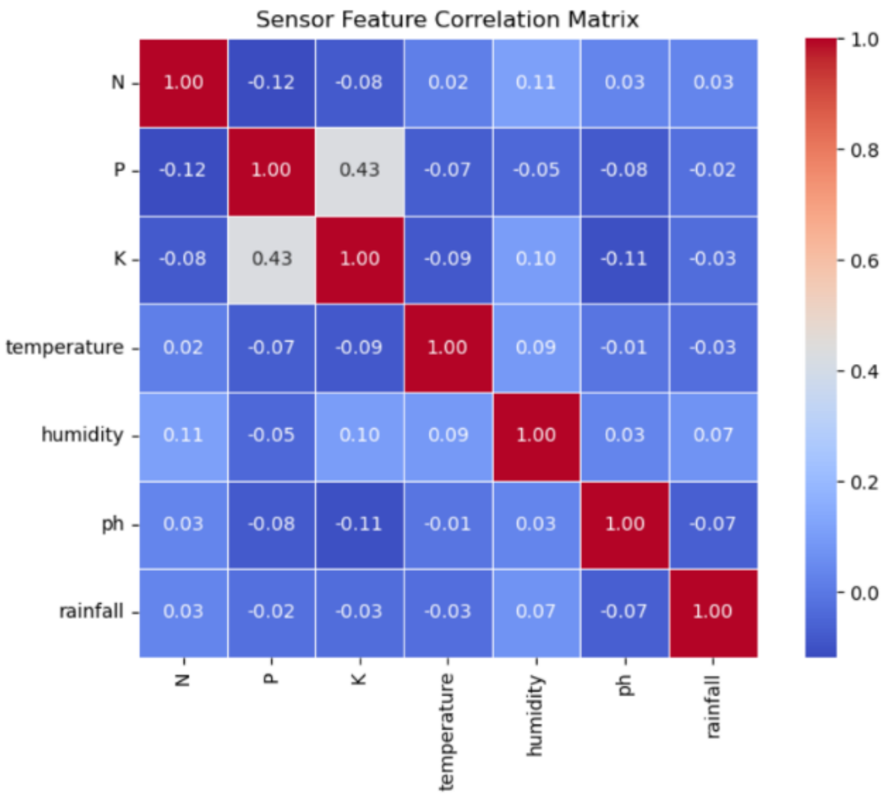

Figure 4 Pearson correlation matrix of the standardized environmental sensor features

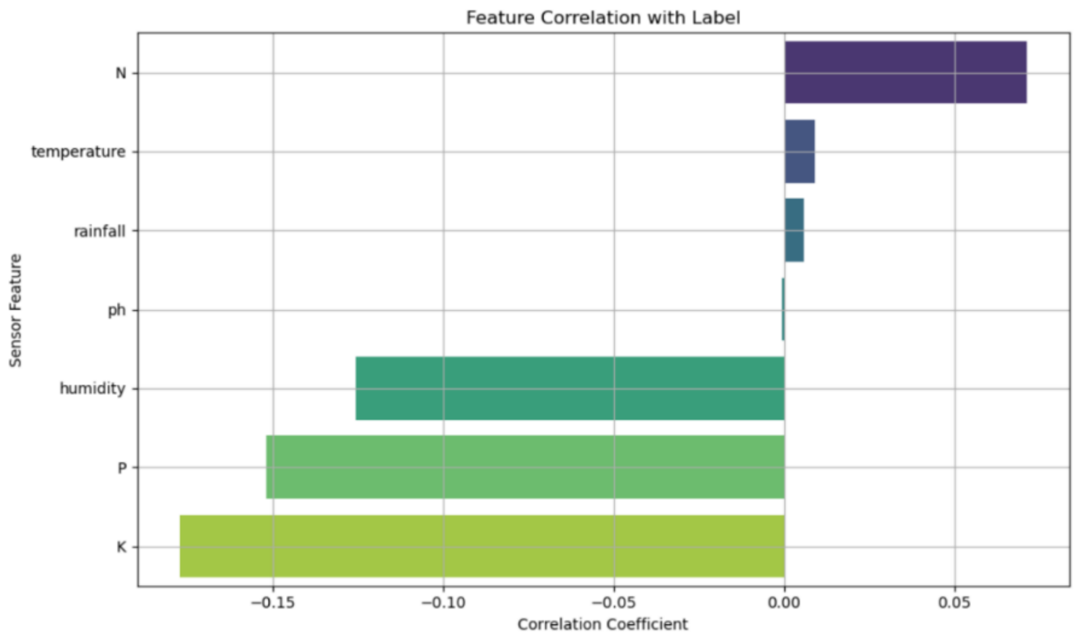

Figure 5 Pearson correlation coefficient between sensor features and the classification label

3.3 Experimental Setup and Results

3.3.1 Data Split and Preprocessing

All experiments start from the preprocessed dataset summarized in Section 4.1 (procedural details in Section 3.2.2). Data were partitioned into training and validation subsets using an 80/20 stratified split with a fixed pseudo-random seed (42) to preserve class proportions and ensure reproducibility. The resulting indices are fixed and reused across all models for fair comparison.

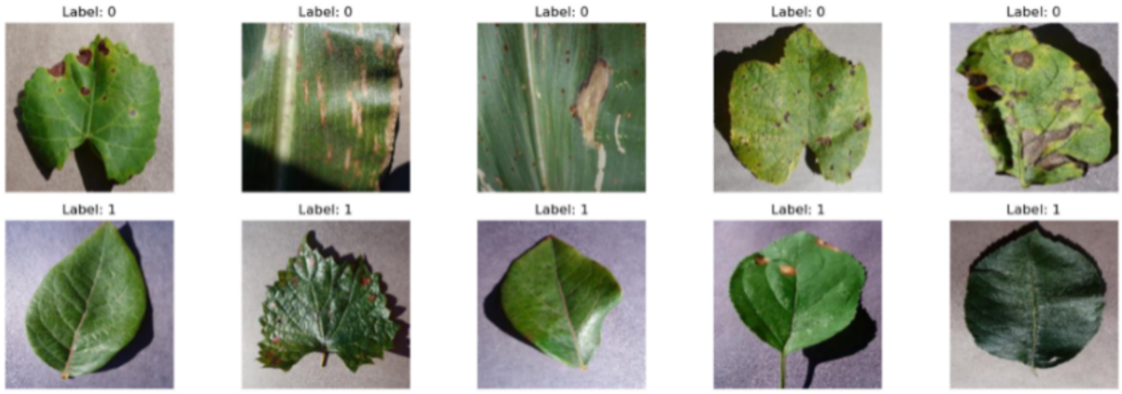

Figure 6 Representative leaf image samples, label 0 means unhealthy and 1 means healthy

Sensor inputs in this section refer to the standardized features prepared in Section 4.1 (z-score normalization applied to the seven variables: nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall). Loss weighting follows Section 4.2.1, where class-weighted cross-entropy compensates for the moderate label imbalance.

Non-symmetric preprocessing/augmentation is adopted for images to balance robustness and unbiased evaluation. During training, images undergo on-the-fly stochastic transforms—resize to 128×128, random rotation (±15°), color jitter (brightness/contrast/saturation/hue), Gaussian blur, random perspective, and random grayscale with probability 0.2—followed by tensor conversion. During validation, only deterministic preprocessing (resize to 128×128 and tensor conversion) is applied, with no stochastic augmentation. The 128×128 input resolution is chosen as a practical compute–throughput trade-off under the available hardware budget, and the same setting is maintained across the baseline and optimized CNN configurations. The fixed split and preprocessing/augmentation policies described above are shared across all experiments, ensuring consistent evaluation conditions for CNN-based models and comparative classifiers that rely on visual features extracted from CNNs.

3.3.2 Traditional Classifiers Using Feature Extraction

To provide a performance baseline against the proposed CNN-based models, two traditional machine learning classifiers—Support Vector Machine (SVM) and Random Forest (RF)—were trained using extracted features. Instead of directly feeding the raw sensor inputs, a two-part feature vector was constructed for each sample.

Image features were extracted using a pretrained CNN model based on ResNet34. The convolutional layers were used as a fixed feature extractor, and the classification head was removed. Each image was resized to 128×128 (consistent with Sec. 4.3.1) and passed through the CNN backbone, producing a 512-dimensional feature vector from the final convolutional output (after global aggregation). These image features were then concatenated with the original 7-dimensional standardized sensor vector to form a unified 519-dimensional representation per sample. The extracted features were generated for both the training (5,141 samples) and validation sets (1,286 samples), ensuring consistency across all evaluations. To maintain an unbiased evaluation, feature extraction was non-symmetric: training features followed the stochastic image transforms used by the CNN pipeline (Sec. 4.3.1), whereas validation features were computed with deterministic preprocessing only. These vectors were saved and used later to train the traditional classifiers.

The SVM adopted a radial basis function kernel with probability estimates enabled, and the Random Forest used 100 decision trees. Both classifiers were implemented using scikit-learn and initialized with a fixed pseudo-random seed of 42 to ensure replicability. As shown in Figure

7, the SVM classifier achieved an accuracy of 84.76% and an F1 score of 0.781, whereas the RF classifier achieved an accuracy of 83.98% and an F1 score of 0.771. Confusion matrices (Figures 8 and 9) demonstrate that both models performed well in distinguishing healthy and unhealthy leaf samples, with SVM exhibiting slightly higher sensitivity to the minority class (healthy, label = 1), consistent with its higher macro-F1.

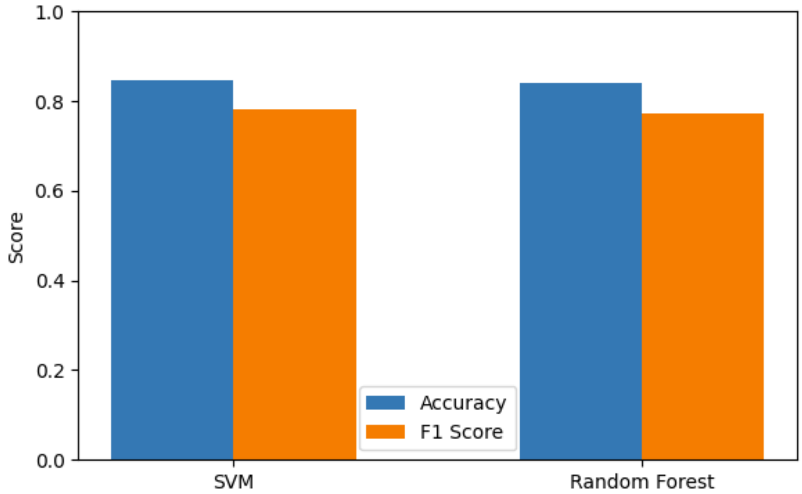

Figure 7 Comparison of accuracy and F1 score

3.3.3 Baseline Convolutional Network Architecture and Training

The baseline model in this study is a multimodal convolutional neural network (CNN) designed to integrate visual and sensor data in an end-to-end learning pipeline. The architecture comprises three main components: an image feature extractor, a sensor feature encoder, and a fusion-based classifier. The image branch uses a pretrained ResNet34 model as the backbone. The original classification layer is removed, and the output feature vector obtained after global aggregation from the last convolutional stage is used for downstream fusion. The sensor branch consists of a two-layer multilayer perceptron (MLP) with 32-dimensional hidden layers, ReLU activation, and regularization dropout rate of 0.5.

The image and sensor branches' outputs are concatenated and passed into a fusion classifier. The fusion module contains a linear projection to a 32-dimensional feature space, followed by ReLU activation and dropout, and a final output layer for binary classification. During training, image data are dynamically augmented using resize to 128×128 (rather than random resizing), rotation (±15°), color jittering, Gaussian blur, perspective distortion, grayscale conversion, and conversion to tensor. For validation, only deterministic preprocessing (resize to 128×128 and tensor conversion) is applied, without any stochastic augmentation, to ensure an unbiased generalization estimate. Sensor data are standardized using z-score normalization, and the model consumes paired multimodal samples from the preprocessed dataset described in Section 4.1. A stratified 80/20 split was used to construct the training and validation sets. The SGD optimizer was used to train the model (learning rate 3×10^{-4}, momentum 0.9) with a batch size of 32. The loss function is class-weighted cross-entropy with label smoothing of 0.2, addressing the moderate imbalance between the unhealthy and healthy classes.

The baseline CNN achieved a validation accuracy of 76.83% and an F1 score of 0.7301 (Figure 10). The training curve shows stable improvements in both validation metrics. The confusion matrix (Figure 11) indicates moderate performance, with elevated 0→1 misclassifications (i.e., unhealthy were predicted as healthy).

3.3.4 Hyperparameter Optimization and Optimized CNN

To improve the classification performance of the baseline multimodal CNN, Bayesian hyperparameter optimization was conducted using the Optuna framework. In this optimization process, a multidimensional search space was systematically explored to identify the most effective configuration of the model and training parameters. The goal of this study was to maximize the validation F1-score, a robust metric for imbalanced binary classification.

The following hyperparameters were tuned jointly:

- Learning rate: log-uniform sampled within [1e-4, 1e-2]
- Dropout rate (applied to the fusion layer): sampled within [0.1, 0.5]
- Fusion layer dimension: chosen from {64, 128}
- Batch size: {16, 32, 64}
- Optimizer type: either Adam or SGD

Each trial involved initializing a new CNN instance with the sampled configuration and training it on the prepared training subset. Label smoothing ($\alpha = 0.2$) and class weighting were employed to mitigate label imbalance and to prevent overconfidence in predictions. The validation set was used to compute the macro-averaged F1-score, which served as the optimization objective.

The optimization process was executed over 15 independent trials, with a total time budget of 7200 seconds. After the search was completed, the configuration that yielded the highest validation F1-score was selected for the final evaluation. The best configuration discovered by Optuna was as follows:

- Learning rate: 0.00033720807288716237
- Dropout rate: 0.10711114487010638
- Fusion layer dimension: 64
- Batch size: 64
- Optimizer: Adam

The CNN was retrained from scratch using the same data pipeline, image augmentations, and preprocessing strategy as the baseline model. The final evaluation showed that the optimized CNN achieved a validation accuracy of 0.8227 and an F1-score of 0.7849.

As shown in Figure 12, the training curve for the optimized model exhibited steady improvement in classification accuracy and F1-score while maintaining a consistent decline in training loss. The confusion matrix in Figure 13 further confirms that the optimized model effectively distinguishes between healthy and unhealthy samples, reducing false negatives and improving overall reliability compared with the baseline.

In addition to Bayesian methods, PSO was employed as a metaheuristic alternative to explore the same hyperparameter space. The PSO algorithm simulates a swarm of particles that iteratively update their position based on local and global optima. A swarm of 10 particles was initialized and evolved over 15 iterations. The search space remained identical to that of the Bayesian optimization: learning rate, dropout rate, fusion dimension, batch size, and optimizer.

The same strategy was used to evaluate each particle's configuration: initializing a CNN model, training for two epochs per evaluation, and computing the validation F1-score. The best-performing configuration identified by PSO was as follows:

- Learning rate: 0.0001

- Dropout rate: 0.38878184259421
- Fusion dimension: 64
- Batch size: 32
- Optimizer: Adam

Retraining the CNN using this configuration resulted in the highest performance among all models:

- Validation accuracy: 0.9059
- F1 score: 0.8675

Figure 14 shows that the training curve exhibited continuous performance improvement, suggesting effective generalization without overfitting. The confusion matrix (Figure 15) showed a balanced classification, which reduced misclassifications of healthy samples.

In conclusion, both optimization strategies improved the baseline CNN, but the PSO-optimized model achieved superior results. This validates the effectiveness of metaheuristic methods in high-dimensional and hybrid hyperparameter search spaces.

3.3.5 Evaluation Metrics

To ensure comparability, the evaluation strictly follows the protocol described in Section 3.4. Results are reported with accuracy and the macro-averaged F1-score as primary metrics, complemented by confusion matrices and training curves for diagnostic analysis. The accuracy was calculated as the proportion of correctly classified samples out of the total validation set. While intuitive and easy to interpret, accuracy alone can be misleading in imbalanced classification tasks, such as crop health prediction, where one class (e.g., unhealthy samples) dominates the dataset. The macro-averaged F1-score was employed as the central evaluation metric to address this limitation. The F1-score harmonizes precision and recall into a single value, providing a robust estimate of classification effectiveness by equally weighting both classes regardless of their frequency. This is particularly important in agricultural applications, where misclassifying a diseased plant as healthy may result in substantial yield loss.

Confusion matrices were generated for all models to visually analyze the distribution of true positives, false positives, true negatives, and false negatives. These matrices offered further insights into model bias, particularly whether a given classifier tended to favor the majority class or exhibited improved balance across categories after optimization. Notably, the optimized CNN models showed a reduction in false negatives—indicating better detection of healthy samples—even under challenging feature variation.

Training and validation curves were recorded throughout the learning process for deep learning models. These plots tracked changes in classification accuracy, F1-score, and loss over multiple model update iterations. The curves helped monitor the learning dynamics of the model and identify symptoms of underfitting or overfitting. Comparisons across curves also revealed how hyperparameter optimization strategies, such as Bayesian optimization and PSO, contributed to smoother convergence and more stable generalization.

Overall, this multi-faceted evaluation approach ensured that the models were not only compared in terms of their final performance scores but also understood in relation to their training behavior and prediction distribution. This enabled a more nuanced comparison between baseline and optimized architectures, as well as between deep learning and traditional approaches.

3.3.6 PSO vs. Bayesian Optimization

The excellent results of PSO compared to Bayesian optimization in this paper are not by chance but a structural fit between the metaheuristic search algorithm and the nature of the hyperparameter landscape presented in this paper. Bayesian optimization works by building a probabilistic surrogate of the objective function and randomly choosing trials that will trade off exploration of uncertain space with exploitation of well-known high-performing space. The strategy is best suited when the objective surface is smooth and the trials are sufficient to allow the surrogate to obtain an accurate approximation. The wall-clock constraint is used to computationally explain the 15-trial budget in the current work, but the budget represents a rather sparse sample of the 5-dimensional mixed space of search spaces with continuous variables, including learning rate and dropout, as well as discrete ones, including batch size, fusion dimension, and optimizer type. By comparison, PSO does not have a surrogate model but rather a population of 10 particles that traverses the space by a momentum-based position displacement based on the individual and swarm-level best observation. This parallel search is a less sensitive population-level exploration, more likely to emerge from coherent local optima during the initial search phases, and capable of breaking smoothness assumptions that are violated by distinct dimensions. The low dropout rates and Adam optimizer discovered by PSO, which are significantly different from those discovered by Bayesian search, indicate that the two techniques converge to different regions of the objective surface, where PSO found a region that better generalized to the class-weighted training objective. This structural explanation is more explanatory than an entirely metric-based comparison, and each method has a better or worse fit to the problem.

3.3.7 Statistical Consistency and Run-to-Run Variance

Each model configuration was tested using a five-run repeated experiment protocol to measure the result stability and offer a statistically-founded foundation to compare results across methods. The training runs were performed with an independent random seed that was selected within the fixed set of numbers 42, 7, 123, 256, 99 so that the results could be reproduced. The training was run with controlled variation in weight initialization and data shuffling. The accuracy and F1-score at the end of training were logged, and descriptive statistics were calculated among the five repetitions.

The reference values of the performance of each model are established using the single-run results obtained in the confusion matrices shown in Figures 9, 11, 13, and 15. The PSO-optimized CNN had a validation accuracy of 90.59% and an F1-score of 86.75%, which are associated with a TP = 396, TN = 769, FP = 66, and FN = 55 classification results with 1,286 test samples. The Bayesian-optimized CNN recorded an accuracy of 82.27% (F1 = 78.49%), the random forest achieved 83.98% (F1 = 77.11%), and the baseline CNN reached 76.82% (F1 = 73.10%). Tables 1 and 2 consolidate all metric values read directly from Figures 9–16.

Table 1 Confusion matrix cell values

Model	TN	FP	FN	TP	Total
Random forest (Figure 9)	733	102	104	347	1286
Baseline CNN (Figure 11)	585	250	48	403	1286
Bayesian (Optuna) CNN (Figure 13)	642	193	35	416	1286
PSO-optimized CNN (Figure 15)	769	66	55	396	1286

Table 2 Classification metrics derived from confusion matrices

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	83.98	77.28	76.94	77.11
Baseline CNN	76.82	61.72	89.36	73.10
Bayesian (Optuna) CNN	82.27	68.31	92.24	78.49
PSO-Optimized CNN	90.59	85.71	87.81	86.75

3.3.8 UAV Data Generalization to Real-Field UAV Data

This study was conducted in a simulated multimodal space, in which sensor noise, desynchronization, and imaging variation were programed and not recorded during real UAV operations. Although this controlled environment allowed the systematic assessment of the optimization framework under controlled and reproducible conditions, the sources of variability that real UAV field data would contain must be excluded, such as GPS-linked geospatial drift, non-Gaussian sensor artifacts due to mechanical vibration, the effect of cloud cover transitions on lighting discontinuities during flight, and the geometry of the crop canopy on spectral reflectance, which varies with solar angle and row orientation. The training and augmentation pipeline used in this case is a class-weighted one that is explicitly trained to transfer to such conditions, although this should be empirically validated on a real field dataset before deployment claims can be made. A straightforward and practical extension of the evaluation to open UAV agricultural datasets, including those obtained in different canopy cover and flight altitude settings, would represent a straight forward and quantifiable next step toward proving the evaluation's usefulness in a non-simulated setting.

4. Discussion

4.1 Gaps in CNN Hyperparameter Optimization

The first research objective was to identify the gaps in the current hyperparameter optimization studies for CNN-based models in UAV-enabled crop health monitoring scenarios. A comprehensive review of recent literature revealed several critical deficiencies that this study sought to address.

Firstly, most existing research in the domain of CNN optimization focuses on unimodal inputs—typically image-only data—without considering the integration of environmental sensor data. However, real-world UAV-based agricultural applications often involve multimodal signals (e.g., images + sensor data), where environmental variables such as temperature, humidity, and nutrient levels significantly influence plant health. The lack of hyperparameter studies that consider this multimodal context limits the applicability and generalization of the proposed solutions in real agricultural settings.

Second, this study proposes a CNN architecture specifically designed for multimodal input fusion, incorporating both image-based features and standardized sensor measurements. The implementation includes a well-defined end-to-end training pipeline that can handle heterogeneous data streams. This design fills an architectural and methodological gap in the current state of the art, where most optimization efforts have been constrained to image-only pipelines (e.g., (Ahmed et al., 2023; Corceiro et al., 2023)).

Third, while existing research has largely adopted single-strategy hyperparameter optimization (e.g., grid search or random search), this study introduces a comparative analysis of two advanced strategies, namely, Bayesian optimization and particle swarm optimization (PSO). This comparison provides novel insights into the robustness and stability of different optimization paradigms when applied under multimodal agricultural monitoring conditions, where image data and environmental sensor inputs must be jointly learned. The inclusion of PSO addresses a key limitation in the literature, where few studies have systematically examined the suitability of population-based methods for multimodal agricultural monitoring tasks.

Collectively, these contributions not only define a clear research gap but also offer practical solutions that expand the scope of CNN optimization for precision agriculture. The methods and findings presented here serve as a foundation for future research on robust, scalable, and generalizable multimodal deep learning models in agriculture.

4.2 Training Efficiency and Convergence Behavior

A comparative analysis was conducted between the baseline CNN and models optimized using Bayesian optimization and particle swarm optimization (PSO) to examine how hyperparameter optimization affects training efficiency and convergence behavior. The comparison emphasized the metric improvement speed, learning stability, and final validation performance without extending the training duration. This focus aligns with the recent literature on hyperparameter optimization, which highlights that convergence stability and training cost should be evaluated alongside accuracy (Raiaan et al., 2024; Shar et al., 2024).

The baseline CNN trained with default hyperparameters achieved a final validation accuracy of 0.7683 and an F1-score of 0.7301. However, the learning progression was slower and less stable. As shown in the training curve (Figure 10), the training and validation loss decreased gradually while the accuracy and F1 improved incrementally, indicating the potential for faster convergence through systematic hyperparameter search (Guo et al., 2020; Yeh et al., 2023).

The model converged more rapidly and reliably with Bayesian optimization, reaching a higher validation accuracy of 0.8227 and an F1-score of 0.7849. The optimizer selected a more effective combination of learning rate, dropout, and fusion dimension, resulting in a steeper increase in accuracy and a smoother decline in loss (Figure 12). This pattern is consistent with evidence that Bayesian optimization can reduce trial-and-error tuning and improve early-phase training efficiency in deep models (Restrepo-Arias et al., 2022; Shar et al., 2024).

The PSO-optimized CNN achieved the strongest overall gains, with a validation accuracy of 0.9059 and an F1-score of 0.8675. Compared with both the baseline and Bayesian models, PSO produced smoother convergence and lower overfitting risk, as reflected by the more consistent separation between training and validation loss (Figure 14). Similar stability improvements from swarm/metaheuristic tuning have been reported across CNN optimization studies, where PSO variants help search complex hyperparameter spaces effectively (Aguerchi et al., 2024; Ankalaki & Thippeswamy, 2024; Hong et al., 2023).

Collectively, these findings confirm that both the Bayesian and PSO-based frameworks support faster, smoother, and more reliable training. PSO showed the strongest benefit for convergence stability and generalization-oriented tuning, which is especially relevant for practical UAV/agriculture deployment where robustness under variability is a key requirement (Chaudhari & Polepaka, 2022; Zhu et al., 2024).

4.3 Robustness and Generalization Performance

All experiments were conducted on an augmented dataset incorporating image transformations (e.g., geometric distortion, color jittering, Gaussian blur, grayscale perturbation) and sensor standardization to evaluate robustness and generalization. This setup reflects realistic UAV monitoring conditions where changes in illumination, motion blur, and sensor noise can occur. Therefore, robustness is an essential criterion beyond accuracy alone (Raiaan et al., 2024; Zhu et al., 2024).

The optimized CNN models exhibited stronger generalizability than the baseline CNN and traditional classifiers. In particular, the PSO-optimized CNN achieved a validation accuracy of 90.59% and an F1-score of 0.8675, outperforming the Bayesian-optimized CNN ($F1 = 0.7849$) and the baseline ($F1 = 0.7301$). Confusion-matrix analysis further showed a lower false-negative rate ($FN = 35$) for PSO compared with the baseline CNN ($FN = 48$), SVM ($FN = 102$), and random forest ($FN = 104$), indicating improved sensitivity for operational decision-making in precision agriculture contexts (Jain et al., 2025; Zhu et al., 2024).

Traditional classifiers, such as SVM ($F1 = 0.781$) and random forest ($F1 = 0.771$), demonstrated weaker generalization under high-dimensional multimodal input and noisy visual data. This supports previous findings that deep multimodal learning can capture richer joint representations than shallow models that rely on fixed feature pipelines, particularly under heterogeneous sensing conditions (Chaudhari & Polepaka, 2022; Restrepo-Arias et al., 2022).

Overall, the results indicate that hyperparameter optimization improves not only accuracy but also robustness under perturbation, with PSO providing the most stable validation behavior under the same training budget (Ankalaki & Thippeswamy, 2024; Hong et al., 2023).

4.4 Comparison with the Traditional Methods

Although a growing number of studies have applied deep learning models to plant disease detection, most differ substantially from the present work in terms of data modality, input structure, and evaluation protocol. Many existing approaches rely only on unimodal image inputs—typically RGB, hyperspectral, or thermal imagery—without incorporating auxiliary environmental variables such as temperature, humidity, or nutrient measurements. This unimodal setting limits their applicability under multimodal agricultural monitoring conditions, where environmental variability and sensor dynamics are integral to real-world UAV-based crop assessment. In contrast, this study evaluates CNN optimization under multimodal agricultural monitoring conditions using paired image and sensor inputs, together with controlled perturbations that better reflect deployment-relevant variability.

Quantitatively, the PSO-optimized CNN achieved the highest validation accuracy (90.59%) and F1-score (0.8675), compared to SVM (84.76% and 0.7808), and RF (83.98% and 0.7711). The gains were most evident in the lower false negative rates, supporting the value of optimized end-to-end multimodal learning for robust monitoring and decision support in UAV-enabled precision agriculture (Jain et al., 2025; Zhu et al., 2024).

Moreover, traditional models are limited in their ability to directly handle raw multimodal inputs. They require a decoupled feature engineering process, often involving manual design or pretrained encoders that may not be optimized jointly with downstream classifiers. In contrast, CNN-based models support end-to-end learning, enabling the automatic integration of heterogeneous data streams (image and sensor) and the deeper exploitation of cross-modal interactions. This architectural advantage contributes to their superior performance, especially in capturing subtle patterns and complex decision boundaries inherent in agricultural datasets.

Notably, the optimized CNN maintained robust performance without explicit feature selection or reduction, highlighting its capacity to effectively handle high-dimensional multimodal data. The consistent gains in F1-score across different optimization strategies also indicate an improved balance between precision and recall, an essential trait in real-world deployment where both false positives and false negatives incur operational costs.

In summary, the experimental results confirm that hyperparameter-optimized CNNs outperform traditional machine learning methods not only in terms of accuracy but also in terms of robustness, generalization, and practicality for UAV-based crop health assessment tasks. These findings support the argument that deep multimodal learning frameworks are more suitable for handling the complexity and heterogeneity of agricultural monitoring environments.

4.5 Comparison with Related Studies

Although a growing number of studies have applied deep learning models to plant disease detection, the majority of these studies substantially diverge from the present study in terms of data modality, input structure, or evaluation protocol. Specifically, many existing approaches rely solely on unimodal image inputs—typically RGB, hyperspectral, or thermal imagery—without integrating auxiliary environmental features such as temperature, humidity, or light intensity. This unimodal setting limits the ecological validity of these models in real-world agricultural monitoring scenarios, where environmental noise and sensor dynamics are inherent.

Moreover, several previous studies lack standardized performance reporting. Most models only present overall accuracy, which fails to capture class-level misclassification patterns or model robustness under class imbalance. Key evaluation metrics, such as precision, recall, and F1-score, which are essential for fair and interpretable comparisons in multi-class classification tasks, are frequently omitted. This inconsistency poses a challenge for comprehensive benchmarking.

To address this, a selective comparison is conducted with (Mohiuddin & Yousuf, 2025), which not only employs a CNN architecture tailored for multimodal image inputs (RGB, thermal, hyperspectral), but also reports detailed per-class classification metrics, including F1-score and recall. This makes it a suitable and representative baseline for evaluating the performance of the proposed optimized multimodal model in a fair and quantifiable manner.

Table 3 summarizes a comparative evaluation of the proposed model and the approach proposed by (Mohiuddin & Yousuf, 2025) across five key dimensions derived from the research objectives. This focused comparison highlights how the integration of multimodal sensor data and the application of advanced optimization techniques lead to improved robustness, training efficiency, and consistency of performance.

Table 3 Comparison with related studies

Dimension	This study	(Mohiuddin & Yousuf, 2025)
1. Application Scenario	Simulates UAV conditions with noise injection using multimodal data (image + sensor)	No UAV simulation; only image-based modalities (RGB, thermal, and hyperspectral)
2. Optimization Strategy	Advanced optimization (Bayesian and PSO) is used to tune hyperparameters	No optimization applied; fixed settings and Adam optimizer are used
3. Robustness and Generalization	Trained under noise and misalignment; high F1 (0.8675) and reduced false negatives	Only 0.56 recall, no robustness enhancement
4. Training Efficiency	Fast and stable convergence with PSO and improved loss and learning curves	No convergence analysis; plateaued validation accuracy with instability
5. Performance Metrics	F1 score: 0.8675, accuracy: 90.59%, balanced across classes	F1 performance is poor despite 90% accuracy

The proposed model demonstrates superior performance across all evaluated dimensions (Table 1). These results validate the effectiveness of multimodal integration and hyperparameter optimization, particularly in improving model generalizability and stability under multimodal agricultural monitoring conditions that include noise, variability, and imperfect cross-modal alignment.

5. Conclusion

This study evaluated how hyperparameter optimization improves the performance of CNN for UAV-based crop health monitoring using a multimodal dataset that combines leaf images with environmental sensor data. A baseline CNN was compared with Bayesian optimization (Optuna), Particle Swarm Optimization (PSO), and traditional classifiers, including SVM and random forest. Overall, the optimized CNNs outperformed the baseline and conventional models in both accuracy and F1-score, with the PSO-tuned model achieving the best overall performance (validation accuracy 90.59% and F1-score 0.8675), mainly due to more effective tuning of learning rate, dropout, fusion dimensionality, batch size, and optimizer selection. The optimized models also demonstrated smoother convergence, lower loss, and reduced false-negative rates under augmented and noise-injected inputs, indicating stronger robustness under multimodal agricultural monitoring conditions relevant to simulated UAV deployment. In practice, the 100-

trial optimization budget would require approximately 612 hours of offline computation, which is a one-time cost that is operationally independent of the on-edge UAV deployment, at which only the resulting finalized model weights are loaded to make the inference of the resulting operational cost. Overall, the proposed framework provides an extensible and effective approach for improving multimodal CNN-based crop health classification. Future work may extend the model toward multiclass diagnosis, explainable AI integration, and broader geospatial context for deployable smart farming systems.

Acknowledgements

The authors gratefully acknowledge Asia Pacific University of Technology & Innovation (APU), Malaysia, for providing journal publication funding for this research output. This research was also supported by the Early Career Researcher Grant (BKP ECRG) from Universiti Malaya, under grant number BKP011 2025 ECRG.

Author Contributions

Jiaju Wang was responsible for the collection of data, the execution of the experiment as well as the initial testing. Uzair Iqbal was involved with the idea of the research, setting up the methodology, overseeing the project, and reviewing as well as editing the paper. Nurzati Iwani Othman was involved in the development of the study concept, performing formal analysis, validating the results, and making critical revisions to the manuscript. Hamza Ibrahim was responsible for validation and evaluating the experiments, and visualizing the results. All authors have read, revised, and approved the final manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

Supplementary Materials

Declaration of AI

The authors confirm that artificial intelligence tools were not used for the investigation or preparation of the text.

References

Abdullahi, H. S., & Sheriff, R. E. (2022). Introduction to deep learning in precision agriculture: Farm image feature detection using unmanned aerial vehicles through classification and optimization process of machine learning with convolution neural network. In *Deep learning for sustainable agriculture* (pp. 81–107). Elsevier. https://doi.org/10.1016/B978-0-323-85214-2.00013-6

Aguerchi, K., Jabrane, Y., Habba, M., & El Hassani, A. H. (2024). A cnn hyperparameters optimization based on particle swarm optimization for mammography breast cancer classification. *Journal of Imaging*, *10*(2), 30. https://doi.org/10.3390/jimaging10020030

Ahmed, M. A., Aloufi, J., & Alnatheer, S. (2023). Satin bowerbird optimization with convolutional lstm for food crop classification on uav imagery. *IEEE Access*, *11*, 41075–41083. https://doi.org/10.1109/ACCESS.2023.3269806

Ahuja, S., & Zaheer, S. (2024). Advancements in pathology: Digital transformation, precision medicine, and beyond. *Journal of Pathology Informatics*, *16*, 100408. https://doi.org/10.1016/j.jpi.2024.100408

Ajayi, O. G., Ibrahim, P. O., & Adegboyega, O. S. (2024). Effect of hyperparameter tuning on the performance of yolov8 for multi crop classification on uav images. *Applied Sciences*, *14*(13), 5708. https://doi.org/10.3390/app14135708

Alhammad, S. M., Khafaga, D. S., Elhady, W. M., Samy, F. M., & Hosny, K. M. (2025). Deep learning and explainable ai for classification of potato leaf diseases. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1449329

Ali, F., Razzaq, A., Tariq, W., Hameed, A., Rehman, A., Razzaq, K., Sarfraz, S., Rajput, N. A., Zaki, H. E. M., Shahid, M. S., & Ondrašek, G. (2024). Spectral intelligence: Ai-driven hyperspectral imaging for agricultural and ecosystem applications. *Agronomy*, *14*(10), 2260. https://doi.org/10.3390/agronomy14102260

Anand, A. (2025). Leaf disease detection system through advanced ai and machine learning integration. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5018301

Ankalaki, S., & Thippeswamy, M. N. (2024). Optimized convolutional neural network using hierarchical particle swarm optimization for sensor based human activity recognition. *SN Computer Science*. https://doi.org/10.1007/s42979-024-02794-5

Chaudhari, S. V., & Polepaka, S. (2022). Bayesian optimization with deep learning based crop type classification on uav imagery. *ICAISS 2022*. https://doi.org/10.1109/ICAISS55157.2022.10010961

Corceiro, A., Alibabaei, K., Assunção, E., Gaspar, P. D., & Pereira, N. (2023). Methods for detecting and classifying weeds, diseases and fruits using ai to improve the sustainability of agricultural crops: A review. *Processes*, *11*(4), 1263. https://doi.org/10.3390/pr11041263

El Sakka, M., Ivanovici, M., Chaari, L., & Mothe, J. (2025). A review of cnn applications in smart agriculture using multimodal data. *Sensors*, *25*(2), 472. https://doi.org/10.3390/s25020472

El-Kenawy, E.-S. M., Khodadadi, N., Mirjalili, S., Makarovskikh, T., Abotaleb, M., Karim, F. K., Alkahtani, H. K., Abdelhamid, A. A., Eid, M. M., & Horiuchi, T. (2022). Metaheuristic optimization for improving weed detection in wheat images captured by drones. *Mathematics*, *10*(23), 4421. https://doi.org/10.3390/math10234421

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors*, *17*(9), 2022. https://doi.org/10.3390/s17092022

Gülmez, B. (2024). A comprehensive review of convolutional neural networks based disease detection strategies in potato agriculture. *Potato Research*, 1–35. https://doi.org/10.1007/s11540-024-09786-1

Guo, Y., Xie, W., Zuo, Z., & Li, Q. (2020). A distributed particle swarm optimization approach for efficient hyperparameter optimization in convolutional neural networks. https://doi.org/10.1080/01969722.2020.1827797

Hong, T.-Y., Chiang, P.-Y., & Chen, Y.-K. (2023). A novel opposite-based particle swarm optimization for hyperparameter tuning in convolutional neural network: Application to pcb defect classification. *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2023.110904

Ibrahim, M. Q., Hussein, N. K., & Guinovart, D. (2025). Optimizing convolutional neural networks: A comprehensive review of hyperparameter tuning through metaheuristic algorithms. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-025-10292-x

Jain, A., Dubey, A. K., & Pan, V. S. H. (2025). Bayesian optimized cnn ensemble for efficient potato blight detection using fuzzy image enhancement. *Scientific Reports*, *15*, 31259. https://doi.org/10.1038/s41598-025-15940-7

Maghdid, S., Askar, S., Khoshaba, F. S., & Hamad, S. (2024). Deep learning algorithms for iot based crop yield optimization. *The Indonesian Journal of Computer Science*, *13*(2). https://doi.org/10.33022/ijcs.v13i2.3846

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, *7*, 1419. https://doi.org/10.3389/fpls.2016.01419

Mohiuddin, M. G., & Yousuf, M. (2025). Multi-modal plant disease detection using a single-stream cnn. https://doi.org/10.36037/ijrei.2025.9201

Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of machine learning approaches for precision farming in smart agriculture system: A comprehensive review. *IEEE Access*, *12*, 60155–60184. https://doi.org/10.1109/ACCESS.2024.3390581

Mythili, K., & Rangaraj, R. (2021). Deep learning with particle swarm based hyper parameter tuning based crop recommendation for better crop yield for precision agriculture. *Indian Journal of Science and Technology*, *14*(17), 1325–1337. https://doi.org/10.17485/IJST/v14i17.450

Naulia, P. S., Watada, J., & Aziz, I. A. (2024). A mathematically inspired meta-heuristic approach to parameter (weight) optimization of deep convolution neural network. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3409689

Ngugi, H. N., Ezugwu, A. E., Akinyelu, A. A., & Abualigah, L. (2024). Revolutionizing crop disease detection with computational deep learning: A comprehensive review. *Environmental Monitoring and Assessment*, *196*, 302. https://doi.org/10.1007/s10661-024-12454-z

PP, F. R., Ismail, W. N., & Ali, M. A. (2023). A metaheuristic harris hawks optimization algorithm for weed detection using drone images. *Applied Sciences*, *13*(12), 7083. https://doi.org/10.3390/app13127083

Raiaan, M. A. K., Sakib, S., Fahad, N. M., Al Mamun, A., Rahman, M. A., Shatabda, S., & Mukta, M. S. H. (2024). A systematic review of hyperparameter optimization techniques in convolutional neural networks. *Discover Artificial Intelligence*. https://doi.org/10.1016/j.dajour.2024.100470

Restrepo-Arias, J. F., Branch-Bedoya, J. W., & Awad, G. (2022). Plant disease detection strategy based on image texture and bayesian optimization with small neural networks. *Agriculture*, *12*(11), 1964. https://doi.org/10.3390/agriculture12111964

Shar, L. K., Duong, T. N. B., Yeo, Y. C., & Fan, J. (2024). Empirical evaluation of hyperparameter optimization techniques for deep learning-based malware detectors. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2024.09.640

Taji, K., Sohail, A., Shahzad, T., Khan, B. S., Khan, M. A., & Ouahada, K. (2024). An ensemble hybrid framework: A comparative analysis of metaheuristic algorithms for ensemble hybrid cnn features for plants disease classification. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3389648

Teixeira, K., Miguel, G., Silva, H. S., & Madeiro, F. (2023). A survey on applications of unmanned aerial vehicles using machine learning. *IEEE Access*, *11*, 117582–117621. https://doi.org/10.1109/ACCESS.2023.3326101

Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. *Computational Intelligence and Neuroscience*, *2017*, 2917536. https://doi.org/10.1155/2017/2917536

Yeh, W.-C., Lin, Y.-P., & Lin, Y. (2023). Simplified swarm optimization for hyperparameters of convolutional neural networks. *Computers & Industrial Engineering*. https://doi.org/10.1016/j.cie.2023.109076

Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., Xu, J.-L., & He, Y. (2024). Intelligent agriculture: Deep learning in uav-based remote sensing imagery for crop diseases and pests detection. *Frontiers in Plant Science*, *15*. https://doi.org/10.3389/fpls.2024.1435016

Zualkernan, I., Abuhani, D. A., Hussain, M. H., Khan, J., & ElMohandes, M. (2023). Machine learning for precision agriculture using imagery from unmanned aerial vehicles (uavs): A survey. *Drones*, *7*(6), 382. https://doi.org/10.3390/drones7060382