EG UK Computer Graphics & Visual Computing (2025)
Y. Sheng and A. Slingsby (Editors)

Non-Invasive Detection and Characterization of Powdery Mildew in Strawberries Using Hyperspectral Imaging and Deep Learning under Poly-tunnel Conditions

Jobin Francis[†1] Maximilian Pircher[1] Philipp Wree[1] and Ali Al Masri[1]

[1] Fraunhofer Institute for Computer Graphics and Research, Joachim-Jungius-Straße 11, 18059, Rostock, Germany

Abstract

Powdery Mildew (Podosphaera aphanis, PM) presents a serious challenge to strawberry cultivation, causing significant yield losses if not controlled early. Globally, PM contributes to substantial reductions in strawberry production, posing economic threats to growers and raising food security concerns. Therefore, the development of accurate, noninvasive detection mechanisms is crucial for effective disease management. This study investigates the application of Hyperspectral Imaging (HSI) for nondestructive PM detection and classification of healthy and PM-infected strawberry leaves grown under a poly-tunnel, an environment that closely simulates real-world agricultural conditions. A hyperspectral camera (350–1000 nm), mounted on a mechanized rail system beneath the poly-tunnel, was used to capture leaf images as it moved linearly above the strawberry canopy. Spectral preprocessing included Savitzky-Golay smoothing (SGS), Standard Normal Variate (SNV), and Multiplicative Scatter Correction (MSC), with the Isolation Forest algorithm applied for outlier removal. A one-dimensional Convolutional Neural Network (1D-CNN) was trained to classify healthy versus PM-infected leaves, achieving 75% accuracy and 84% precision. The model outperformed traditional classifiers such as Random Forest (RF), Decision Tree (DT), and Partial Least Squares Discriminant Analysis (PLS-DA). Among all tested pipelines, the SGS+MSC+1D-CNN combination yielded the highest performance. This study highlights the feasibility and effectiveness of integrating HSI with deep learning for robust disease detection under semi-controlled conditions, laying the groundwork for scalable, real-time plant health monitoring in precision agriculture.

CCS Concepts

• ***Computing methodologies*** → *Artificial Intelligence;* • ***Image and Video Acquistion*** → *Hyperspectral Imaging;*

1. Introduction

Strawberries are a widely grown fruit that is well-liked by consumers worldwide due to their flavor and high nutritional value, which includes vitamin C, folate, and phenolic compounds [WFJ*23], [CSL*24]. Furthermore, with about 9 million tonnes produced globally, strawberries represent a significant economic commodity. However, a variety of fungal diseases infects the strawberry cultivation further, resulting in 20 to 30% of the total production [WFJ*23]. Powdery Mildew (PM), due to its ease of infection, is a type of disease that poses a serious threat to strawberry production lines [SCH*21]. PM attacks leaves, flowers, and fruits one after the other, which, if the infection is serious, can result in significant crop loss [SCH*21].

Because of the significant impact of PM, effective methods and management are required to control infection spread in strawberry fields. The most common method for suppressing PM is to apply fungicides as a spray; however, PM is prone to developing fungicide resistance [LMS10]. On the other hand, excessive application may result in environmental degradation and the accumulation of toxic substances in soil, which is not ethically acceptable [KG11].

The traditional method of monitoring plant disease is to seek the assistance of an expert to detect the presence of diseases. However, this method is not cost-effective, requires a lot of labor and time, and accurate predictions are prone to errors [SCH*21]. Another common practice is laboratory testing, such as plant leaflet grafting or electron microscopy, among others [ZOY*23]. These methods are more accurate than visual inspection, but they are also time-consuming, destructive, and necessitate precision instruments and rigorous operations [ZOY*23]. As a result, an accurate and non-destructive identification technology for detecting diseases at an early stage is critical and required for the strawberry production systems to prevent the spread of diseases and also to reduce the use of fungicides.

† Corresponding author: jobin.francis@igd-r-extern.fraunhofer.de

DOI: 10.2312/cgvc.20251204

Even though machine vision using conventional RGB is fast and non-invasive and capable of detecting visible lesions, it is incapable of detecting the physiological status at early infection stages, since changes in plant metabolism and tissue structure occur before visual symptoms appear [XLSS22]. This limitation highlights the potential of Hyperspectral Imaging (HSI), a relatively new and emerging technique that combines the strengths of both imaging and spectroscopy, which is able to provide continuous and narrow-band image information as well as spectral data information for each pixel [ZOY*23]. Because of the combination of spectral and imaging techniques, HSI has a high potential for early disease diagnosis, as it can determine biochemical and physiological responses such as leaf pigment compositions, water content, and so on [XLSS22].

The recent studies, when HSI is integrated with chemometric techniques, have demonstrated significant potential in plant disease detection across various crops. Zhang et al. [ZOY*23] explored the identification of gray mold and anthracnose in strawberry leaves by leveraging spectral fingerprint features alongside vegetation indices. Their study employed chemometric models such as Back Propagation Neural Networks (BPNN), Support Vector Machines (SVM), and Random Forests (RF). Similarly, Wu et al. [WFJ*23] proposed a gray mold detection framework for strawberries incorporating textural features, spectral data, and multiple vegetation indicators. Machine learning models used in this research included Extreme Learning Machine (ELM), K-Nearest Neighbors (KNN), and SVM. Xuan et al. [XLSS22] focused on the early detection of PM in wheat, utilizing Gray Level Co-occurrence Matrix (GLCM) for textural feature extraction in addition to spectral analysis, and applied Partial Least Squares Discriminant Analysis (PLS-DA) for classification. In another relevant study, Jung et al. [JKK*22] developed a hybrid HSI and 3D Convolutional Neural Network (3D-CNN) framework for classifying gray mold in strawberry leaves, using training data spanning infected, asymptomatic, and healthy tissue. These studies, which adopt similar methodological frameworks, provide foundational support for the current investigation. Moreover, comprehensive reviews such as [RML*23], [NBBH23] on HSI-based plant disease detection offer valuable insights and data that further contextualize this work.

However, most of the aforementioned studies have been conducted under controlled laboratory conditions. The primary advantage of such an environment is the stability it offers. To put it another way, the factors such as illumination, sample orientation (e.g., leaf positioning), and ambient conditions remain consistent throughout data acquisition. This controlled setup simplifies the hyperspectral imaging process and helps maintain calibration accuracy. In contrast, real-world or field conditions introduce a variety of challenges, including fluctuations in illumination, environmental factors such as wind and temperature, and potential inconsistencies in sample orientation. These variables can significantly affect the quality and reliability of the hyperspectral data, potentially altering calibration and impacting model performance. Given that these plant diseases typically occur in open-field settings, it is crucial to investigate detection methods under such realistic conditions. Therefore, one of the main objectives of this study is to evaluate the performance of a hyperspectral imaging-based disease detection model in real-time, field-like environments. Specifically, this research involves acquiring hyperspectral data from strawberry plants cultivated under a poly-tunnel, as illustrated in Figure 1. The main objectives of this study are as follows:

- To acquire hyperspectral data from strawberry plants in a semi-controlled, real-world environment (poly-tunnel).
- To extract and analyze spectral features of healthy and PM-affected strawberry leaves for disease detection.
- To develop a classifier model for effective discrimination between healthy and diseased leaves.
- To evaluate the model's performance and compare it with state-of-the-art classifiers under near-real-time conditions.
- To assess the robustness and practical applicability of HSI-based disease detection beyond laboratory settings.

2. Materials and Methods

2.1. Plant cultivation

The strawberry leaf samples used in this experiment were acquired from a strawberry field managed by Fraunhofer IGD, Rostock, Germany. The strawberry plants were grown on a few racks of long meters beneath a poly-tunnel. In this experiment, the plants were cultivated without the PM inoculation process, and powdery mildew naturally developed on the strawberry leaves, becoming severe over the course of several weeks. This is an investigation into the typical symptoms of powdery mildew that impact plants. The severity at various phases was then tracked and recorded. As depicted in Figure 1, the camera, mounted on a semi-mobile trolley that travels along overhead rails in a poly-tunnel, captures images at regular intervals with centimeter-precise localization in the visible and near-infrared ranges at hyperspectral resolutions.

Figure 1: *Strawberry plants cultivated under a poly-tunnel and imaged using a hyperspectral camera mounted on a moving rail mechanism*

2.2. Hyperspectral Image Acquisition and Processing

2.2.1. Hyperspectral imaging system

A Cubert X20 was employed for the image acquisition of the strawberry leaves, which ranged in wavelength from 350 to 1000 nm. The Cubert X20 is an Ultra-Violet Visible-Near Infrared (UV Vis-NIR) hyperspectral video camera, and its detailed specifications

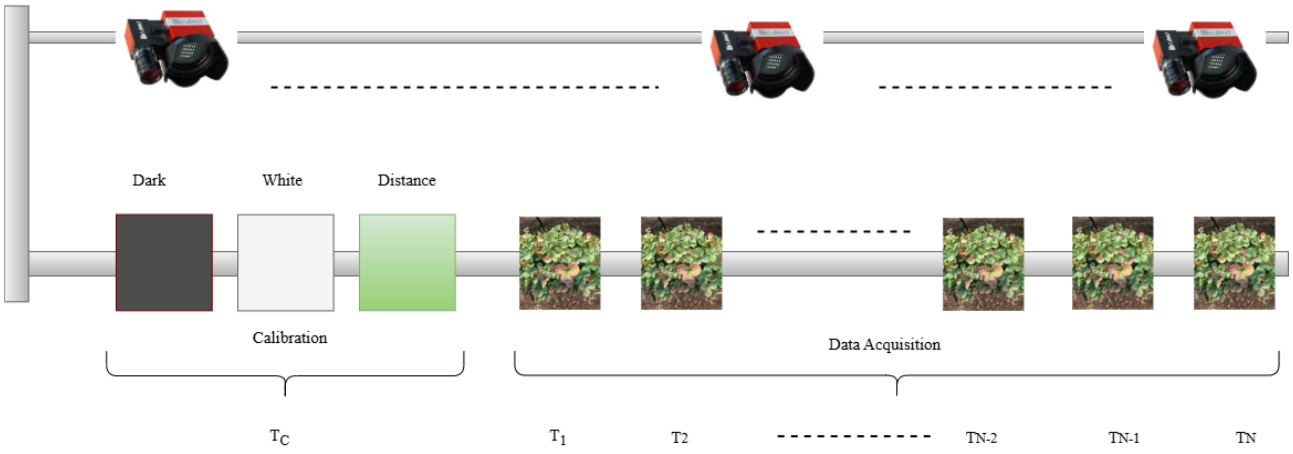


Figure 2: *Schematic representation of the protocol used for calibration and data acquisition employed in this study*

Table 1: *Specifications of Cubert X20 camera*

No	Parameters	Value	No	Parameters	Value
1	Spectral range (nm)	350-1000	5	Field of view	35°
2	Spatial pixels	410 × 410	6	Max frame rate	4 Hz
3	No. of bands	164	7	Data depth	12 bit
4	Spectral sampling (nm)	4	8	Bit depth	16

are given in Table 1. The Cubert X20 was then calibrated prior to data acquisition for the data samples (strawberry leaves). The dark, white, and distance calibrations were done sequentially and recorded. The dark calibration of the equipment was accomplished by covering the lens with an opaque cap [XLSS22]. The white reference image, on the other hand, is obtained using a rectangular Spectralon tile with 99.9% reflectance and the lens that has been opened after dark calibration. Following that, distance calibration was done by determining and fixing the distance between the camera and the object. The camera calibration is a required and inevitable step because the collected HSI images require radiometric correction before being used for investigations [JWT*21].

Hence, to determine the hyperspectral reflectivity of the image, the following equation needs to be applied as [CSL*24],

$$I_c = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \tag{1}$$

where I_{C} represents the calibrated image, I_{raw} represents the raw image, I_{dark} represents the dark reference image, and I_{white} represents the white reference image.

2.2.2. Calibration Protocol

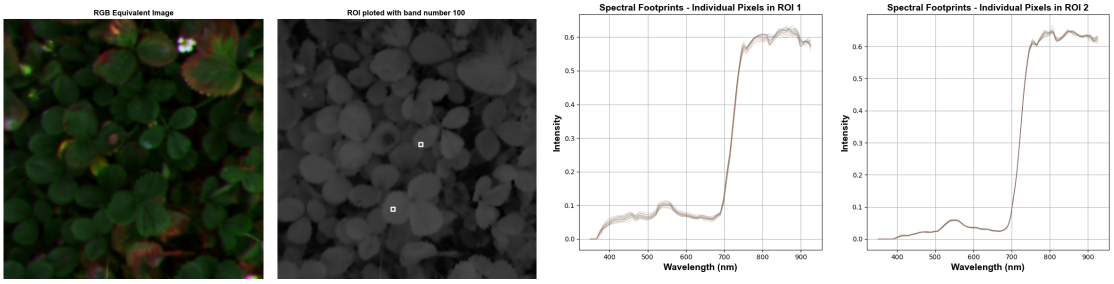
Given that the primary objective of this experiment was to detect the PM disease and quantify its detrimental effects on strawberry leaves, a dedicated acquisition protocol was designed and employed. Prior to image collection, a strict calibration procedure was performed using white and dark reference standards, as well as distance calibration, to ensure consistent and repeatable spectral measurements. An overview of the entire procedure is illustrated in Figure 2.

As shown in Figure 2, during acquisition, the HSI camera was mounted on a mechanized rail system mounted above the strawberry plants, allowing the camera to traverse the length of the polytunnel in a continuous and uniform manner. However, once the camera began moving along this rail, slight fluctuations in the calibration setup were observed due to variations in ambient lighting, leaf orientation, and mechanical vibrations inherent to the camera's motion. Then, let T_C represents the calibration period prior to imaging. Following this, hyperspectral data were sequentially captured at time points T_1, T_2, T_3, and so forth, as the camera moved along the rail. Let T_N denotes the time at which spectral deviations from the calibrated conditions exceed an acceptable threshold. This threshold was determined by comparing the spectral signatures measured at each time point to a reference spectrum of healthy, uninfected leaves. Once such a deviation was detected (e.g., at T_N), all subsequent data were deemed unreliable for accurate analysis.

Accordingly, all images acquired after the T_{N+1} frame were discarded, ensuring that only spectrally consistent and well-calibrated data were used for further processing and modeling. This strict protocol guaranteed that the modeling was based exclusively on data collected under stable calibration conditions, improving the robustness and accuracy of the PM disease detection system.

2.2.3. Spectra extraction

The spectral data extraction was carried out using the Cuvis SDK provided by Cubert, which offers wrappers for Python, MATLAB, and C++. For this study, the Python interface of the Cuvis SDK was employed, along with relevant Python libraries for data processing. As shown in *Eqn* (1), the calibrated hyperspectral reflectance data were computed using measurements obtained from dark, white, and distance calibration frames. To ensure consistency, the reflectance values were normalized to fall within the $[0, 1]$ range. Given that the hyperspectral camera operates with a 12-bit depth, normaliza-



(a) RGB equaivalent image of a single HSI data (b) ROI selection for heathy and PM affected leaf areas (c) Corresponding spectra for ROI 1 (PM affected) (d) Corresponding spectra for ROI 2 (Healthy)

Figure 3: *Figures illustrate the various steps involved in spectra extraction of the strawberry leaves*

tion was performed by dividing the raw pixel values by 4095, the maximum possible intensity value.

Further, regions of interest (ROI) corresponding to healthy and infected areas were manually identified. From each hyperspectral data cube, the spectral bands corresponding to red, green, and blue wavelengths were concatenated to create an RGB-equivalent image, facilitating visual inspection and ROI selection. Based on these RGB composites, the specific regions representing each class were delineated. Subsequently, the spectral data from the selected ROIs were extracted and stored separately for further analysis. The entire procedure of extracting spectra is illustrated in Figure 3. Figure 3 (a) is the RGB-equivalent image created using the aforementioned procedure, in which the PM-affected area is clearly visible. Then, Figure 3 (b) is the grayscale image captured at 794 nm (100^{th} band), where ROIs taken from the healthy leaf area as well as the PM-affected area are displayed. Finally, Figure 3 (c) and Figure 3 (d) represent the mean spectra plotted against the wavelengths for these two ROIs, where ROI 1 denotes the mean spectra from the PM-affected leaf area, whereas ROI 2 represents the mean spectra taken from the healthy leaf area.

2.2.4. Hyperspectral data preprocessing

To reduce the influence of noise and light scattering in the acquired spectral data, a structured preprocessing pipeline was implemented. The process begins with the application of the Isolation Forest algorithm [LTZ08], an unsupervised outlier detection method that isolates anomalous observations based on random partitioning. This step targets and removes spectral outliers potentially caused by noisy measurements, variable illumination, or other inconsistencies not representative of healthy or diseased leaf samples.

Following outlier detection, Savitzky–Golay Smoothing (SGS) [WLM*23] was applied to reduce high-frequency noise while preserving the original shape and features of the spectra. Subsequently, two preprocessing combinations were evaluated for data modeling:

- SGS followed by Standard Normal Variate (SGS+SNV)
- SGS followed by Multiplicative Scatter Correction (SGS+MSC)

Both SNV and MSC are designed to correct multiplicative and additive effects due to scattering, aiming to normalize each spectrum relative to the dataset mean.

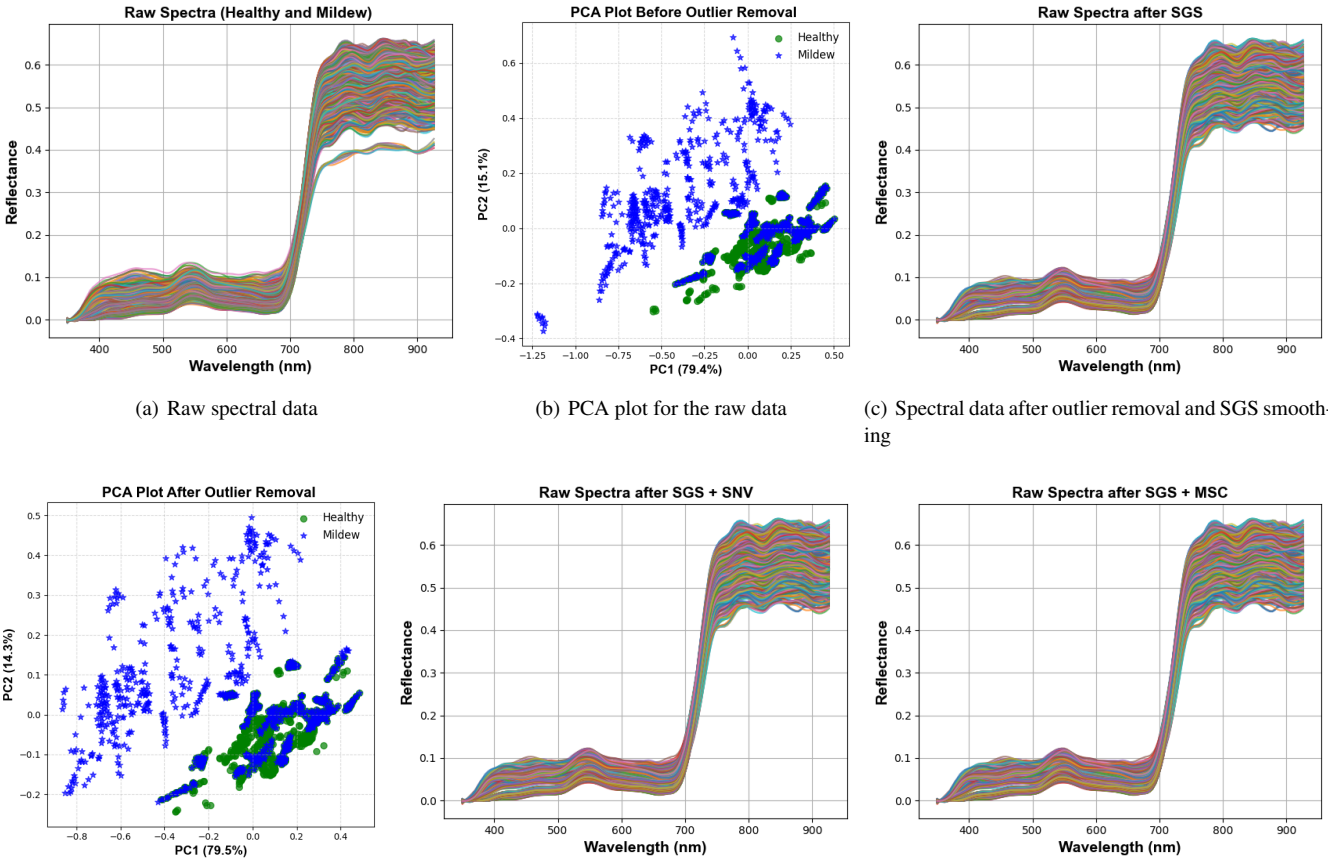
These preprocessing strategies were employed to enhance spectral data quality and improve the reliability of downstream classification models. Figure 4 (a) represents total spectral data acquired for this experiment, and Figure 4 (b) represents its corresponding Principal Components Analysis (PCA) [ASS21] plot. Then, Figure 4 (c) denotes the spectral data after outlier removal using Isolation Forest and SGS, and Figure 4 (d) represents its respective PCA plot. Similarly, Figures 4 (e) and 4 (f) are the spectral data plotted after the SGS+SNV and the SGS+MSC preprocessing steps, respectively.

2.3. Classification model development

Machine learning models combined with hyperspectral imaging have been widely adopted in plant disease diagnosis. Traditional machine learning techniques such as SVM, KNN, Naïve Bayes (NB), and PLS-DA have been commonly used to identify plant diseases and estimate their severity [JWT*21]. However, these methods rely heavily on manual feature extraction and tend to perform well only under specific conditions. At the same time, studies have increasingly incorporated deep learning techniques, which offer the advantage of automatic feature extraction and have demonstrated higher accuracy compared to traditional approaches [ASE23].

Nowadays, deep learning has become a preferred choice for disease identification, driven by advances in computational power, larger storage capacities, and the growing availability of large, labeled datasets. Deep learning with CNN has emerged as a prominent topic in plant disease identification research [ASE23]. CNNs are commonly used in deep learning to classify images, recognize objects, and segment semantically. CNN is an enhanced version of Artificial Neural Networks (ANNs) that is widely used for image and spectrum analysis. A CNN's convolutional layers extract features from the incoming data. The convolutional layers consist of a filter and an activation function that classifies images of leaves as healthy or disease-affected.

In this experiment, a one-dimensional CNN (1D-CNN) was developed to classify hyperspectral signatures of healthy and powdery mildew-affected strawberry leaves. The model consisted of an input layer accepting 1D spectral vectors of 145 bands, followed by two convolutional layers with 64 and 128 filters, respectively, each



(a) Raw spectral data

(b) PCA plot for the raw data

(c) Spectral data after outlier removal and SGS smoothing

(d) PCA plot after after outlier removal and SGS smoothing

(e) SGS+SNV out

(f) SGS+MSC out

Figure 4: *Figures illustrate raw spectral data, outlier detection, SGS, SNV and MSC results*

using a kernel size of 3 and ReLU activation. Batch normalization layers were applied after each convolutional layer to stabilize and accelerate training. Max-pooling layers with a pool size of 2 were used to progressively reduce feature dimensions. The extracted features were flattened and passed through a fully connected (dense) layer with 64 neurons and ReLU activation, followed by a dropout layer (rate = 0.3) to prevent overfitting. Finally, a sigmoid-activated output neuron was used for binary classification. The model was trained using the Adam optimizer and binary cross-entropy loss.

3. Results and Discussions

3.1. Spectral Analysis

In this experiment, the full hyperspectral wavelength range was utilized to analyze the spectral behavior of strawberry leaves. Spectral absorption varies with molecular structure; thus, infection-induced changes in cellular composition and molecular structure lead to noticeable differences in the spectral patterns of healthy versus PM-infected leaves [JWT*21]. These spectral variations are often consistent and distinct—serving as a form of chemical fingerprint—which allows not only the detection of disease occurrence but also the assessment of infection severity.

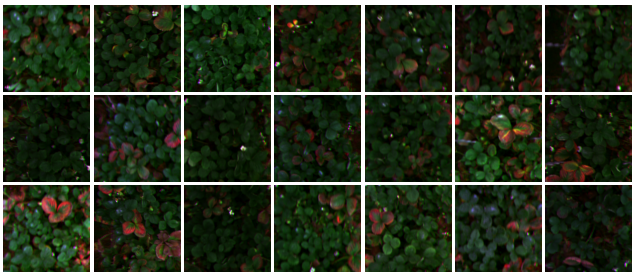


Figure 5: *Figure illustrates a few sample images of strawberries from acquired from the strawberry field*

Figure 5 presents sample images captured by the HSI camera during data acquisition. As shown, visible distinctions can be observed between healthy and PM-infected leaves. Figure 6 illustrates

the mean spectral signatures of healthy and PM-affected leaf areas, derived from the sample image in the first row and second column of Figure 5. As depicted in Figure 6, healthy leaves typically exhibit low reflectance in the 400–700 nm region of the spectrum. Within this range, a characteristic green peak appears between 500 and 570 nm, including a nitrogen absorption band near 550 nm [WFJ*23]. Additionally, a distinct reflectance valley is observed around 670 nm, attributed to strong chlorophyll absorption. This is followed by a sharp rise at the red edge and a generally high reflectance in the near-infrared (NIR) region.

In contrast, PM-infected leaf regions display significant spectral deviations due to altered photochemical activity and changes in pigment content associated with disease progression [XLSS22]. These differences are evident in the comparative mean spectra, clearly confirming the theoretical expectations. The spectral signatures of healthy and infected leaf areas thus highlight key features that enable reliable differentiation using hyperspectral imaging.

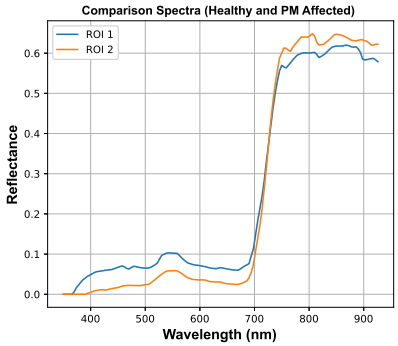


Figure 6: *Mean Spectra of healthy leaf and PM-infected leaf plotted against the wavelength*

3.2. Evaluation metrics

The models' performance in distinguishing PM-affected strawberry leaves was evaluated using the confusion matrix, precision, recall, F_1-score, and accuracy using the equations mentioned below [CSL*24].

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{F}_1\text{-score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

where , TP, TN, FP, FN represent the true positive, true negative, false positive, and false negative data samples, respectively

3.3. Performance evaluation results for the classification models

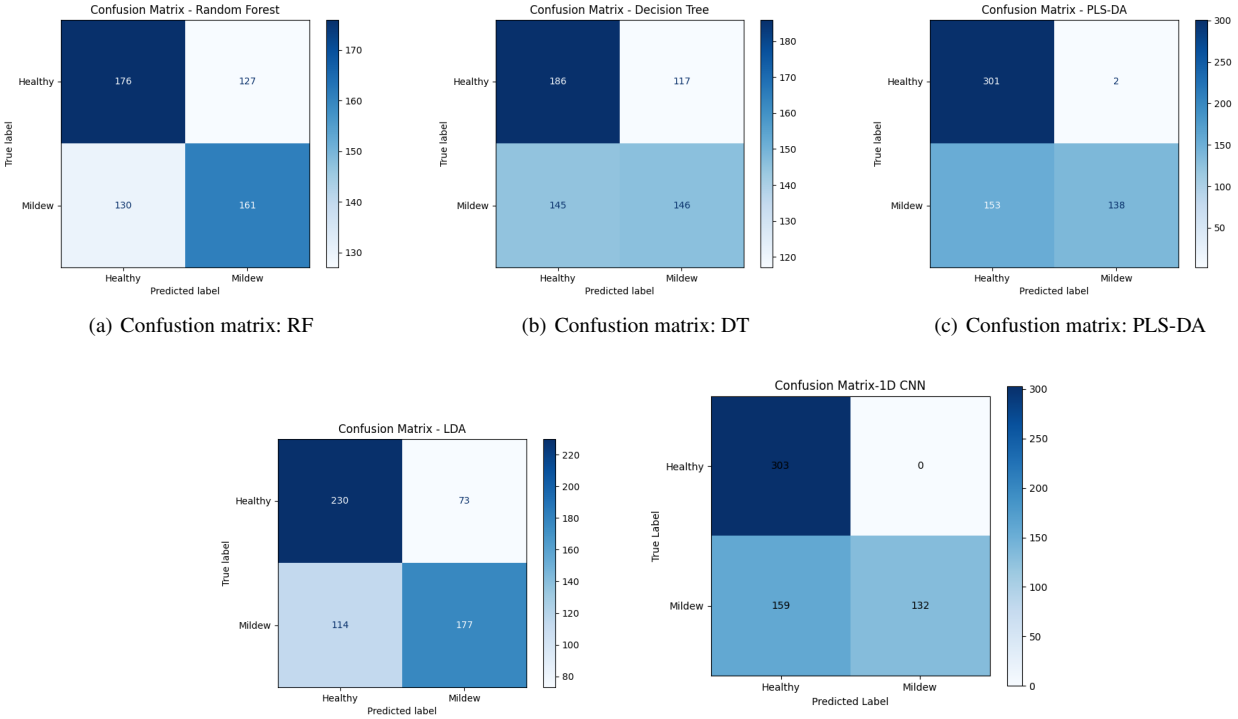
In this experiment, the accuracy of the 1D-CNN model converged within 30 epochs, demonstrating efficient training. As the number of training iterations increased, a steady decrease in training loss was observed, indicating effective learning. In addition to the 1D-CNN, traditional machine learning models such as RF, DT, LDA, and PLS-DA were employed for comparative analysis.

Table 2: *Performance evaluation results of the classifier models*

Model	Preprocessing	Precision	Recall	F_1-score	Accuracy
RF	SGS+SNV	0.56	0.54	0.55	0.57
	SGS+MSC	0.56	0.55	0.56	0.57
DT	SGS+SNV	0.54	0.49	0.52	0.55
	SGS+MSC	0.56	0.50	0.53	0.56
PLS-DA	SGS+SNV	0.75	0.72	0.64	0.74
	SGS+MSC	0.77	0.72	0.65	0.74
LDA	SGS+SNV	0.71	0.61	0.65	0.69
	SGS+MSC	0.71	0.61	0.65	0.69
1D-CNN	SGS+SNV	0.82	0.73	0.71	0.73
	SGS+MSC	**0.84**	0.73	0.71	**0.75**

The performance comparison of all models is summarized in Table 2, with the best results highlighted in bold. It was observed that the RF and DT models yielded relatively poor evaluation metrics and did not achieve satisfactory classification performance. In contrast, the PLS-DA model produced acceptable results, showing improvement over RF and DT across key performance indicators with the accuracy of 0.74, and the precision of 0.77. Among all evaluated models, the 1D-CNN employed delivered the accuracy and precision values as 0.74 and, 0.84 respectively. Hence, the 1D-CNN yielded the most consistent and superior performance, outperforming the traditional methods in terms of accuracy and other evaluation metrics, confirming its suitability for the classification task in this study. Furthermore, the confusion matrices for all the models obtained are shown in Figure 7 (a) to Figure 7 (e).

Even though the proposed framework yielded promising results compared to other machine learning models, there remains potential for further improvement. One contributing factor is that, while the 1D-CNN employed in this study effectively captured spectral features, it is inherently limited to spectral information and does not account for spatial patterns that could provide additional context for classification. Incorporating architectures capable of extracting both spectral and spatial features, such as 2D- or 3D-CNNs; could enhance the model's ability to localize disease symptoms more accurately, especially under heterogeneous field conditions. Furthermore, environmental variations during image acquisition—such as changes in lighting, leaf orientation, and background interference—can influence spectral signatures. Although standard preprocessing techniques help to minimize these effects, future systems may benefit from adaptive correction mechanisms or real-time calibration strategies to ensure the reliability and consistency of hyperspectral data in variable field environments.



(a) Confustion matrix: RF (b) Confustion matrix: DT (c) Confustion matrix: PLS-DA

(d) Confustion matrix: LDA (e) Confustion matrix: 1D-CNN

Figure 7: *Figures illustrate the various confusion matrices obtained for the classifier models*

4. Challenges and Future Perspectives

The integration of hyperspectral imaging (HSI) and deep learning for the detection of powdery mildew in strawberry leaves opens promising avenues in precision agriculture. However, several technical aspects encountered during this study present valuable opportunities for refinement and future exploration. One key area for future enhancement involves expanding the diversity and volume of hyperspectral data. Although the current dataset effectively supported model development, incorporating a broader range of samples across different growth stages, environmental conditions, and plant varieties could further improve model generalization and robustness. Such diversity not only enhances classification performance but also ensures adaptability to real-world agricultural settings where variability is the norm.

Another consideration lies in the high dimensionality of hyperspectral data, which includes hundreds of spectral bands. While these bands offer rich information, they also introduce spectral redundancy and potential noise. In this study, preprocessing techniques such as SGS, SNV, and MSC were employed to improve signal quality. Nevertheless, further optimization through advanced feature selection or dimensionality reduction methods or attention-based band selection may yield even more discriminative features and improve model efficiency. The detection of powdery mildew, particularly at early infection stages, presents a subtle challenge due to minimal visible and biochemical changes in the plant tissue. These changes often result in subtle spectral variations that are difficult to capture without highly sensitive models. Integrating complementary data modalities, such as RGB or thermal imaging, alongside hyperspectral data, may enhance the system's ability to identify early disease symptoms with greater precision.

In future work, efforts will be directed toward expanding the dataset to capture a broader spectrum of conditions, exploring multimodal sensing approaches, and employing more sophisticated model architectures that integrate spatial and spectral learning. Emphasis will also be placed on real-time applicability through model optimization and deployment strategies suitable for field environments. These developments aim to bridge the gap between experimental success and practical deployment, bringing the benefits of hyperspectral technology closer to everyday agricultural decision-making.

5. Conclusions

This study investigated the feasibility of hyperspectral imaging for detecting and classifying healthy and powdery mildew (PM)-affected leaves in plants grown under a poly-tunnel environment. The results demonstrated that changes in spectral reflectance patterns were strongly correlated with physiological alterations caused by PM infection. A one-dimensional Convolutional Neural Network (1D-CNN) was implemented as the primary classification model and achieved an accuracy of 75% on the test dataset over the

full spectral range. The performance of the 1D-CNN model was compared with several conventional classifiers, including Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), Random Forest (RF), and Decision Tree (DT). The 1D-CNN consistently outperformed these methods. Spectral preprocessing techniques such as Savitzky-Golay Smoothing (SGS), Standard Normal Variate (SNV), and Multiplicative Scatter Correction (MSC) were employed to improve data quality, while Isolation Forest was used for outlier detection. Among all tested configurations, the SGS + MSC + 1D-CNN model achieved the best classification performance. The findings highlight the potential of combining hyperspectral imaging with deep learning for early disease detection under protected cultivation systems. Future work will aim to improve model generalizability by incorporating a larger and more diverse dataset, including different stages of PM infection.

References

[ASE23] AHMAD A., SARASWAT D., EL GAMAL A.: A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology 3* (2023), 100083. doi:https://doi.org/10.1016/j.atech.2022.100083. 4

[ASS21] ANOWAR F., SADAOUI S., SELIM B.: Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review 40* (2021), 100378. doi:https://doi.org/10.1016/j.cosrev.2021.100378. 4

[CSL*24] CHUN S.-W., SONG D.-J., LEE K.-H., KIM M.-J., KIM M. S., KIM K.-S., MO C.: Deep learning algorithm development for early detection of botrytis cinerea infected strawberry fruit using hyperspectral fluorescence imaging. *Postharvest Biology and Technology 214* (2024), 112918. doi:https://doi.org/10.1016/j.postharvbio.2024.112918. 1, 3, 6

[JKK*22] JUNG D.-H., KIM J. D., KIM H.-Y., LEE T. S., KIM H. S., PARK S. H.: A hyperspectral data 3d convolutional neural network classification model for diagnosis of gray mold disease in strawberry leaves. *Frontiers in plant science 13* (2022), 837020. doi:https://doi.org/10.3389/fpls.2022.837020. 2

[JWT*21] JIANG Q., WU G., TIAN C., LI N., YANG H., BAI Y., ZHANG B.: Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Physics & Technology 118* (2021), 103898. doi:https://doi.org/10.1016/j.infrared.2021.103898. 3, 4, 5

[KG11] KALIA A., GOSAL S.: Effect of pesticide application on soil microorganisms. *Archives of Agronomy and Soil Science 57*, 6 (2011), 569–596. doi:10.1080/03650341003787582. 1

[LMS10] LEBEDA A., MCGRATH M. T., SEDLÁKOVÁ B.: Fungicide resistance in cucurbit powdery mildew fungi. *Fungicides 11*, 2 (2010), 221–246. 1

[LTZ08] LIU F. T., TING K. M., ZHOU Z.-H.: Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (2008), pp. 413–422. doi:10.1109/ICDM.2008.17. 4

[NBBH23] NOSHIRI N., BECK M. A., BIDINOSTI C. P., HENRY C. J.: A comprehensive review of 3d convolutional neural network-based classification techniques of diseased and defective crops using non-uav-based hyperspectral images. *Smart Agricultural Technology 5* (2023), 100316. doi:https://doi.org/10.1016/j.atech.2023.100316. 2

[RML*23] RAYHANA R., MA Z., LIU Z., XIAO G., RUAN Y., SANGHA J. S.: A review on plant disease detection using hyperspectral imaging. *IEEE Transactions on AgriFood Electronics 1*, 2 (2023), 108–134. doi:10.1109/TAFE.2023.3329849. 2

[SCH*21] SHIN J., CHANG Y. K., HEUNG B., NGUYEN-QUANG T., PRICE G. W., AL-MALLAHI A.: A deep learning approach for rgb image-based powdery mildew disease detection on strawberry leaves. *Computers and electronics in agriculture 183* (2021), 106042. doi:https://doi.org/10.1016/j.compag.2021.106042. 1

[WFJ*23] WU G., FANG Y., JIANG Q., CUI M., LI N., OU Y., DIAO Z., ZHANG B.: Early identification of strawberry leaves disease utilizing hyperspectral imaging combing with spectral features, multiple vegetation indices and textural features. *Computers and Electronics in Agriculture 204* (2023), 107553. doi:https://doi.org/10.1016/j.compag.2022.107553. 1, 2, 6

[WLM*23] WANG J., LIN T., MA S., JU J., WANG R., CHEN G., JIANG R., WANG Z.: The qualitative and quantitative analysis of industrial paraffin contamination levels in rice using spectral pretreatment combined with machine learning models. *Journal of Food Composition and Analysis 121* (2023), 105430. doi:https://doi.org/10.1016/j.jfca.2023.105430. 4

[XLSS22] XUAN G., LI Q., SHAO Y., SHI Y.: Early diagnosis and pathogenesis monitoring of wheat powdery mildew caused by blumeria graminis using hyperspectral imaging. *Computers and Electronics in Agriculture 197* (2022), 106921. doi:https://doi.org/10.1016/j.compag.2022.106921. 2, 3, 6

[ZOY*23] ZHANG B., OU Y., YU S., LIU Y., LIU Y., QIU W.: Gray mold and anthracnose disease detection on strawberry leaves using hyperspectral imaging. *Plant Methods 19*, 1 (2023), 148. doi:https://doi.org/10.1186/s13007-023-01123-w. 1, 2