Early Detection of PVY and PLRV in Potato Crops Using Hyperspectral Imaging

Tessa Postma, Intern Computer Vision & Artificial Intelligence
Supervisor: Willem Dijkstra

Summer 2026

Introduction

Potato Virus Y (PVY) and Potato Leafroll Virus (PLRV) are two of the most damaging viral diseases in European potato production, with PVY alone capable of reducing yields by up to 80%. Early detection is critical yet challenging: physiological changes occur weeks before visible symptoms appear.

Hyperspectral Imaging (HSI) captures hundreds of narrow spectral bands simultaneously, providing access to the complete spectral signature of plant tissue. This study presents a machine learning pipeline for cross-variety, early detection of both viruses using a Specim FX-17 camera (224 spectral bands, 900-1700 nm) across 9 potato varieties and 2 growth stages [1-2].

Abstract

A machine learning pipeline combining leaf segmentation with XGBoost classification detects PVY and PLRV in potato crops using hyperspectral imaging. Individual leaf features are extracted and aggregated into a plant-level profile capturing canopy-wide spectral statistics. LDA-guided selection of the most discriminative spectral bands outperforms the full spectrum. Cross-variety generalisation across multiple potato varieties and growth stages remains the main challenge.

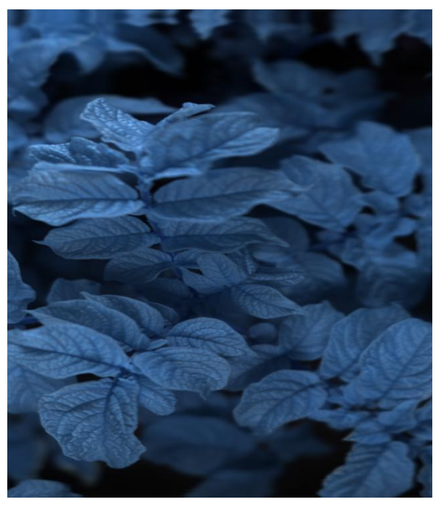


Figure 1. RGB preview of a potato plant from the hyperspectral dataset, generated from selected near-infrared bands.

Materials and Methods

Hyperspectral images of 9 potato varieties across 2 growth stages were acquired using a Specim FX-17 (224 bands, 900–1700 nm). Preprocessing included flat-field correction, spectral cropping, ℓ2-normalisation, and Savitzky-Golay filtering. Linear Discriminant Analysis (LDA) was applied to select the spectral bands that best separate the three disease classes. XGBoost was used as the classifier with a 1.5× PVY class penalty, evaluated using both plant-wise train/test splits and Leave-One-Variety-Out (LOVO) cross-validation across 8 varieties.

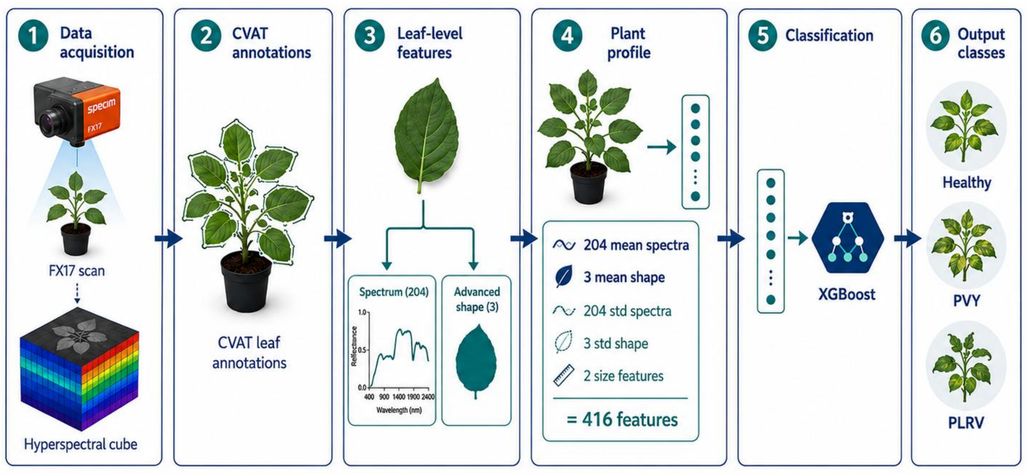


Figure 2. Overview of the workflow: from data acquisition and leaf annotation to leaf-level feature extraction, plant profile aggregation, and XGBoost classification into three disease classes.

Experiments

Five key experiments were conducted: a feature ablation study across 12 leaf-level feature combinations, a spectral band selection experiment comparing LDA (Fig. 3), PCA, and full spectrum alternatives (Table 1), a canopy size ablation testing the effect of filtering small and large leaves, a temporal drift experiment across growth stages, and LOVO cross-validation to evaluate cross-variety generalisation.

Acknowledgements

Supported by the European Union within the JTF-call project DigiAgro3. Thanks to NHL Stenden Professorship CVAI for dataset access, annotation tools, and computing resources.

NHL STENDEN
computer vision & artificial intelligence

Table 1. Spectral resolution ablation results for both growth stages. LDA Top 20 bands consistently outperforms the full 204-band spectrum in May (F1: 0.792) and June (F1: 0.866).

Band Configuration	May F1	May Accuracy	June F1	June Accuracy
LDA Top 20 Bands	**0.792**	**76.2%**	**0.866**	**88.5%**
Full Spectrum (204 Bands)	0.742	71.4%	0.662	69.2%
Every 10th (20 Bands)	0.707	71.4%	0.725	76.9%
LDA Top 15 Bands	0.626	61.9%	0.704	76.9%
Range A+B (42 Bands)	0.603	61.9%	0.768	80.8%
PCA Top 20 Bands	0.556	57.1%	0.718	73.1%
LDA Top 10 Bands	0.543	57.1%	0.765	80.8%
Range B Only (21 bands)	0.483	52.4%	0.737	76.9%
Range A Only (21 bands)	0.440	52.4%	0.527	61.5%

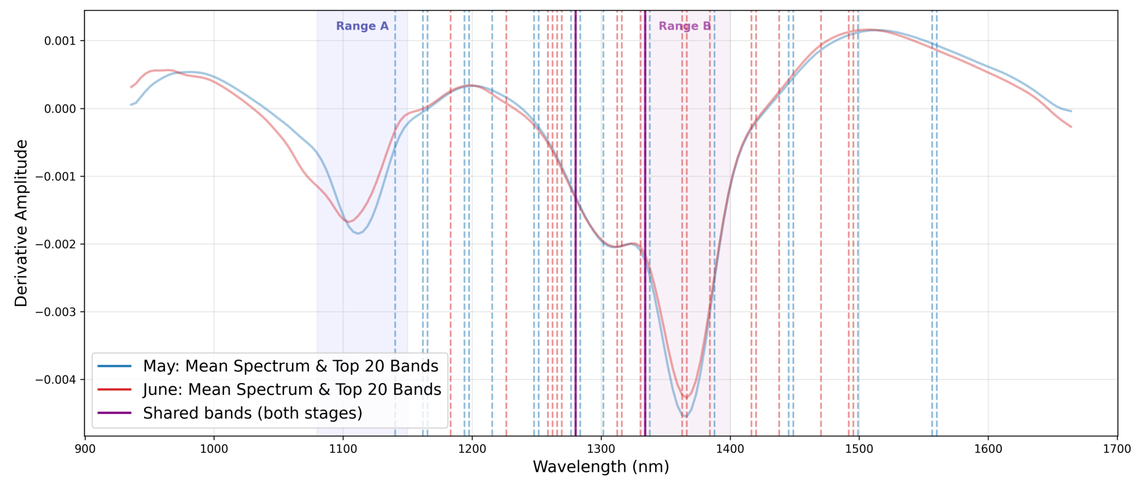


Figure 3. Top 20 LDA-selected discriminative bands for both growth stages overlaid on the mean first-derivative normalised reflectance spectrum. Blue dashed lines indicate May bands, red dashed lines indicate June bands, and purple solid lines indicate shared bands (1280 and 1334 nm).

Results

- **Compact leaf features**: Mean + Advanced Shape (207 features) matched the full 822-feature set at 4x lower dimensionality.
- **Spectral bands**: LDA Top 20 bands achieved 76.2% acc. (May) and 88.5% acc. (June), outperforming the full spectrum (Fig. 4, Table 1), confirming the curse of dimensionality, additional bands introduce noise rather than signal.
- **Canopy size matters:** Filtering both small and large leaves consistently degraded performance, confirming the full canopy must be retained.
- **Temporal drift**: May model on June data: 49.4% acc. June model on May data: 28.6%, a single model is not sufficient.
- **Cross-variety**: LOVO accuracies of 55.6% (May) and 46.8% (June) show generalisation remains the main challenge.

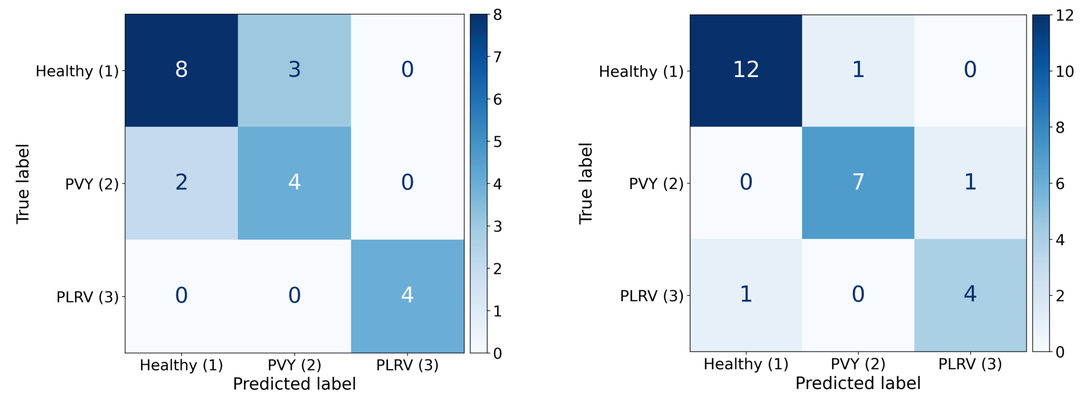


Figure 4. Confusion matrices for the optimal XGBoost models (Top 20 LDA bands, 1.5× PVY penalty, no filter). Left: early growth stage (May, 76.2%, F1: 0.792). Right: late growth stage (June, 88.5%, F1: 0.866).

Conclusions

This study shows that hyperspectral imaging with LDA-guided band selection can detect PVY and PLRV, but cross-variety generalisation and temporal drift remain key challenges before field deployment is feasible. Future work includes improving generalisation and exploring cheaper sensors targeting the most informative wavelengths.

References

[1] Griffel et al., Evaluation of artificial neural network performance for classification of potato plants infected with potato virus Y using spectral data on multiple varieties and genotypes, Smart Agricultural Technology, 2023.
[2] Polder et al., Potato virus Y detection in seed potatoes using deep learning on hyperspectral images, Frontiers in Plant Science, 2019.