Early Detection of Cracking in Grapes under Field Conditions using Hyperspectral Imaging and Machine Learning

By Yovel Gani

Supervised by:

Dr. Alon Sela, Prof. Victor Alchanatis, Prof. Irad Ben-Gal

August 2026

A thesis submitted toward the degree of Master of Science (M.Sc.)

Abstract

Grape cracking is a severe physiological disorder that leads to substantial economic losses in table-grape production by reducing marketable yield and increasing susceptibility to secondary infections such as sour rot. Early, non-destructive detection is therefore critical for enabling timely management interventions. This study investigates the feasibility of hyperspectral imaging (HSI; VNIR 400–1000 nm) combined with machine learning for crack detection under field conditions, considering both early- and late-stage acquisitions (initial/minimal vs. advanced cracking). All spectral analyses were restricted to 450–925 nm and employed per-spectrum Standard Normal Variate (SNV) normalization.

A multi-stage analytical framework was developed, including pixel-level classification, whole-image crack detection, and wavelength selection. Pixel-level spectral signatures of healthy and cracked berries were manually annotated and used to train an XGBoost classifier, achieving high discrimination performance (accuracy $\approx 97\%$, AUC ≈ 0.99) and confirming the existence of a distinct spectral separation between tissue states.

Direct application of pixel-level predictions to field scenes resulted in background-induced misclassification. To address this limitation, a whole-image crack-detection pipeline was developed, aggregating pixel-level CRACK probabilities using spatial post-processing (thresholding, morphological filtering, and patch/blob-based decision rules). The system was evaluated using a spatially disjoint row-based calibration–test split and reported with emphasis on the Cracked class. On the held-out test row, the pipeline achieved $\mathrm{F1}_{\mathrm{CRACK}} = 0.704$ under early-stage acquisition and $\mathrm{F1}_{\mathrm{CRACK}} = 0.818$ under late-stage acquisition.

Wavelength selection experiments demonstrated that comparable detection performance can be retained using reduced spectral-band subsets, indicating the potential for compact sensing configurations suitable for field deployment. Overall, the results establish the feasibility of HSI-based crack detection under realistic vineyard conditions and provide a methodological foundation for operational, non-destructive monitoring tools.

Contents

1 Introduction **4**

1.1 The Problem of Fruit Cracking 4

1.2 Economic Impact 4

1.3 Research Motivation 4

1.3.1 Main Research Question 6

1.4 Research Objectives and Hypotheses 6

1.4.1 Research Objectives 6

1.4.2 Contributions 6

2 Literature Review **8**

2.1 The Challenge of Grape Cracking: Physiology and Economic Impact . . . 9

2.2 Biomechanical Mechanisms and Cultivar Susceptibility 10

2.3 Bridging the Gap: The Need for Advanced Sensing in Complex Environments 11

2.4 Hyperspectral Imaging for Agricultural Applications 12

2.5 Machine Learning Paradigms for Hyperspectral Data 14

2.6 Related Work: From Visible Defect Detection to Early Warning 15

2.7 Research Gap and Study Positioning 16

3 Research Objectives and Hypotheses **17**

3.1 Research Objectives 17

3.2 Research Hypotheses 17

4 Materials and Methods **20**

4.1 Data Acquisition 20

4.1.1 Study Area and Experimental Design 20

4.1.2 Imaging Systems 22

4.1.3 Data Collection Protocol 25

4.1.4 Dataset Statistics 26

4.1.5 Pixel-Level Dataset Preparation 28

4.1.6 Expansion to Multi-Class Dataset 29

4.1.7 Whole-Image Dataset Construction and Experimental Design . . 30

4.2 Modeling and Evaluation Methodology 31
4.2.1 Global Spectral Preprocessing 32
4.2.2 Cross-Validation and Data Partitioning Strategy 33
4.2.3 Spectral Discriminability and Informative Wavelength Analysis . . 33
4.2.4 Pixel-Level Classification Framework 33
4.2.5 Anomaly Detection Framework 35
4.2.6 Whole-Image Detection and Aggregation Pipeline 37
4.2.7 Wavelength Selection 39
4.2.8 Evaluation Protocols and Performance Metrics 40
4.2.9 Reproducibility and Implementation Details 40

5 Experimental Results **41**
5.1 Overview of Experimental Results 41
5.1.1 Baselines 41
5.2 Spectral Separability and Informative Wavelengths 42
5.2.1 Mean Spectral Signatures of Healthy and Cracked Tissue 42
5.2.2 Spectral Separability Across Wavelength Regions 43
5.3 Pixel-Level Binary Classification 44
5.4 Pixel-Level Multi-Class Classification 45
5.5 Autoencoder-Based Anomaly Detection 46
5.6 Whole-Image Crack Detection 48
5.7 Wavelength Selection via Backward Feature Selection 51
5.7.1 BFS Trajectories and Subset Selection 51
5.7.2 End-to-End Wavelength-Selection Results 53
5.7.3 BFS Wavelength-Selection Stability Analysis 54

A Supplementary Results **62**
A.1 Cross-Validation Statistics and Full Precision Tables 62
A.1.1 Pixel-Level Classification: Binary and Multi-Class 62
A.1.2 Autoencoder-Based Anomaly Detection 62
A.2 Regional VIS–NIR Reflectance Summaries 62
A.3 Binary Pixel-Level Classification (Balanced) 65
A.4 Multi-Class Pixel-Level Classification (Balanced) 66
A.5 Autoencoder Anomaly Detection (Balanced) 67
A.5.1 One-Class Binary Formulation 67
A.5.2 Crack-Only Training Formulation 67
A.5.3 Non-Crack Training Formulation 68
A.6 Wavelength Selection: Generalization and Performance Analysis 69
A.6.1 BFS Thresholds (Full Precision) 69

A.6.2 Wavelength Results (Full Precision) 69
A.6.3 Generalization Analysis 69
A.6.4 Wavelength Overlap Across Seeds 69

List of Figures

2.1 The formula for reflectance Blackburn, 2007. 12
2.2 Hyperspectral Imaging (HSI) data cube illustrating the spectral signature of a pixel. 13
2.3 Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.). 13

4.1 Map of the experimental vineyard showing the five randomized blocks (A–E) and the four irrigation regimes applied within each block. 21
4.2 Hyperspectral imaging setup and data acquisition using the Specim IQ system under vineyard conditions. 22
4.3 Thermal imaging system and example thermal data acquired from grape clusters under vineyard conditions. 24
4.4 Graphical user interface of the HSI – RGB annotation tool. (1) Hyperspectral image visualized as an RGB composite with a SAM2-generated segmentation overlay, (2) wavelength-selectable grayscale HSI band with the same SAM2 segmentation mask, and (3) RGB image of the same grape cluster captured on the same day using a Canon EOS R5. (4) Spectral signature extracted from a single pixel within the SAM2-segmented region, and (5) autoencoder reconstruction of the same spectral signature, discussed later in this work. 28
4.5 Protocol overview for the row-based split, training and calibration on Row 1 Late-stage data, held-out evaluation on Row 2, and post-inference reporting stratified by acquisition stage. 31
4.6 Whole-image crack-detection pipeline from pixel-level CRACK probabilities to image-level classification via spatial aggregation. 38

5.1 Mean spectral signatures (± standard deviation) for healthy and cracked tissue across the VIS–NIR range. 42
5.2 Wavelength-wise spectral separability metrics between healthy and cracked tissue, including absolute mean difference, Cohen's d, and Fisher score. . 43

5.3 BFS trajectories for the whole-image pipeline: CRACK-class PR–AUC (primary objective) and CRACK-class F1 versus the number of retained wavelengths (n; full baseline n=159). 51
5.4 BFS threshold-marker visualization on CRACK metrics. Markers indicate the maximum and the first n where performance drops more than 0.5% and 1.0% below that maximum. 52
5.5 Distribution and popularity of the top-30 selected wavelengths across five varying-split seeds. 54
5.6 Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-30 subsets. 55
5.7 Stability heatmap for the top-30 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$). Each cell indicates how many seeds selected a wavelength within that 5 nm bin. 56
5.8 Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-30 subsets. 57
5.9 Distribution and popularity of the top-11 selected wavelengths across five varying-split seeds. 58
5.10 Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-11 subsets. 59
5.11 Stability heatmap for the top-11 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$; 0.5% drop setting). Each cell indicates how many seeds selected a wavelength within that 5 nm bin. 60
5.12 Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-11 subsets. 61

A.1 Regional mean reflectance (±1 SD) and absolute mean class difference for the VIS and NIR ranges. 64

List of Tables

1.1 Estimated financial losses from fruit cracking in major crops.(Alchanatis et al., 2022). 5

4.1 Pixel-level datasets. 26
4.2 Whole-image datasets. 26
4.3 Exact dataset statistics used in the experiments. For whole-image inference, Row 1 is used for calibration/tuning and Row 2 for held-out testing. 27
4.4 Pixel-level exports: dataset composition. 27
4.5 Multi-class pixel-level dataset: exported pixels by class. 27
4.6 Multi-class dataset: class definitions. 30
4.7 Autoencoder training hyperparameters. 36

5.1 Pixel-level classification metrics (unbalanced, binary setting). Truncated mean values. Full cross-validation statistics in Appendix A.1. 44
5.2 Pixel-level classification metrics (unbalanced, multi-class). Truncated mean values. Full cross-validation statistics in Appendix A.1. 45
5.3 Pixel-level autoencoder anomaly detection (binary setting). Truncated mean values. Full statistics in Appendix A.1. 46
5.4 Pixel-level autoencoder anomaly detection (crack-only training). Truncated mean values. Full statistics in Appendix A.1. 47
5.5 Pixel-level autoencoder anomaly detection (all non-CRACK training). Truncated mean values. Full statistics in Appendix A.1. 47
5.6 Dataset composition for full-image crack-detection experiments under early- and late-stage acquisition conditions (initial/minimal vs. advanced cracking). Row 1 was used for calibration, while Row 2 served as an independent test set. 48
5.7 Full-image crack detection performance for early- and late-stage acquisition conditions. F1 (Cracked) is the primary metric; the Cracked class is positive. 49
5.8 Full-image crack detection using autoencoder-based anomaly detection. F1 (Cracked) is the primary metric; the Cracked class is positive. 50
5.9 Summary of BFS wavelength-count thresholds for CRACK metrics. (Full precision in Appendix A.6.1) 51

5.10 Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions. (Full precision in Appendix A.6.2) 53

A.1 Pixel-level classification metrics (unbalanced training set, binary). Values are reported as mean ± standard deviation across folds where available. Precision/recall and ROC–AUC refer to the *Cracked* class (positive class); F1 is the weighted F1-score. 62
A.2 Pixel-level classification metrics (**unbalanced** training set, multi-class). Values are reported as mean ± standard deviation across folds. 63
A.3 Pixel-level autoencoder anomaly detection: binary setting (non-cracked as normal class). Values are reported as mean ± standard deviation across LOGO folds. 63
A.4 Pixel-level autoencoder anomaly detection: trained on CRACK pixels (multi-class setting). Values are mean across LOGO folds. 63
A.5 Pixel-level autoencoder anomaly detection: trained on all non-CRACK classes (multi-class setting). Single group-wise train–test split. 63
A.6 Summary statistics of reflectance values across spectral regions for healthy and cracked tissue. 63
A.7 Regional spectral separability metrics between healthy and cracked tissue. Mean values are averaged across wavelengths within each range, while peak values indicate the maximum separability and the corresponding wavelength. 64
A.8 Pixel-level classification metrics (balanced training set). Values are reported as mean ± standard deviation across folds where available. Precision/recall and ROC–AUC refer to the *Cracked* class (positive class); F1 is the weighted F1-score. 65
A.9 Pixel-level classification metrics (**balanced** training set). Values are reported as mean ± standard deviation across folds. 66
A.10 Pixel-level anomaly detection performance using autoencoders trained on non-cracked samples (balanced sampling). Values are reported as mean ± standard deviation across folds. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class). 67
A.11 Pixel-level anomaly detection performance using autoencoders trained exclusively on *Cracked* pixels and evaluated in a multi-class setting (balanced sampling). Values are reported as mean ± standard deviation across LOGO folds. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class). 67

A.12 Pixel-level anomaly detection performance using autoencoders trained on all semantic classes except *Cracked* and evaluated in a multi-class setting (balanced sampling). Values are reported for the held-out test split. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class). 68
A.13 Summary of BFS wavelength-count thresholds for CRACK metrics (full precision). 69
A.14 Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions (full precision). 69
A.15 Model Generalization Analysis: Comparison between Calibration and Test MCC for wavelength subsets. 70
A.16 Pairwise Jaccard similarity of top-30 wavelengths across seeds (varying split). Mean Jaccard = 0.493. 70
A.17 Pairwise Jaccard similarity of top-11 wavelengths across seeds (varying split). Mean Jaccard = 0.398. 70

Chapter 1

Introduction

1.1 The Problem of Fruit Cracking

Fruit cracking is a significant physiological disorder affecting a wide range of commercially important crops, including sweet cherries, tomatoes, pomegranates, and particularly table grapes (Vitis vinifera L.). The phenomenon is characterized by the rupture of the fruit's skin (cuticle and epidermis), exposing the internal flesh to the environment. These cracks render the fruit unmarketable and create entry points for pathogens, leading to rapid decay from secondary infections such as sour rot. Cracking typically occurs during the final stages of fruit ripening, a period when the fruit is most valuable. It is primarily caused by an imbalance between the turgor pressure within the fruit and the mechanical strength of its skin, a condition often exacerbated by environmental factors like rain, high humidity, and irrigation practices.

1.2 Economic Impact

The economic consequences of grape cracking are substantial. Direct losses can amount to up to 30% of the marketable yield. For table grapes, this translates to an estimated financial loss of €625 per ton, culminating in a total industry loss of approximately €175 million annually, based on European market data (Table 1.1). Beyond direct yield loss, cracking negatively impacts fruit appearance and quality, leading to price reductions of 20-50% and a shortened shelf-life, further diminishing profitability for growers.

1.3 Research Motivation

Current methods for managing cracking risk are reactive and often insufficient. There is a critical need for a non-destructive, reliable method to detect early signs of cracking susceptibility before visible fissures appear. Hyperspectral imaging (HSI) has emerged

Table 1.1: Estimated financial losses from fruit cracking in major crops.(Alchanatis et al., 2022).

Fruit	**Prod. (K t)**	**Coverage (K ha)**	**Price (€/kg)**	**Loss (%)**	**Fin. Loss (€/t)**	**Total loss (M€)**
Table grapes	280	7.0	2.5	25	625	175.0
Citrus	2300	58.5	1.1	20	220	506.0
Pomegranate	43	1.0	3.3	25	825	35.5
Sweet cherry	560	37.3	2.0	20	400	224.0

as a powerful tool in precision agriculture for assessing plant health and fruit quality due to its ability to capture detailed spectral information beyond the visible range. This research aims to investigate the potential of HSI and machine learning to address the challenge of grape cracking.

This thesis addresses two related but distinct analysis settings. *Identification/detection* concerns recognising existing crack tissue based on its hyperspectral signature and can be naturally posed as a pixel-level classification problem when representative labels are available. *Early-warning screening*, in contrast, targets pre-symptomatic field assessment under realistic vineyard conditions, where direct pixel-wise segmentation of complex scenes is often unstable: background clutter (e.g., leaves, wood, and artificial materials) and acquisition-dependent domain shifts (illumination, viewing geometry) can dominate individual pixel-level outputs and lead to unreliable image-level decisions. For this reason, the early-warning setting is approached through robust single-acquisition representations that aggregate information at the grape-cluster/image level. In particular, early-stage acquisitions are not a focal point for clear field identification because reliable field indicators are often not yet present, whereas late-stage acquisitions typically allow clearer field identification.

Terminology (Early vs. Late Acquisition). In this thesis, *detection/identification* refers to classification at the time of measurement using hyperspectral signatures. The terms *early-stage* and *late-stage* denote acquisition-stage subsets relative to crack progression within the same cluster, as defined by vineyard monitoring records:

- **Late-stage:** Images acquired after the first recorded cracking event, representing an advanced cracking condition.
- **Early-stage:** The earliest available image associated with the onset of cracking for the same cluster, representing the initial observable stage.

Both stages involve detection of crack-related spectral signatures at the time of measurement. These terms describe crack status at the time of acquisition, not a future outcome or pre-symptomatic prediction.

1.3.1 Main Research Question

Can hyperspectral imaging and machine learning models be used to detect and identify cracking-related status at the time of measurement under field conditions, including both early- and late-stage acquisitions (initial/minimal vs. advanced cracking)?

1.4 Research Objectives and Hypotheses

1.4.1 Research Objectives

1. **Identification:** To develop a high-accuracy classification model to differentiate between healthy and cracked grape tissue at the pixel level using their hyperspectral signatures.

2. **Whole-image screening:** To develop and evaluate an image-level crack-detection pipeline by combining pixel-level inference with spatial aggregation and post-processing, and to assess performance under early- and late-stage acquisition conditions (initial/minimal vs. advanced cracking).

3. **Wavelength selection:** To quantify the performance–complexity trade-off of reduced spectral-band subsets and identify stable candidate wavelengths to support compact sensing configurations.

1.4.2 Contributions

The main contributions of this thesis are as follows:

- A consistent VIS–NIR preprocessing protocol was established and applied throughout the study by restricting spectra to 450–925 nm and using per-spectrum SNV normalization, enabling direct comparability across experiments.

- A field-acquired hyperspectral dataset was curated at both pixel and whole-image levels, including manual pixel labeling and full-scene vineyard acquisitions, and was used to benchmark supervised classifiers under leave-one-group-out protocols and row-separated calibration/test evaluation, demonstrating reliable crack discrimination at the spectral-signature level (Chapter 5, Section 5.3).

- The limits of direct pixel-level segmentation in realistic vineyard scenes were analyzed by tracing error sources such as mixed pixels, illumination/BRDF effects, and spectrally confounding background materials, motivating a shift toward spatial aggregation (Chapter **??**, Section **??**).

- A whole-image crack-detection pipeline was developed by combining pixel-level inference with probability thresholding, morphological filtering, and patch/blob-based aggregation, with post-processing parameters optimized on calibration data and evaluated under early- vs. late-stage acquisition conditions (initial/minimal vs. advanced cracking; Chapter 5, Section 5.6).
- Backward Feature Selection (BFS) was used to quantify the performance–complexity trade-off of wavelength reduction and to assess selection stability across random seeds, providing a basis for compact band subsets and sensor-design implications (Chapter 5, Sections 5.7 and 5.7.3).

Chapter 2

Literature Review

This chapter establishes the theoretical and technological foundation for the non-destructive detection of grape cracking. The review is structured as a logical progression from the biological nature of the problem to the proposed computational solutions:

- **The Biological & Economic Context:** Explores grape cracking as a physiological disorder, detailing its biomechanical mechanisms and the substantial economic impact that necessitates proactive monitoring.
- **Sensing Technologies:** Introduces precision agriculture concepts and highlights Hyperspectral Imaging (HSI) as a powerful non-destructive sensing modality to address this agricultural need.
- **Computational Approaches:** Examines relevant machine learning paradigms, as extracting actionable insights from HSI data requires advanced analytics.
- **Related Work & Research Gap:** Reviews existing literature on fruit defect detection, culminating in the identification of the critical research gap addressed by this thesis: the absence of robust, field-ready frameworks for early-stage crack detection.

2.1 The Challenge of Grape Cracking: Physiology and Economic Impact

Fruit cracking is a major physiological disorder in table grapes (*Vitis vinifera* L.) characterized by the physical rupture of the berry skin, which severely degrades fruit quality and marketability (Guirao et al., 2025). This physical failure of the skin typically occurs during the late ripening stages (post-veraison). Veraison represents a critical transition phase marked by berry softening, shifts in water and solute transport, and rapid tissue expansion (Guirao et al., 2025; Monteiro et al., 2025). Cracking is primarily driven by a biomechanical imbalance: as internal turgor pressure increases—often due to rapid water uptake—the skin lacks sufficient elasticity to withstand the resulting stress (La Spada et al., 2024; Yu et al., 2025).

The economic impact of this disorder is substantial. Even minor skin fissures diminish visual quality and accelerate postharvest decay, leading to increased water loss, tissue browning, and higher susceptibility to secondary infections such as sour rot (Seo et al., 2022; Xu et al., 2025; C. Zhang et al., 2025). Crucially, because sour rot is highly contagious and spreads rapidly to adjacent healthy berries, a single cracked grape can act as a catalyst for a broader, cluster-wide outbreak (Hall et al., 2018). This cascading infection effect drastically multiplies crop losses, making early-stage monitoring essential to prevent localized damage from becoming an epidemic. Consequently, cracking results in significant preharvest rejection rates and postharvest waste (Santos et al., 2023). In the European table grape industry alone, annual financial losses from cracking are estimated at approximately €175 million, assuming a standard 25% yield loss (Alchanatis et al., 2022).

Currently, agricultural management relies heavily on reactive interventions, such as late-stage foliar sprays (e.g., calcium or salicylic acid), to strengthen the cuticle (Karantzi et al., 2021; La Spada et al., 2024). However, these strategies are often insufficient because they are applied late in the ripening process, when physiological vulnerability is already high and mechanical resistance is compromised (Chang & Keller, 2021; Chang et al., 2019; Yu et al., 2025). This limitation underscores a critical industry need for proactive, monitoring-based strategies that can identify cracking risk before macroscopic damage appears (Alchanatis et al., 2022).

2.2 Biomechanical Mechanisms and Cultivar Susceptibility

Understanding the physiological processes of the berry skin is essential for identifying when and how cracking develops. A primary driver of cracking is the dynamic relationship between water uptake and internal pressure. Even minor fluctuations in soil moisture or surface wetness can cause a rapid increase in berry volume. When mesocarp expansion outpaces skin extensibility—especially near maturity when cuticular deposition slows—tensile stress accumulates. This reduces the skin's elastic reserve and increases the likelihood of splitting Khadivi-Khub, 2014.

This mechanical stress initially manifests as microcracks within the cuticle, which acts as both a diffusion barrier and a protective shell. Microcracks can form days before visible splitting occurs. They propagate through structurally weak regions of the skin under mechanical stress or excessive hydration (e.g., from high humidity or rain) Chang and Keller, 2021; Emmons and Scott, 1998; Peschel and Knoche, 2005. Conversely, heat and drought conditions can induce structural fatigue through repeated cycles of dehydration and rehydration. Because these early microcracks are not visually apparent, sensitive spectral tools are required to detect them.

Beyond mechanical factors, cracking susceptibility is modulated by plant hormones and cultivar-specific traits. Hormones such as gibberellins (GA_3) and abscisic acid (ABA), which are routinely used in table grape viticulture, alter skin biomechanics, fruit firmness, and cell wall architecture, indirectly influencing the risk of cracking Castellarin and other, 2016; Lichter et al., 2014. Cultivars with thin cuticles or rapid sugar accumulation, such as the 'Scarlotta Seedless' evaluated in this study, exhibit heightened sensitivity to temperature fluctuations and hydration-induced damage A. Author et al., 2022; B. Author et al., 2024; Chang and Keller, 2021. Once the cuticle barrier is breached, the fruit becomes highly vulnerable to secondary diseases such as sour rot Hall et al., 2018. Therefore, the onset of pre-symptomatic microcracking represents a pivotal point of failure. This underscores the necessity for non-destructive optical methods, such as hyperspectral imaging, to identify these latent structural vulnerabilities.

2.3 Bridging the Gap: The Need for Advanced Sensing in Complex Environments

Detecting grape cracking before it becomes visually apparent is a highly complex challenge that cannot be reliably addressed through traditional visual agronomic observation. As discussed in the previous sections, the critical physiological shifts—such as internal pressure imbalances and micro-structural failures—occur hidden from the naked eye. Furthermore, these subtle changes happen in the vineyard, an environment characterized by immense natural variability. Addressing a biological problem of this complexity requires moving beyond conventional methods and adopting innovative, data-driven frameworks capable of capturing invisible signals.

While standard computer vision (CV) using simple RGB cameras has been widely adopted in agriculture for tasks like yield estimation or identifying obvious external defects Patrício and Rieder, 2018; H. Tian et al., 2020, it is fundamentally ill-equipped for early crack detection. Standard cameras primarily capture surface color information. In addition, their performance heavily degrades under the dynamic lighting, shadows, and complex backgrounds typical of a real-world vineyard Benos et al., 2021; Dhanya et al., 2022. Because standard RGB imaging cannot probe beneath the skin of the fruit, it completely misses the pre-symptomatic physiological stress that precedes actual cracking.

To overcome these limitations, the complexity of the problem must be met with equally sophisticated tools. Solving this requires a paradigm shift: coupling non-destructive, deep-sensing technologies with advanced artificial intelligence. Specifically, there is a critical need for sensing modalities that can capture broad spectral information—revealing both the chemical and structural state of the tissue—paired with machine learning algorithms capable of decoding this high-dimensional data Cao et al., 2025. This necessity forms the core motivation for transitioning to Hyperspectral Imaging (HSI) as an innovative solution for early defect detection.

2.4 Hyperspectral Imaging for Agricultural Applications

Hyperspectral imaging (HSI) overcomes the limitations of conventional vision systems by capturing detailed spectral information across a broad and continuous range of electromagnetic wavelengths. While traditional RGB cameras record reflected light in only three broad bands (red, green, and blue) and multispectral imaging (MSI) captures a limited discrete set (typically 3 to 15 bands), HSI acquires reflectance data across hundreds of narrow, contiguous spectral bands, often with a resolution finer than 5 nm Benelli et al., 2020; Mehta, 2018. This produces a three-dimensional data cube where each spatial pixel contains a dense, continuous spectral vector, enabling the detection of both surface and subsurface physicochemical properties Hong et al., 2025.

The fundamental metric in this analysis is reflectance, defined as the ratio of radiation reflected by a surface to the radiation it receives. This normalization ensures that the measurement reflects the intrinsic chemical and micro-topographical properties of the material rather than fluctuations in illumination intensity Blackburn, 2007. Figure 2.1 formalizes this relationship.

$$Reflectance = \phi out \ / \ \phi in$$
$$\text{where } \phi = \text{radiation}$$

Figure 2.1: The formula for reflectance Blackburn, 2007.

When exposed to light, different materials absorb and reflect specific wavelengths based on their molecular bonds, creating a unique "spectral signature" Shaw and Manolakis, 2002. This allows HSI to reliably distinguish between tissues that appear identical to the naked eye but differ physiologically. Figure **??** illustrates this principle, demonstrating how variations in moisture content produce systematic shifts in the spectral signature.

In agricultural applications, the visible and near-infrared (VIS–NIR) spectrum is particularly valuable. As shown in Figure 2.3, this range (typically 400–1000 nm) spans both the visible region—which correlates strongly with plant pigments like chlorophyll—and the NIR region, which is highly sensitive to water content and internal cellular structure Ball, 2007; T. Wang and Zhao, 2024; Y. Zhang et al., 2022; Zwinkels, 2015.

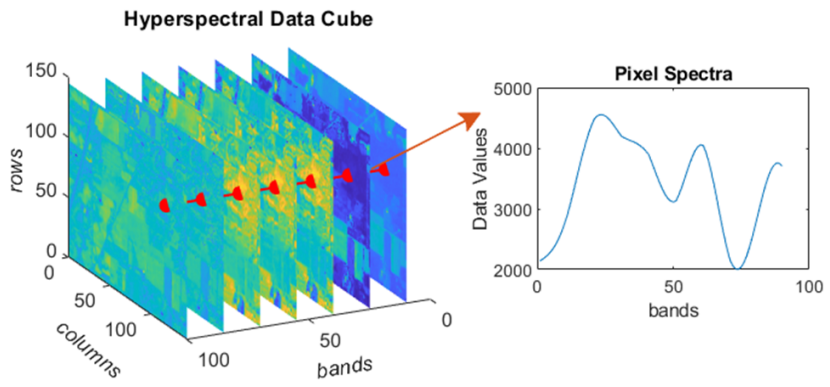


Figure 2.2: Hyperspectral Imaging (HSI) data cube illustrating the spectral signature of a pixel.

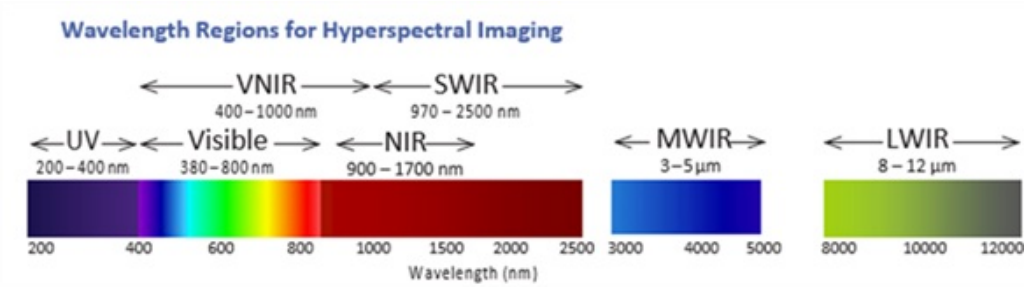


Figure 2.3: Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.).

This dual sensitivity makes VIS–NIR HSI uniquely suited for early defect detection in fruit. Because grape cracking is fundamentally driven by water uptake and cuticular structural failure, the high sensitivity of the NIR bands to moisture and cellular integrity allows for the detection of physiological stress and microcracking before physical symptoms emerge Abade et al., 2025; Liu et al., 2023; Q. Sun et al., 2024. Consequently, HSI has been successfully employed to assess internal quality attributes (such as soluble solid content and firmness) and detect both external and internal defects, including bruises, decay, and scars, across various fruit species Akter et al., 2025; Kanwal et al., 2025; Min et al., 2023; P. Tian et al., 2023; Yang et al., 2025; X. Zhang, Li, et al., 2024; Zhu et al., 2025.

By combining the rich spectral data of HSI with advanced machine learning techniques, it is possible to transition from post-symptomatic visual inspection to proactive, field-ready monitoring. This establishes the necessary technological foundation for the early-warning crack detection framework proposed in this thesis.

2.5 Machine Learning Paradigms for Hyperspectral Data

Extracting actionable insights from high-dimensional hyperspectral data requires robust analytical frameworks. Traditional machine learning (ML) approaches, such as Partial Least Squares (PLS) and Support Vector Machines (SVM), have been widely adopted in agricultural applications due to their effectiveness on small-to-medium datasets and their interpretability García-Vera et al., 2024; Ram et al., 2024. These models heavily rely on established spectral preprocessing pipelines—including Savitzky–Golay smoothing and Standard Normal Variate (SNV) normalization—to mitigate sensor noise and illumination variability Hasanzadeh et al., 2022; Khaled et al., 2023. A major advantage of classical ML is its compatibility with feature extraction and wavelength selection techniques (e.g., CARS), which reduce spectral redundancy and computational complexity, making the models highly suitable for fast, field-ready applications Z. Zhang et al., 2024.

In recent years, deep learning architectures have emerged as powerful alternatives capable of exploiting complex non-linear structures and spatial contexts without hand-crafted features Chen et al., 2024. However, hyperspectral deep learning in agriculture faces severe practical constraints: it requires massive amounts of densely annotated data, which is rarely available for pixel-level field tasks. Furthermore, the high dimensionality of HSI combined with limited field data significantly increases the risk of overfitting and domain shift Attri et al., 2023. Consequently, classical pipelines—combining explicit spectral preprocessing, interpretable models, and wavelength selection—remain operationally attractive and highly competitive for agricultural datasets.

Beyond algorithm selection, the choice of modeling paradigm dictates system robustness. Pixel-level modeling enables fine-grained discrimination of subtle biochemical variations, which is critical for detecting micro-scale pre-symptomatic stress. Yet, pixel-wise approaches are highly sensitive to background clutter (e.g., soil, leaves), mixed pixels, and severe class imbalances typical of early-stage defect detection Ram et al., 2024. To overcome these practical field challenges, modern workflows increasingly rely on spatial aggregation strategies, translating sensitive pixel-level classifications into conservative, robust image-level or region-level decisions Chen et al., 2024; Fass et al., 2025.

2.6 Related Work: From Visible Defect Detection to Early Warning

Prior applications of HSI in fruit quality analysis can be broadly categorized into two objectives: the detection of already manifested defects, and the early identification of latent physiological disorders. HSI is a well-established and highly accurate tool for the former, successfully employed in postharvest inspection to detect visible blemishes, mechanical injuries, and advanced decay across diverse fruit types Min et al., 2023; Vignati et al., 2023; B. Wang et al., 2023; C. Wang et al., 2025.

More recently, research has shifted toward pre-symptomatic detection. Several studies have demonstrated that HSI can capture latent spectral changes associated with internal disorders, such as early-stage bitter pit in apples Dong et al., 2025 and initial pathogen stress like gray mold or bacterial leaf spots before visual symptoms appear Haghbin et al., 2023; Ou et al., 2024; X. Zhang, Vinatzer, and Li, 2024; Y. Zhang et al., 2024.

Despite these promising results, the vast majority of both visible and early-detection studies remain confined to highly controlled laboratory or postharvest environments. Operational field deployment introduces immense environmental variability, including dynamic ambient lighting, complex backgrounds, and shifting plant phenology, all of which distort spectral signatures and degrade model transferability Vignati et al., 2023; C. Wang et al., 2025. Because dense data annotation under such chaotic field conditions is prohibitively difficult, true preharvest frameworks that provide early-warning signals under realistic vineyard conditions remain critically scarce Haghbin et al., 2023; Y. Zhang et al., 2024.

2.7 Research Gap and Study Positioning

The literature review reveals a significant conceptual and methodological disconnect. On one hand, the physiological and biomechanical mechanisms driving grape cracking are well documented, establishing that internal pressure imbalances and micro-structural failures precede macroscopic skin rupture. On the other hand, while hyperspectral imaging has proven highly capable of detecting latent internal defects, its application has been largely restricted to reactive, post-symptomatic identification or highly controlled laboratory settings.

Consequently, a critical research gap persists: **there is currently no established, non-destructive framework capable of providing proactive, early-warning detection of grape cracking susceptibility under realistic preharvest field conditions.**

The present study is directly positioned to bridge this gap. By integrating VIS–NIR hyperspectral imaging with optimized classical machine learning and spectral wavelength selection, this research moves beyond reactive defect detection. Unlike prior works, the proposed framework emphasizes early-stage identification under actual vineyard variability. By explicitly linking highly sensitive pixel-level spectral classification with robust whole-image spatial aggregation, this thesis introduces a novel, field-ready early warning tool designed to identify the physiological onset of grape cracking before it translates into irreversible economic loss.

Chapter 3

Research Objectives and Hypotheses

3.1 Research Objectives

The primary objective of this research is to develop and validate hyperspectral imaging (HSI) methods for crack-related *detection/identification at the time of measurement* in table grapes ('Scarlotta' cultivar) under field conditions, spanning both early-stage (absent-to-minimal/incipient cracking) and late-stage (advanced, macroscopically visible cracking) acquisition conditions. The work bridges the gap between spectroscopic research on cracking mechanisms and practical, non-destructive agricultural screening by leveraging imaging, machine learning, and robust whole-image aggregation.

3.2 Research Hypotheses

The research is guided by the following hypotheses, structured hierarchically to reflect a progression from physical feasibility to algorithmic performance, robustness, early detection capability, and spectral efficiency.

Hypothesis 1: Spectral Distinguishability (Physical Basis)

This hypothesis examines whether a fundamental physical difference exists in the spectral reflectance properties of healthy and cracked grape tissues.

- H_0**:** There is no statistically significant difference in the spectral signatures of healthy and cracked tissues in 'Scarlotta' grapes.
- H_1**:** Statistically significant differences exist in the spectral signatures of healthy versus cracked grape skins at specific wavelengths, providing a physical basis for subsequent algorithmic discrimination.

Hypothesis 2: Pixel-Level Classification Feasibility

This hypothesis evaluates whether machine learning algorithms can effectively learn the spectral differences identified in Hypothesis 1 under controlled, pixel-level conditions.

- H_0**:** A machine learning model trained on hyperspectral data cannot achieve classification performance, as measured by standard metrics (e.g., F1-score, precision/recall, or MCC), significantly higher than a random or baseline classifier at the pixel level.
- H_1**:** The proposed machine learning framework can successfully classify individual pixels as *Cracked* or *Healthy* with high performance, effectively learning the underlying spectral decision boundary.

Hypothesis 3: Whole-Image Detection and Robustness

This hypothesis assesses the robustness of the trained models when applied to realistic whole-image hyperspectral scenes containing complex background elements not present in the clean pixel dataset.

- H_0**:** The model fails to distinguish between the spectral signatures of cracks and complex background clutter (e.g., leaves, woody tissue, or plastic), resulting in a high false positive rate when applied to full hyperspectral images.
- H_1**:** The spectral signatures of cracks are sufficiently distinct from background clutter to allow the model to effectively suppress noise and detect cracks within full-image scenes with high precision.

Hypothesis 4: Early Detection Capability

This hypothesis investigates the ability of the whole-image framework to detect cracking at its onset, prior to the appearance of visually observable symptoms.

- H_0**:** The spectral changes associated with the early onset of cracking are too subtle to be reliably detected, limiting the model's applicability to late-stage damage.
- H_1**:** The spectral model is sufficiently sensitive to detect early physiological changes and micro-cracks, enabling early-stage identification of clusters exhibiting crack-related spectral signatures at the time of measurement (even when macroscopic cracking is not yet clearly visible).

Hypothesis 5: Spectral Efficiency and Operational Feasibility

This hypothesis examines whether the detection capabilities demonstrated in Hypotheses 3 and 4 can be preserved when using a reduced subset of informative spectral bands.

- H_0**:** Reducing the number of spectral bands results in a significant degradation in detection performance, either in whole-image scenarios or during early-stage detection, implying that the full hyperspectral spectrum is required.
- H_1**:** A carefully selected subset of informative wavelengths captures the key physiological markers of cracking, enabling robust detection in both whole-image and early-stage scenarios using cost-effective multispectral sensing configurations.

This thesis is organized into several interconnected chapters. The *Literature Review* chapter provides a comprehensive overview of grape cracking mechanisms and relevant precision agriculture techniques. The *Materials and Methods* chapter describes the experimental design, data acquisition procedures, hyperspectral imaging systems, and modeling approaches used in this study. The *Results* chapter presents the experimental findings, including pixel-level classification, whole-image crack detection, and wavelength selection experiments. The *Discussion* chapter interprets these results in the context of early crack detection, physiological relevance, and practical deployment considerations. Finally, the *Conclusions and Future Work* chapter summarizes the main contributions of the study and outlines directions for future research.

Chapter 4

Materials and Methods

4.1 Data Acquisition

4.1.1 Study Area and Experimental Design

The study was conducted in the Lachish region of Israel, an area characterized by a semi-arid climate with a hot, dry summer and a mild winter. These conditions make the region a central hub for grape cultivation (viticulture) in Israel. The experiment was conducted in a commercial table grape vineyard (*Vitis vinifera* L., cv. 'Scarlotta'), in collaboration with the "Tali Grapes" brand. The vineyard was grafted onto 1103 Paulsen rootstock and was planted in 2018.

Figure 4.1 summarizes the block structure and irrigation assignment used throughout the study, clarifying the spatial provenance of the samples used later for row-based calibration and testing.

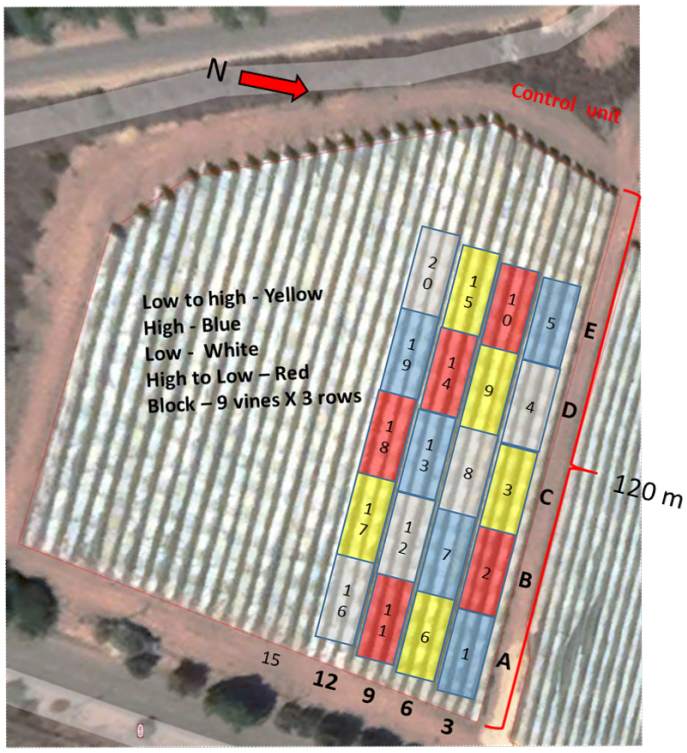


Figure 4.1: Map of the experimental vineyard showing the five randomized blocks (A–E) and the four irrigation regimes applied within each block.

The experimental design followed a Randomized Complete Block Design (RCBD), including five replicates (Blocks A–E). Within each block, four different irrigation regimes were tested, resulting in a total of 20 experimental plots (5 blocks $\times$ 4 treatments).

Each experimental plot consisted of three consecutive rows of nine vines each (27 vines per plot). To minimize edge effects and water spillover between irrigation regimes, only vines from the middle row of each plot were selected for measurement. From each middle row, 12 representative grape clusters were tagged, yielding a total sample size of 240 clusters (20 plots $\times$ 12 clusters).

The four irrigation regimes tested were:

- **High Irrigation (Blue):** 796 mm of water throughout the season.
- **Low Irrigation (White):** 376 mm of water throughout the season.
- **Reducing Irrigation (Red):** A transition from high to low irrigation (total 532 mm), beginning at the post-veraison stage.
- **Increasing Irrigation (Yellow):** A transition from low to high irrigation (total 645 mm), beginning at the post-veraison stage.

Irrigation commenced on April 29, 2024, and was applied daily or every other day via a drip irrigation system.

The experiment continued until the end of September 2024. Physiological and yield measurements—including stem water potential, stomatal conductance, fruit properties, and percentage of healthy clusters—were performed weekly. Physiological data were collected using a Scholander Pressure Bomb and a Porometer, and soil moisture was measured using Soil Moisture Sensors.

4.1.2 Imaging Systems

To study grape cracking in a comprehensive way, a multi-modal imaging setup was used. The system combined three imaging sensors placed at the same location, each capturing different types of information from a different part of the electromagnetic spectrum. Together, these sensors provided complementary spectral, spatial, and thermal data, allowing a more complete analysis of grape berry condition and cracking development. All images were collected from a fixed distance of approximately 1 m from the grape clusters to ensure consistent scale and alignment across all measurements.

Hyperspectral Imaging System (HSI)

Hyperspectral data were collected using the **Specim IQ** (Specim, Spectral Imaging Ltd., Oulu, Finland), a portable snapshot hyperspectral camera designed for field applications. The camera operates in the Visible and Near-Infrared (VNIR) spectral range (400–1000 nm) and produces 3D datacubes of size $512 \times 512 \times 204$ for each snapshot. Figure 4.2 documents the practical acquisition setup under vineyard conditions; this field configuration is important context for the later performance gap between pixel-level and whole-image inference under background clutter and illumination variability.

(a) Field deployment of the Specim IQ system in the vineyard.

(b) Hyperspectral image acquisition of grape clusters using the Specim IQ.

Figure 4.2: Hyperspectral imaging setup and data acquisition using the Specim IQ system under vineyard conditions.

- **Spectral Resolution:** The system captures 204 contiguous spectral bands with a Full Width at Half Maximum (FWHM) of 7 nm and a spectral sampling rate of approximately 3 nm. The high spectral fidelity is ensured by a signal-to-noise ratio (SNR) of $> 400:1$.

- **Spatial Resolution:** The sensor has a resolution of 512×512 pixels. With a $31° \times 31°$ Field of View (FOV), at a 1 m working distance, the camera covers a physical area of approximately 0.55 m $\times$ 0.55 m. This configuration yields a theoretical spatial resolution of approximately **1.08 mm/pixel**, sufficient for resolving individual berries and larger cracks.

- **Radiometric Calibration:** A spectrally flat, diffuse white reference panel (Spectralon®, $\approx 99\%$ reflectance) was included in every captured scene. This allowed subsequent normalization of raw radiance data into relative reflectance values, accounting for varying natural illumination conditions in the vineyard.

RGB Imaging System

To provide high-resolution visual context and capture fine surface textures that might be missed by the lower spatial resolution of the HSI system, a high-end digital camera was used.

- **Camera & Sensor: Canon EOS R5**, equipped with a 45-megapixel full-frame (36×24 mm) CMOS sensor, producing 14-bit RAW images with a resolution of 8192×4608 pixels.

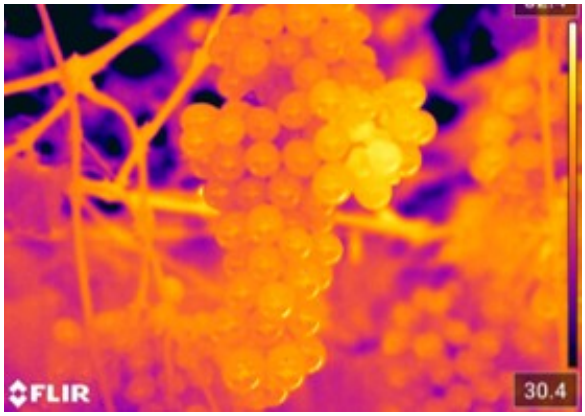
- **Optics:** The camera was fitted with a **35 mm** prime lens, providing a wide Field of View of 54.4° (horizontal) $\times$ 37.8° (vertical) and sufficient depth of field to keep the entire cluster in focus.

- **Spatial Resolution:** At the 1 m working distance, this system achieved an extremely high theoretical spatial resolution of approximately **0.13 mm/pixel**. This level of detail is critical for detecting micro-cracks (hairline fissures) and minute alterations in the berry skin cuticle before they become visible to the naked eye or the HSI sensor.

Thermal Imaging System

Canopy and grape cluster temperatures, which are widely used as indirect indicators of plant water status, stomatal conductance, and transpirational cooling dynamics, were monitored using a high-definition thermal imaging system.
Figure 4.3 illustrates the thermal modality acquired alongside Theramal Camera (FLIR T1020).

(a) FLIR T1020 thermal camera used for vineyard measurements.



(b) Thermal image of a grape cluster showing surface temperature variations.

Figure 4.3: Thermal imaging system and example thermal data acquired from grape clusters under vineyard conditions.

- **Camera: FLIR T1020** (Teledyne FLIR LLC, Wilsonville, OR, USA), a high-performance uncooled microbolometer specifically designed for precision research.
- **Sensor & Optics:** The camera has a thermal resolution of 1024×768 pixels with a 17 μm pixel pitch, operating in the Long-Wave Infrared (LWIR) spectral range (7.5–14 μm). It was equipped with a standard **34 mm** lens (adjustable focus including autofocus), providing a FOV of 45° (H) × 34° (V).
- **Performance:** The system boasts a high thermal sensitivity (Noise Equivalent Temperature Difference, NETD) of < 20 mK (< 0.02°C) at 30°C, enabling the detection of subtle temperature gradients across the grape cluster.
- **Spatial Resolution:** At the 1 m working distance, the theoretical thermal spatial resolution was approximately **0.81 mm/pixel**, allowing for accurate temperature extraction from individual berries while minimizing mixed-pixel effects from the surrounding air or leaves.

Due to the high dimensionality of the acquired data, individual hyperspectral images were substantially larger than the corresponding RGB and thermal images. On average, a single hyperspectral datacube occupied approximately **200 MB**, compared to approximately **11 MB** for a high-resolution RGB image and **2.5 MB** for a thermal image. In total, the imaging campaign resulted in a dataset of approximately **575 GB** of raw data, highlighting the computational and storage challenges associated with multi-modal hyperspectral analysis.

4.1.3 Data Collection Protocol

Data collection was conducted during concentrated field days, typically once a week (Mondays or Thursdays), throughout the growing season from June to the end of September. To ensure consistent lighting conditions, field operations were strictly scheduled between 08:00 and 17:00.

Daily Schedule and Logistics

Each collection day began at 05:30 at the Volcani Institute with equipment organization and verification. Departure to the vineyard in Lachish occurred at approximately 06:00. Upon arrival at 07:00–08:00, the imaging systems were mounted on their respective tripods and calibrated.

Sampling Strategy

Due to significant differences in acquisition time between modalities, a split sampling strategy was employed:

- **RGB Imaging:** All 240 marked grape clusters (12 clusters $\times$ 20 plots) were imaged first during each field day to capture them under relatively uniform early-morning lighting.
- **Hyperspectral Imaging (HSI):** Due to the complex acquisition process—which required re-positioning the camera and white reference panel for each cluster, followed by a dedicated spectral focusing procedure—each HSI scan took approximately 2-3 minutes. Consequently, only a subset of 120 clusters was imaged per session. These clusters were selected to represent a balanced cross-section of the experimental blocks, ensuring full dataset coverage over alternating sessions. HSI acquisition typically commenced around 10:00 AM when sunlight was more stable and intense.
- **Thermal Imaging:** Thermal image acquisition was purposely scheduled for later in the day (starting around 12:00 PM), after RGB imaging was complete. This timing leverages the high ambient temperatures typical of midday in the semi-arid Negev region to maximize thermal contrast on the berry surface. Specifically, cracked or decaying areas, often characterized by exposed moist tissue or conversely by localized dehydration, exhibit different thermal signatures compared to healthy, intact skin due to variations in evaporative cooling rates. Capturing images during peak heat enhances these temperature differentials, improving the detectability of such defects.

Field Procedure

To ensure accurate re-imaging, a consistent labeling system was implemented. Target clusters were marked with laminated, color-coded tags attached via plastic cable ties. The tags featured a numeric identifier (1–60), which was repeated across treatments but distinguished by the specific color assigned to each irrigation regime. This design allowed for unambiguous identification and durability against humidity. Furthermore, vine locations were georeferenced using a GPS module integrated into the thermal camera.

Although only the RGB and HSI datasets were analyzed in the present study, all imaging systems were spatially co-aligned and temporally synchronized during field acquisition to facilitate potential future analyses.

Table 4.1: Pixel-level datasets.

Dataset	Labels	Purpose	Method	H.
Binary	Healthy; Cracked	Spectral differences: intact vs damaged	RGB Pixel → SAM2 → HS segment extract	H1, H2
Multi-Class (10)	Cracked Healthy Leaf Background and more...	Spectral signatures: all materials	RGB Pixel → SAM2 → HS segment extract	H3

Table 4.2: Whole-image datasets.

Dataset	Acquisition Stage	Purpose	Method	H.
Late-Stage	Advanced cracking	Detection under visible crack conditions	Binary label	H3
Early-Stage	Initial onset	Detection at early crack-development stage	Binary label	H4

Note: Early/late refer to acquisition-stage subsets at different crack-progression levels (time of measurement). HSI+RGB pairs. Pixel: segment→label→extract. Image: direct labeling.

4.1.4 Dataset Statistics

Dataset	Split / Variant	Samples	Positive (CRACK)	Negative (REGULAR)
Pixel-Level Binary	Exported labeled pixels	94,166	24,548	69,618
Pixel-Level Multi-Class	Exported labeled pixels	1,189,041	24,548	69,618
Whole-Image Early	Row 1 calibration	172	32	140
Whole-Image Early	Row 2 test	60	32	28
Whole-Image Late	Row 1 calibration	160	35	125
Whole-Image Late	Row 2 test	60	34	26

Table 4.3: Exact dataset statistics used in the experiments. For whole-image inference, Row 1 is used for calibration/tuning and Row 2 for held-out testing.

Pixel-level exports.

Dataset	Labeled Pixels	Source
Binary	94 166	80 segments across 19 hyperspectral images
Multi-class	1 189 041	227 segments across 54 hyperspectral images

Table 4.4: Pixel-level exports: dataset composition.

Multi-class exported pixels by class.

Class	Pixels	Description
BACKGROUND	378,194	background pixels (e.g., soil and mixed background).
WHITE_REFERENCE	270,159	White reference panel pixels.
TRIPOD	137,494	Tripod and imaging setup components.
LEAF	120,552	Vine foliage.
BRANCH	78,654	Woody vine structures.
REGULAR	69,618	Intact grape tissue without visible defects.
PLASTIC	66,185	Artificial vineyard elements.
BURNT_PIXEL	36,676	Saturated or corrupted pixels.
CRACK	24,548	Grape tissue exhibiting visible cracking symptoms.
IRON	6,961	Metallic elements.

Table 4.5: Multi-class pixel-level dataset: exported pixels by class.

4.1.5 Pixel-Level Dataset Preparation

To enable efficient creation of the pixel-level dataset—required to test spectral differences between healthy and cracked grape tissues (Hypothesis 1) and to assess pixel-level classification feasibility using machine learning (Hypothesis 2)—a custom graphical user interface (GUI) was developed in Python. The tool integrates data visualization, interactive annotation, and model inference in a single platform (see Figure 4.4). The dataset generation workflow proceeded as follows:

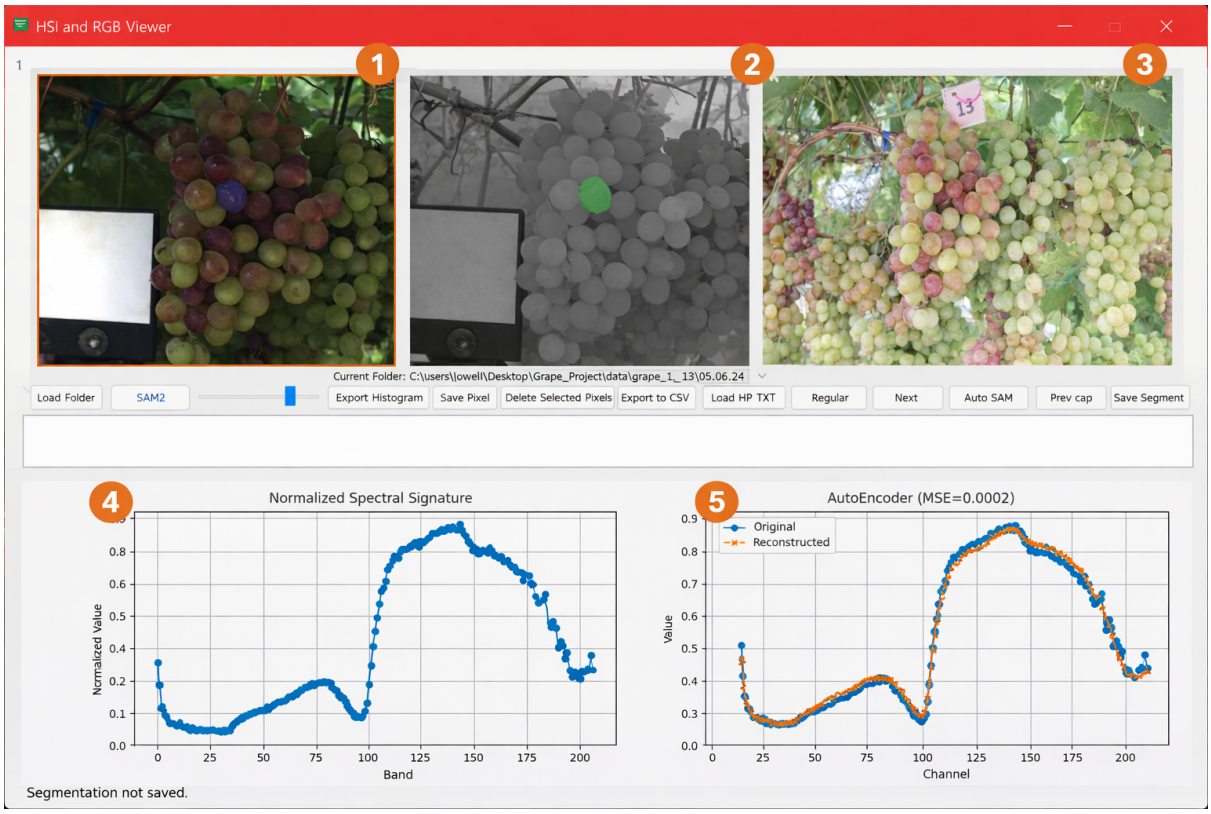


Figure 4.4: Graphical user interface of the HSI – RGB annotation tool. (1) Hyperspectral image visualized as an RGB composite with a SAM2-generated segmentation overlay, (2) wavelength-selectable grayscale HSI band with the same SAM2 segmentation mask, and (3) RGB image of the same grape cluster captured on the same day using a Canon EOS R5. (4) Spectral signature extracted from a single pixel within the SAM2-segmented region, and (5) autoencoder reconstruction of the same spectral signature, discussed later in this work.

1. **Image Loading and Visualization:** The tool displayed the hyperspectral datacube alongside corresponding RGB images (HSI camera and Canon EOS R5). A slider enabled interactive exploration of all 204 spectral bands (400–1000 nm).

2. **Interactive Pixel Sampling:** The user selected a seed pixel by clicking anywhere on a target object in either the RGB or HSI image. For the binary dataset, berries exhibiting clear cracked or healthy states were selected.

3. **SAM2 Segmentation on RGB:** The seed pixel was passed to **Segment Anything Model 2 (SAM2) ravi2024sam2** as a point prompt. SAM2 operated on the RGB image to generate a binary segmentation mask of the selected object. This mask was then spatially aligned and applied to the corresponding hyperspectral datacube to identify the same region.

4. **HSI Data Extraction and Labeling:** The tool extracted spectral signatures from the hyperspectral datacube for all pixels within the segmented region. The user assigned a class label (e.g., 'Cracked' or 'Regular' for binary classification) to the object, and the annotated spectral data were stored.

The final exported binary pixel-level dataset used in this work contains 94,166 labeled pixels (24,548 CRACK; 69,618 REGULAR), as summarized in Table 4.3.

4.1.6 Expansion to Multi-Class Dataset

The binary pixel-level dataset contained two classes: healthy grape tissue and cracked grape tissue. A multi-class pixel-level dataset was constructed to include additional classes representing non-grape objects present in vineyard scenes. This dataset was designed for training multi-class classifiers and whole-image semantic segmentation models, as specified in Hypothesis 3.

The same annotation workflow was applied to acquire spectral samples from background objects. Seed pixels were selected on target objects (leaves, branches, soil, artificial materials, imaging equipment) in the RGB or HSI images. SAM2 segmented each selected object on the RGB image, and the resulting mask was applied to the hyperspectral datacube. Spectral signatures were extracted for all pixels within each segmented region. Each region was assigned a class label, and the labeled spectral samples were stored.

The multi-class dataset comprised ten classes:
The class distribution and dataset size are reported in Table 4.5. The multi-class dataset was exported as labeled spectral samples and used for training supervised multi-class classification models.

Class	Description
CRACK	Grape tissue with visible surface cracks
REGULAR	Intact grape tissue
LEAF	Green vine foliage
BRANCH	Woody vine structures (stem, cane)
BACKGROUND	Soil, ground, and mixed non-object pixels
PLASTIC	Plastic vineyard elements (covers, ties)
TRIPOD	Camera tripod components
WHITE_REFERENCE	Calibration panel pixels
BURNT_PIXEL	Saturated or overexposed pixels
IRON	Metal objects (wires, stakes)

Table 4.6: Multi-class dataset: class definitions.

4.1.7 Whole-Image Dataset Construction and Experimental Design

Dataset Composition

The whole-image dataset is defined at the *(grape_id, week_date)* level, where each record corresponds to a single hyperspectral image of a grape cluster. Sample counts and class distributions for all experimental subsets are reported in Table 4.3.

Spatial Partitioning

A row-based spatial partitioning scheme was applied to ensure physically disjoint datasets. All records acquired from vineyard Row 1 were assigned to the calibration set, while records from Row 2 were reserved as a held-out evaluation set. The two rows are spatially separated within the experimental vineyard, minimizing spatial autocorrelation and preventing information leakage between sets.

Subset Usage Protocol

The calibration set (Row 1) was used for model training and for internal parameter calibration of the whole-image inference procedure. The held-out test set (Row 2) was used exclusively for final performance evaluation and was not involved in model training, parameter tuning, or threshold selection.

Training Data Definition

The whole-image pipeline used a pretrained multi-class pixel-level classifier. The pixel-level model was trained on spectral signatures (pixels) sampled from Late-stage images in vineyard Row 1 only. Early-stage records were excluded from pixel-level training.

Whole-image hyperparameter tuning was performed using Row 1 only. Tuned parameters included the spatial aggregation and post-processing settings used to convert pixel-level

scores into image-level outputs (e.g., patch aggregation thresholds, crack-pixel criteria, and blob filtering parameters such as minimum/maximum region area).

Inference on the Held-out Evaluation Set

After training and calibration, the finalized model and inference parameters were applied to the held-out evaluation set (Row 2) without modification.

Post-Inference Acquisition-Stage Stratification

Following inference, outputs were stratified by acquisition stage for reporting. Stage labels were assigned according to vineyard monitoring records:

1. **Late-stage:** Images acquired after the first recorded cracking event, representing an advanced cracking condition.
2. **Early-stage:** The earliest available image associated with the onset of cracking for the same cluster, representing the initial observable stage.

Both Early-stage and Late-stage results were generated using the same trained pixel-level model. Whole-image inference parameters were tuned separately for Early-stage and Late-stage using Row 1 and then applied to the corresponding stage-specific records in the held-out evaluation set (Row 2).

Row split → Train pixel-level model on pixels from Late-stage (Row 1) → Tune full-image aggregation hyperparameters on Row 1 (separately for Early/Late) → Freeze parameters → Inference on held-out evaluation set (Row 2) → Stage-specific reporting (Early/Late)

Figure 4.5: Protocol overview for the row-based split, training and calibration on Row 1 Late-stage data, held-out evaluation on Row 2, and post-inference reporting stratified by acquisition stage.

4.2 Modeling and Evaluation Methodology

This section describes the modeling framework and evaluation protocols employed to analyze hyperspectral data for grape crack detection and early-warning screening. The methodology is structured to progressively assess spectral discriminability, define appropriate problem formulations, and evaluate model generalization from pixel-level classification to whole-image detection, while maintaining reproducibility and methodological consistency across experiments.

4.2.1 Global Spectral Preprocessing

All hyperspectral data were subjected to a unified and fixed preprocessing pipeline prior to any modeling or evaluation. This preprocessing was applied consistently across all datasets, models, and experimental settings to ensure methodological uniformity, reproducibility, and fair comparability between experiments **thenkabail2016hyperspectral**.

Spectral Range Selection

To avoid the inclusion of spectral regions characterized by consistently low signal-to-noise ratios, the raw spectral signatures were restricted to a predefined wavelength range of 450–925 nm. Preliminary analysis of the data revealed high variance at the spectral extremities (400–450 nm and 925–1000 nm). This observation aligns with previous performance evaluations of the Specim IQ camera; notably, Behmann et al. **behmann2018specim** reported increased noise levels in the 400–450 nm range and significant reflectance artifacts in the near-infrared region (specifically above 925 nm) under outdoor conditions. Consequently, this effective operating range was selected once and retained throughout the study to avoid experiment-specific preprocessing decisions and to ensure consistency across all analyses **bioucas2013hyperspectral**.

This wavelength window was treated as a *global* design choice rather than a tunable hyperparameter: it was fixed prior to model development and kept identical for all cross-validation folds, feature-selection experiments, and final held-out tests.

Quality Control and Outlier Handling

Hyperspectral datacubes were analyzed in normalized reflectance units, where spectral values are expected to lie within the physically valid range $[0, 1]$. Due to field acquisition conditions and imperfect radiometric normalization, isolated violations of this range may occur. Isolated deviations from this range were retained and handled through the global SNV normalization procedure, which mitigates multiplicative and additive effects without discarding spectra.

Per-Spectrum Normalization

Following quality control, each remaining spectral vector was normalized independently using *Standard Normal Variate* (SNV) on a per-pixel basis. In this transformation, each spectrum was mean-centered and scaled by its standard deviation to reduce additive and multiplicative effects associated with illumination variability, surface geometry, and scattering, while preserving relative spectral shape. SNV is widely used in close-range hyperspectral imaging for reducing non-chemical variability without distorting discriminative spectral features **barnes1989snv**; **mishra2017close**; **spath2024separating**.

4.2.2 Cross-Validation and Data Partitioning Strategy

To assess generalization to unseen samples, Leave-One-Group-Out (LOGO) cross-validation was applied at the grape-cluster acquisition level. LOGO was selected due to the limited number of available hyperspectral images, which precluded a simple held-out test set while still requiring rigorous evaluation of cluster-level generalization. All pixels extracted from the same hyperspectral image (corresponding to a unique grape cluster) were assigned exclusively to either the training set or the test set within each fold. This grouping strategy ensured that evaluation was performed on spectral signatures from previously unseen clusters, preventing spatial or spectral leakage between partitions.

For multi-class settings with explicit background/distractor classes, a domain-aware variant was used: LOGO folds were defined over grape-containing acquisitions, while non-grape background samples were assigned to a fixed grouped holdout split (80/20) to prevent leakage across semantically distinct domains.

Across all reported protocols, leakage prevention was enforced by design: the wavelength restriction (450–925 nm) is fixed globally, SNV is computed independently per spectrum, and any data-dependent operation that requires fitting parameters or selecting features is carried out using only the training partition of the relevant protocol (or Row 1 for whole-image calibration) and then applied to held-out groups (or Row 2) without adaptation.

4.2.3 Spectral Discriminability and Informative Wavelength Analysis

Prior to model development, a wavelength-wise separability assessment was conducted to verify that cracked and healthy grape tissues exhibit distinguishable spectral signatures and to localize the spectral regions contributing to class contrast across the VIS–NIR range.

Separability was quantified directly from the pixel-level spectra using three model-agnostic metrics computed at each wavelength: absolute mean difference $|\Delta\mu|$, Cohen's d effect size, and a Fisher-style discriminant ratio. These measures capture both effect magnitude and variance-normalized class separation.

The analysis was used solely to characterize the spectral distribution of cracked–healthy contrast and to support interpretation of the spectral-signature trends reported in the Results. No wavelength subset was selected based on these metrics.

4.2.4 Pixel-Level Classification Framework

Can cracked and healthy grape tissues be spectrally distinguished at the individual pixel level?

Pixel-level classification was performed to evaluate whether spectral signatures of individual pixels could discriminate cracked grape tissue from healthy tissue and background materials. Each pixel was treated as an independent observation represented by its SNV-normalized reflectance spectrum.

Problem Formulation and Class Definitions

Two classification settings were evaluated:

- **Binary classification:** Pixels were assigned to one of two classes—*Healthy* (negative class, y=0) or *Cracked* (positive class, y=1)—based on visual assessment of the corresponding berry tissue during manual annotation.
- **Multi-class classification:** The binary formulation was extended to include explicit background classes representing non-berry materials commonly present in vineyard scenes (leaves, branches, soil, and artificial supports). Three multi-class evaluation modes were investigated:
  - *CRACK VS REST*: Binary formulation treating *Cracked* as positive.
  - *3-CLASSES*: Grouping samples into *Healthy*, *Cracked*, and *not-grape*.
  - *FULL MULTI-CLASS*: Preserving all 10 original class labels.

Feature Representation and Model Training

Each pixel was represented by its SNV-normalized reflectance vector across the 450–925 nm spectral range (159 bands at ∼3 nm spacing). No wavelength selection or feature reduction was applied prior to classification; the full spectral signature was used as input to enable assessment of intrinsic separability without imposed constraints.

The following supervised learning algorithms were evaluated:

- Partial Least Squares Discriminant Analysis (PLS-DA)
- Logistic Regression with L1 regularization
- Support Vector Machine with RBF kernel (SVM-RBF)
- Random Forest
- Gradient-boosted decision trees (XGBoost, used as baseline)
- Shallow multi-layer perceptron (MLP)

XGBoost was adopted as the baseline classifier for all primary comparisons due to its robustness to high-dimensional correlated features.

Both balanced and unbalanced training regimes were investigated. Under the unbalanced regime, the natural class distribution was preserved. Under the balanced regime, all classes were undersampled to match the minority class count. For binary classification, majority-class samples were randomly undersampled. For multi-class settings, segment-proportional undersampling was applied: each class was capped at the *Cracked* class count, with samples drawn proportionally from each segment (mask region) to preserve

within-class diversity. Models were trained using default or lightly tuned hyperparameters; no extensive hyperparameter search was performed.

Cross-Validation Protocol

Generalization was evaluated using the Leave-One-Group-Out (LOGO) protocol at the grape-cluster acquisition level, as described in §4.2.2. For the multi-class formulation, the domain-aware variant was applied.

Evaluation Metrics

Classification performance was assessed using the metrics and reporting conventions detailed in §4.2.8. The primary ranking metric was **PR–AUC** for the *Cracked* class; supplementary metrics included Accuracy, Balanced Accuracy, Precision, Recall, F1, and ROC–AUC.

4.2.5 Anomaly Detection Framework

Can crack damage be detected as a spectral anomaly without explicit class labels?

One-Class Formulation

An anomaly detection framework was implemented at the pixel level using a one-class learning paradigm. Models were trained exclusively on spectral signatures designated as *normal*, and unseen samples were flagged as anomalous based on deviation from this learned distribution.

Definition of Normality per Variant

Three experimental configurations were evaluated, each defining normality differently:

- **Binary one-class:** Normal class comprised Healthy/Regular pixels; anomalies were Cracked pixels.
- **Multi-class (all except Cracked):** Normal class comprised all nine non-*Cracked* classes; anomalies were *Cracked* pixels.
- **Multi-class (Cracked only):** Normal class comprised *Cracked* pixels exclusively; anomalies were all other classes.

Data Partitioning and Leakage Control

Cross-validation strategies were adapted to each variant's data structure:

- **Binary one-class:** Leave-One-Group-Out (LOGO) at the grape-cluster acquisition level (§4.2.2).

- **Multi-class (all except Cracked):** Single grouped train/test split (80/20) stratified by acquisition.
- **Multi-class (Cracked only):** Domain-aware cross-validation combining LOGO over *Cracked*-containing acquisitions with a fixed grouped 80/20 split for non-*Cracked* samples to prevent segment leakage.

For binary one-class IsolationForest (contamination = 0.05) was applied to remove outliers from the training normal set prior to autoencoder fitting.

Spectral Preprocessing

Pixels were preprocessed as described in §4.2.1 (SNV normalization, 450–925 nm range, 159 bands).

Autoencoder Architecture and Training

All anomaly detection experiments employed a fully connected autoencoder with symmetric encoder–decoder structure. The encoder applied ReLU activations, BatchNorm, and dropout ($p = 0.2$) in the first two layers. Several bottleneck sizes were evaluated.

Let $x \in \mathbb{R}^D$ denote a preprocessed spectrum and $\hat{x}$ its reconstruction. Training minimized $\mathrm{MSE}(x, \hat{x})$, with hyperparameters listed in Table 4.7.

Parameter	**Value**
Optimizer	Adam
Learning rate	10^{-3}
Weight decay	10^{-5}
LR scheduler	ReduceLROnPlateau (patience=5, factor=0.5)
Early stopping	patience=15
Max epochs	100
Batch size	256

Table 4.7: Autoencoder training hyperparameters.

Anomaly Scoring and Thresholding

The anomaly score was defined as the per-sample reconstruction mean squared error:

$$e(x) = \frac{1}{D} \|x - \hat{x}\|_2^2. \tag{4.1}$$

For each fold or split, an operating threshold τ was computed as the 99th percentile of reconstruction errors on the training partition of the normal class. A sample was classified as anomalous if $e(x) > \tau$.

Integration into Whole-Image Inference

For full-image inference, the multi-class autoencoder variant trained on CRACK samples only was selected and exported. This model was integrated into the full-image inference pipeline, transforming reconstruction errors into probability-like crack scores via sigmoid mapping:

$$p_{\text{CRACK}}(x) = \frac{1}{1 + \exp\left(\frac{e(x)-\tau}{\alpha}\right)}, \tag{4.2}$$

where $e(x)$ is the reconstruction error for spectrum x, τ is the training threshold (99th percentile of CRACK class errors), and α is a temperature parameter ($\approx 0.3\tau$) controlling transition sharpness. The temperature parameter α was fixed and not optimized. This mapping produces a smooth S-curve: low reconstruction errors (CRACK-like) map to high probability, while high errors (anomalous) map to low probability. Per-pixel scores were subsequently processed through standard post-processing and aggregation stages.

4.2.6 Whole-Image Detection and Aggregation Pipeline

How can spatial context stabilize noisy pixel-level outputs into reliable image-level decisions?

The whole-image pipeline aggregates pixel-wise crack probability maps produced by the trained multi-class pixel-level classifier into image-level decisions. All upstream pixel-level models were fixed at their trained states and were not retrained or tuned at the image level.

Pipeline Architecture and Spatial Aggregation

The whole-image aggregation pipeline follows a sequential process, illustrated in Figure 4.5, and consists of the following stages:

- **Input Hyperspectral Image:** Each input consists of a full hyperspectral datacube ($H \times W \times B$) acquired under field conditions.
- **Pixel-Level Probability Map Generation:** The trained multi-class pixel-level classifier produces a dense pixel-wise crack probability map for the *CRACK* class, with values in the range $[0, 1]$.
- **Morphological Closing and Region Formation:** Spatially adjacent pixels with elevated crack probabilities are grouped into connected regions (BLOBs) using morphological closing operations, enforcing spatial coherence.
- **Blob Size Filtering:** Connected regions were filtered using minimum and maximum area constraints.

- **Patch-Based Division and Decision:** The image is divided into fixed-size patches. A patch is marked as crack-positive if the proportion of crack-related pixels within the patch exceeds a predefined threshold.
- **Image-Level Decision:** The final image-level decision is obtained by aggregating patch-level outcomes. An image is classified as *Cracked* if at least one patch is identified as crack-positive; otherwise, it is classified as *Non-Cracked*.

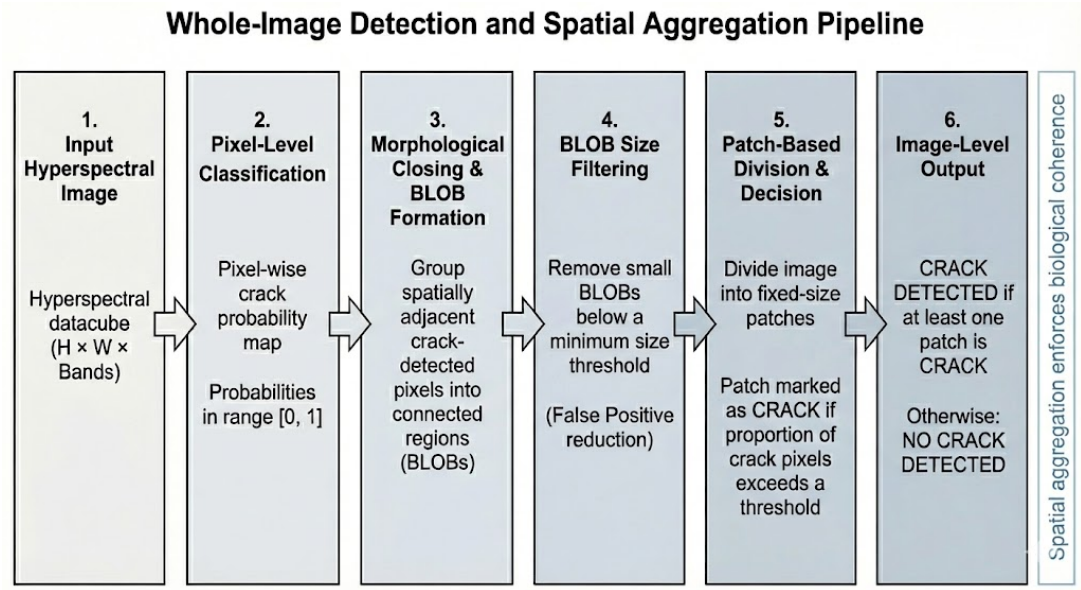


Figure 4.6: Whole-image crack-detection pipeline from pixel-level CRACK probabilities to image-level classification via spatial aggregation.

Whole-Image Dataset and Evaluation Setup

Whole-image evaluation was conducted on the row-partitioned inference dataset described in Section 4.1.7, following the spatial partitioning and subset usage protocol defined therein (Row 1 calibration, Row 2 held-out test). Exact sample counts and class balance (early/late) are reported in Table 4.3.

Evaluation Across Disease Progression Stages

To assess robustness across different acquisition stages, whole-image performance was evaluated separately for early-stage and late-stage samples. These labels refer exclusively to the crack-progression state at the time of measurement, not to temporal forecasting. Late-stage images contained clusters with clearly visible cracks, whereas early-stage images corresponded to initial onset.

4.2.7 Wavelength Selection

How many wavelengths are needed to maintain crack-detection performance?

This stage evaluates whether a reduced subset of wavelengths can maintain crack-detection performance while preserving physical interpretability for potential multispectral sensor design. BFS was applied exclusively at the pixel level; whole-image inference used pixel-level models trained on either the full or reduced spectral representation without additional wavelength adaptation at the image level.

Wavelength selection was evaluated separately for binary (*Cracked* vs. *Healthy*) and multi-class (with background classes) settings, as formulated in §4.2.4.

Spectral Preprocessing

Pixels were preprocessed as described in §4.2.1 (SNV normalization, 450–925 nm range, 159 bands).

Classification Model

XGBoost was used as the fixed classifier for all wavelength selection experiments, consistent with the baseline model adopted in the pixel-level classification framework (§4.2.4).

Two Experimental Settings

Wavelength selection was performed in two independent experiments:

- **Binary setting:** Wavelengths selected for *Cracked* vs. *Healthy* classification.
- **Multi-class setting:** Wavelengths selected for multi-class classification with background classes.

Train-Test Split Protocol

A single grouped 80/20 hold-out split was defined to prevent leakage, with grouping at the grape-cluster level in the binary setting and at the cluster level for background classes in the multi-class setting. At each BFS iteration, XGBoost was re-trained on the 80% training partition, while the held-out 20% test partition remained fixed and served as the invariant selection criterion.

Backward Feature Selection (BFS)

A deterministic wrapper-based backward feature selection (BFS) procedure was applied, starting from the full set of 159 wavelengths and iteratively removing the least informative band at each step. At each iteration, XGBoost was re-trained on the 80% training partition and evaluated on the fixed 20% hold-out set, using PR–AUC–CRACK as the

selection criterion. This procedure yielded a monotonic elimination path and identified the minimal wavelength subset that preserved near-baseline performance.

To assess the stability of the selected wavelength subsets, the BFS procedure was repeated five times on the same fixed grouped 80/20 train–test split (split seed = 42), varying only the random seed supplied to XGBoost (model seeds 1–5). The data partition and grouping remained invariant across runs.

4.2.8 Evaluation Protocols and Performance Metrics

Model training and model-selection procedures were evaluated using PR–$\mathrm{AUC}_{\mathrm{CRACK}}$ as the primary ranking metric across all pixel-level classification, anomaly detection, and wavelength-selection experiments, reflecting the strong class imbalance at the pixel level. Whole-image crack detection was evaluated using $\mathrm{F1}_{\mathrm{CRACK}}$ as the primary system-level metric. Unless explicitly subscripted, reported F1 refers to the weighted F1-score.

4.2.9 Reproducibility and Implementation Details

All experiments were executed in Python using fixed random seeds (default seed = 42) to ensure deterministic data splits and model training. Cross-validation partitions and held-out test sets were kept invariant across runs. Experiment outputs were stored in timestamped directories together with configuration files and cross-validation manifests to enable exact reruns of the reported pipelines.

For wavelength selection (BFS) stability analysis, the primary BFS run employed model seed = 42 and split seed = 42. Five additional repetitions used the same split seed (42) but varied model seeds (1–5), allowing quantification of wavelength selection robustness with respect to XGBoost training randomness alone, independent of data splitting variability.

Chapter 5

Experimental Results

5.1 Overview of Experimental Results

This chapter reports quantitative results following the experimental pipeline described in Section 4.2: spectral characterization, pixel-level classification, whole-image detection, and wavelength-selection analysis.

5.1.1 Baselines

Each set of results is reported relative to the explicit baseline defined below:

- **Pixel-level classification baseline.** The baseline model is **XGBoost** trained on the **full SNV-normalized hyperspectral signature** (450–925 nm, 159 bands; see Section 4.2.1). Binary-task performance is reported in the XGBoost rows of Table 5.1. Results for extended labeling and balancing regimes appear in Table 5.2, with balanced-training baselines provided in Appendix A.

- **Whole-image crack-detection baseline.** Whole-image detection was evaluated using the aggregation pipeline described in Section 4.2.6, applied to pixel-level probability maps from the full-spectrum model without wavelength selection. Early/late calibration and held-out test performance are summarized in Table 5.7, with dataset composition reported in Table 5.6.

- **BFS wavelength-selection reference.** BFS performance is reported as a function of the **number of retained wavelengths** relative to the **full-spectrum baseline**. Elimination trajectories are shown in Figures 5.3 and 5.4, with derived wavelength-count thresholds summarized in Table 5.9. Whole-image test-set results at representative wavelength counts, including the 159-band baseline, are reported in Table 5.10.

5.2 Spectral Separability and Informative Wavelengths

Per-band spectral separability between healthy and cracked tissue was assessed prior to modeling, addressing Hypothesis 1. Preprocessing followed Section 4.2.1.

5.2.1 Mean Spectral Signatures of Healthy and Cracked Tissue

Mean spectral signatures ($\pm$SD) for healthy and cracked tissue are shown in Figure 5.1.

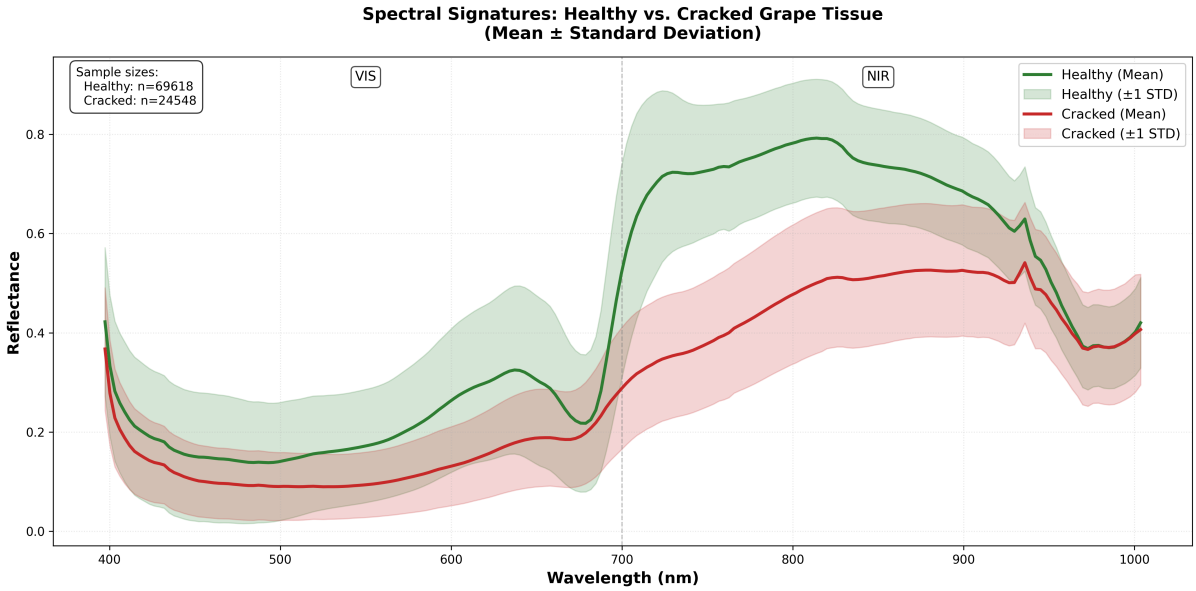


Figure 5.1: Mean spectral signatures ($\pm$ standard deviation) for healthy and cracked tissue across the VIS–NIR range.

5.2.2 Spectral Separability Across Wavelength Regions

Cohen's d **cohen1988statistical** ($d = |\mu_1 - \mu_2| / \sigma_{\text{pooled}}$) and Fisher Score **duda2001pattern** (inter-class to intra-class variance ratio) were computed per band; results are shown in Figure 5.2.

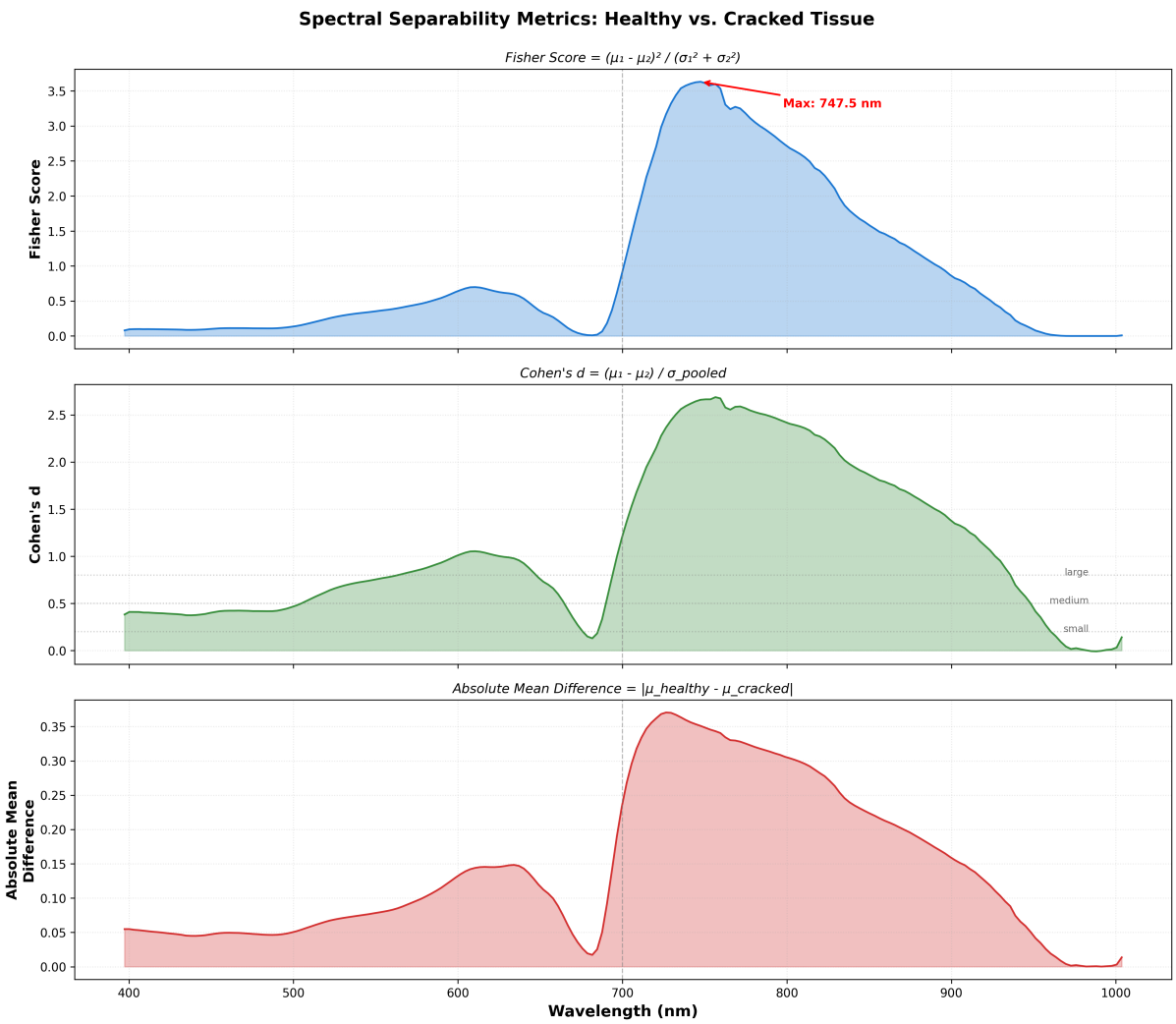


Figure 5.2: Wavelength-wise spectral separability metrics between healthy and cracked tissue, including absolute mean difference, Cohen's d, and Fisher score.

Separation is substantially higher in the NIR, with peak discriminability at 727–757 nm. Regional VIS–NIR summaries are in Appendix A.2.

5.3 Pixel-Level Binary Classification

Supervised pixel-level classification addresses Hypothesis 2, following the LOGO cross-validation protocol described in Section 4.2.2. PR–AUC$_{\text{CRACK}}$ is the primary metric.

Table 5.1 reports performance across LOGO folds (Section 4.2.2; unbalanced class distribution). Balanced-training results are in Appendix A.

Table 5.1: Pixel-level classification metrics (unbalanced, binary setting). Truncated mean values. Full cross-validation statistics in Appendix A.1.

Model	Accuracy	Precision (Cracked)	Recall (Cracked)	Weighted F1	ROC–AUC (Cracked)
PLS-DA	0.992	0.999	0.968	0.992	1.000
Logistic Regression (L1)	0.996	0.989	0.994	0.996	0.999
SVM (RBF)	0.993	0.994	0.978	0.993	0.999
Random Forest	0.996	0.996	0.990	0.996	0.999
XGBoost	0.994	0.995	0.979	0.994	1.000

All five models achieve PR–AUC$_{\text{CRACK}} \geq 0.987$ (Table 5.1), supporting Hypothesis 2.

5.4 Pixel-Level Multi-Class Classification

Binary detection was extended to three labeling strategies (*crack_regular_rest*, *crack_vs_rest*, *multi_class*) as specified in Section 4.2; PR–AUC$_{\text{CRACK}}$ is the primary metric. Balanced-training results are in Appendix A.

Table 5.2: Pixel-level classification metrics (unbalanced, multi-class). Truncated mean values. Full cross-validation statistics in Appendix A.1.

Label setup	Model	Accuracy	Balanced Acc.	Precision (Cracked)	Recall (Cracked)	Macro-F1	ROC–AUC	PR–AUC
crack_regular_rest								
	PLS-DA	0.508	0.687	0.882	0.753	0.581	0.997	0.863
	Logistic Regression (L1)	0.853	0.868	0.965	0.966	0.801	0.999	0.983
	Random Forest	0.917	0.917	0.894	0.939	0.890	0.997	0.962
	XGBoost	0.921	0.916	0.980	0.949	0.883	0.999	0.990
	MLP (Small)	0.917	0.916	0.953	0.972	0.875	0.999	0.991
crack_vs_rest								
	PLS-DA	0.512	0.694	0.884	0.762	0.589	0.997	0.866
	Logistic Regression (L1)	0.857	0.870	0.966	0.965	0.804	0.999	0.984
	Random Forest	0.914	0.915	0.887	0.936	0.885	0.997	0.961
	XGBoost	0.920	0.915	0.980	0.950	0.882	0.999	0.990
	MLP (Small)	0.916	0.915	0.952	0.973	0.874	0.999	0.991
multi_class								
	PLS-DA	0.967	0.978	0.720	0.992	0.901	0.997	0.967
	Logistic Regression (L1)	0.997	0.995	0.967	0.997	0.992	0.999	0.997
	Random Forest	0.996	0.994	0.967	0.996	0.990	0.999	0.997
	XGBoost	0.997	0.994	0.981	0.990	0.992	0.999	0.997
	MLP (Small)	0.990	0.989	0.924	0.986	0.969	0.996	0.989

The *multi_class* formulation yields the highest PR–AUC$_{\text{CRACK}}$ (≥ 0.967); LR, RF, and XGBoost all exceed 0.99 (Table 5.2).

5.5 Autoencoder-Based Anomaly Detection

In the one-class binary setting, the autoencoder achieved PR–AUC$_{\text{CRACK}}$ ranging from 0.925 to 0.970 (Table 5.3).

When trained on CRACK pixels as the normal class, PR–AUC$_{\text{CRACK}} \approx 1.0$ (Table 5.4), indicating trivial reconstruction of CRACK signatures. Training on all non-CRACK classes with CRACK defined as the anomaly yielded PR–AUC$_{\text{CRACK}} \approx 0.01$ (Table 5.5).

Supervised models outperform autoencoders for CRACK detection. Additional autoencoder formulations are reported in Appendix A.1.

Table 5.3: Pixel-level autoencoder anomaly detection (binary setting). Truncated mean values. Full statistics in Appendix A.1.

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)
Autoencoder (64–32–16)	0.896	0.822	0.997	0.901	0.981	0.955
Autoencoder (128–64–32)	0.859	0.754	0.997	0.862	0.967	0.925
Autoencoder (64–32–8)	0.892	0.812	0.997	0.896	0.979	0.951

Table 5.4: Pixel-level autoencoder anomaly detection (crack-only training). Truncated mean values. Full statistics in Appendix A.1.

Architecture	Accuracy	Precision (non-CRACK)	Recall (non-CRACK)	F1 (non-CRACK)	ROC–AUC (non-CRACK)	PR–AUC (non-CRACK)
Autoencoder (64–32–16)	0.966	0.999	0.966	0.982	0.985	0.999
Autoencoder (128–64–32)	0.986	0.999	0.986	0.993	0.984	0.999
Autoencoder (64–32–8)	0.962	0.999	0.963	0.980	0.983	0.999

Table 5.5: Pixel-level autoencoder anomaly detection (all non-CRACK training). Truncated mean values. Full statistics in Appendix A.1.

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)
Autoencoder (64–32–16)	0.996	0.009	0.111	0.017	0.457	0.008
Autoencoder (128–64–32)	0.998	0.011	0.012	0.012	0.451	0.008
Autoencoder (64–32–8)	0.998	0.020	0.031	0.025	0.454	0.011

Overall, autoencoder-based formulations yielded lower PR–AUC$_{\text{CRACK}}$ than the supervised XGBoost baseline and were therefore not adopted for subsequent pixel-level or whole-image detection stages.

5.6 Whole-Image Crack Detection

Using the protocol described in Section 4.2.6, whole-image detection was assessed under early- and late-stage conditions, addressing Hypothesis 3. The primary metric is $\mathbf{F1_{CRACK}}$. The baseline for whole-image comparison is the full-spectrum XGBoost aggregation pipeline (159 bands, 450–925 nm) without wavelength selection, as defined in Section 5.1.1. Dataset composition is in Table 5.6; supervised and autoencoder results are in Tables 5.7 and 5.8, respectively.

Table 5.6: Dataset composition for full-image crack-detection experiments under early- and late-stage acquisition conditions (initial/minimal vs. advanced cracking). Row 1 was used for calibration, while Row 2 served as an independent test set.

Row	Stage	Split	Total	Unique	Pos	Neg
Row 1	Early	Calibration	172	60	32	140
Row 1	Late	Calibration	160	60	35	125
Row 2	Early	Test	60	60	32	28
Row 2	Late	Test	60	60	34	26

Table 5.7: Full-image crack detection performance for early- and late-stage acquisition conditions. F1 (Cracked) is the primary metric; the Cracked class is positive.

Stage	Split	TP	FP	TN	FN	Acc	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	F2 (Cracked)	Specificity (Healthy)	NPV (Healthy)	MCC
Late	Calibration	34	5	120	1	0.963	0.966	0.872	0.971	0.919	0.950	0.960	0.992	0.897
Late	Test	27	5	21	7	0.800	0.801	0.844	0.794	0.818	0.804	0.808	0.750	0.598
Early	Calibration	23	6	134	9	0.913	0.838	0.793	0.719	0.754	0.732	0.957	0.937	0.703
Early	Test	19	3	25	13	0.733	0.743	0.864	0.594	0.704	0.633	0.893	0.658	0.504

On the independent test set (Table 5.7), under late-stage conditions the supervised pipeline achieves $\mathrm{F1}_{\mathrm{CRACK}} = 0.818$. Under early-stage conditions $\mathrm{F1}_{\mathrm{CRACK}} = 0.704$.

Table 5.8: Full-image crack detection using autoencoder-based anomaly detection. F1 (Cracked) is the primary metric; the Cracked class is positive.

Stage	Split	TP	FP	TN	FN	Acc	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	F2 (Cracked)	Specificity (Healthy)	NPV (Healthy)	MCC
Late	Calibration	26	7	118	9	0.900	0.843	0.788	0.743	0.765	0.751	0.944	0.929	0.702
Late	Test	22	7	19	12	0.683	0.689	0.759	0.647	0.698	0.667	0.731	0.613	0.375
Early	Calibration	19	17	123	13	0.826	0.736	0.528	0.594	0.559	0.579	0.879	0.904	0.452
Early	Test	13	13	15	19	0.467	0.471	0.500	0.406	0.448	0.422	0.536	0.441	−0.058

On the independent test set (Table 5.8), the autoencoder-based pipeline yields lower $\mathrm{F1}_{\mathrm{CRACK}}$ than the supervised classifier in both stages (Late: $\mathrm{F1}_{\mathrm{CRACK}} = 0.698$ vs. $\mathrm{F1}_{\mathrm{CRACK}} = 0.818$; Early: $\mathrm{F1}_{\mathrm{CRACK}} = 0.448$ vs. $\mathrm{F1}_{\mathrm{CRACK}} = 0.704$).

The effect of wavelength selection on the whole-image pipeline is evaluated in the following section using backward feature selection.

5.7 Wavelength Selection via Backward Feature Selection

BFS was applied using the protocol described in Section 4.2.7, optimizing PR–AUC$_{\text{CRACK}}$ on the whole-image pipeline. The full-spectrum baseline is 159 wavelengths (450–925 nm).

5.7.1 BFS Trajectories and Subset Selection

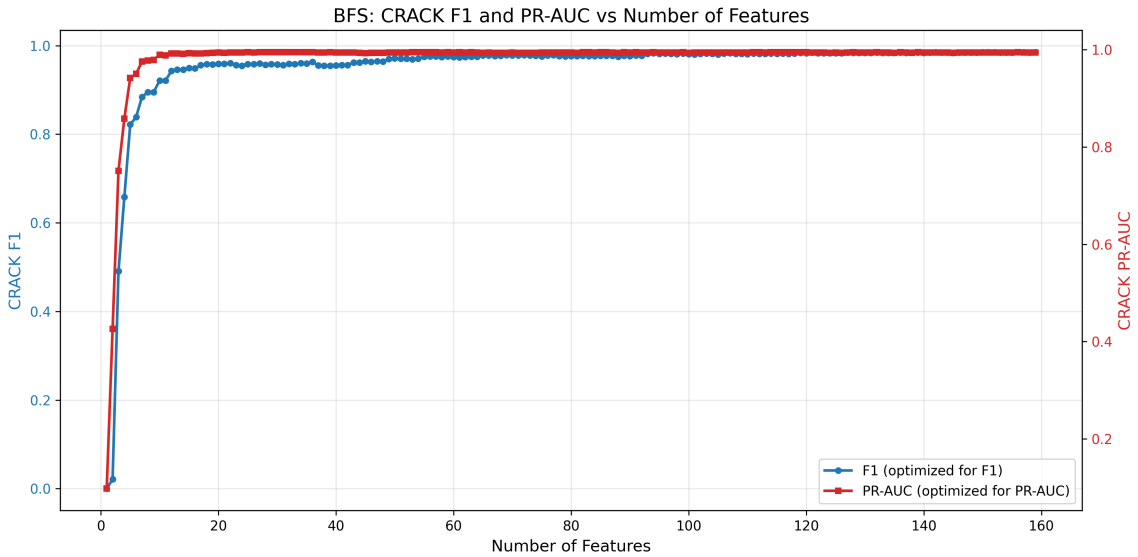
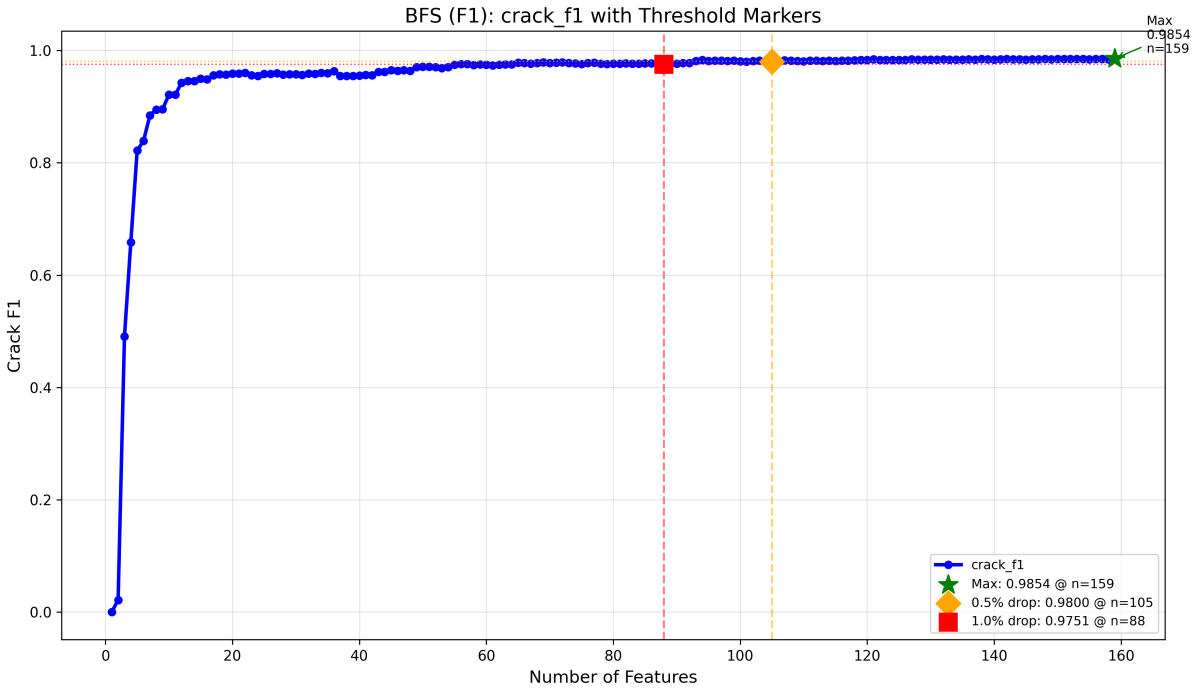


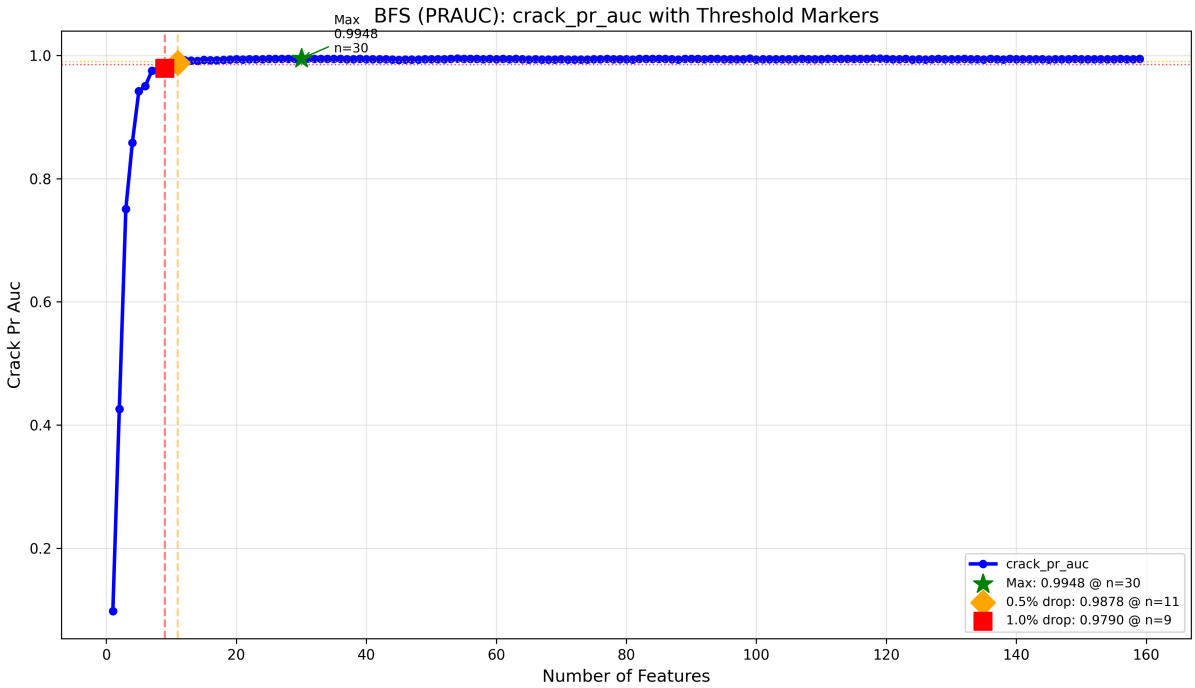
Figure 5.3: BFS trajectories for the whole-image pipeline: CRACK-class PR–AUC (primary objective) and CRACK-class F1 versus the number of retained wavelengths (n; full baseline n=159).

Table 5.9: Summary of BFS wavelength-count thresholds for CRACK metrics. (Full precision in Appendix A.6.1)

Metric	Max score	$n_{\max}$	0.5% drop: n	score	1.0% drop: n	score
CRACK PR–AUC	0.994	30	11	0.987	9	0.979
CRACK F1	0.985	159	105	0.980	88	0.975



(a) CRACK-class F1 vs. number of wavelengths (n), with markers at the maximum and at 0.5%/1.0% drop points.



(b) CRACK-class PR–AUC (primary BFS objective) vs. number of wavelengths (n), with the same threshold markers.

Figure 5.4: BFS threshold-marker visualization on CRACK metrics. Markers indicate the maximum and the first n where performance drops more than 0.5% and 1.0% below that maximum.

5.7.2 End-to-End Wavelength-Selection Results

Table 5.10 reports whole-image detection metrics for representative wavelength subsets, evaluated separately for early- and late-stage acquisitions on the held-out test set.

Table 5.10: Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions. (Full precision in Appendix A.6.2)

Stage	# Wavelengths	Accuracy	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	MCC
Early	9	0.783	0.790	0.880	0.687	0.771	0.587
Late	9	0.800	0.805	0.866	0.764	0.812	0.605
Early	11	0.750	0.757	0.840	0.656	0.736	0.519
Late	11	0.816	0.811	0.828	0.852	0.840	0.625
Early	30	**0.816**	**0.821**	**0.888**	**0.750**	**0.813**	**0.644**
Late	30	**0.833**	**0.834**	**0.875**	0.823	0.848	**0.665**
Early	159	0.800	0.803	0.857	**0.750**	0.800	0.607
Late	159	**0.833**	0.825	0.833	**0.882**	**0.857**	0.659

The top-30 subset achieves $F1_{CRACK} = 0.8136$ (Early) and 0.8485 (Late), matching the 159-band baseline in Early stage (0.8000) and approaching it in Late stage (0.8571); the top-11 subset yields $F1_{CRACK} = 0.7368$ (Early) and 0.8406 (Late). Additional metrics and generalization-gap analysis are provided in Appendix A.6.

5.7.3 BFS Wavelength-Selection Stability Analysis

Cross-seed wavelength overlap statistics are reported in Appendix A.6.

Wavelength Popularity and Stability – Top 30

The top-30 wavelength subsets demonstrate strong cross-seed consistency in their selection patterns, with dominant spectral regions remaining stable under coarser binning schemes. The top-30 subset exhibits substantially higher cross-seed stability compared to smaller subsets such as the top-11. The consistently selected regions are predominantly located in the near-infrared (NIR) portion of the spectrum. Figures 5.5–5.8 provide comprehensive visualization of this cross-seed wavelength stability.
Exact wavelength identities and cross-seed overlap statistics are provided in Appendix A.6.

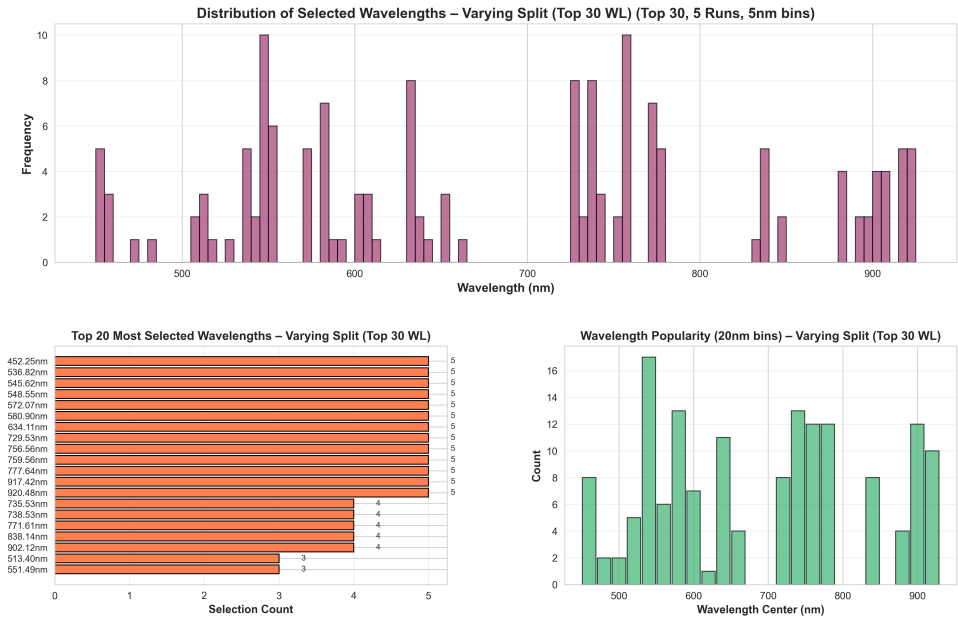


Figure 5.5: Distribution and popularity of the top-30 selected wavelengths across five varying-split seeds.

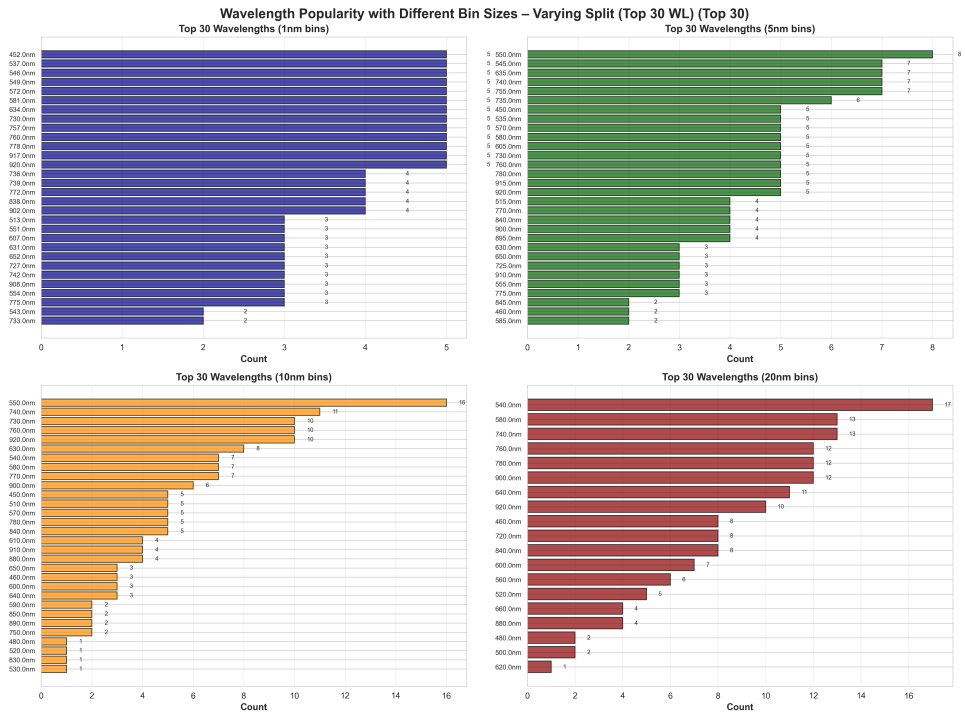


Figure 5.6: Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-30 subsets.

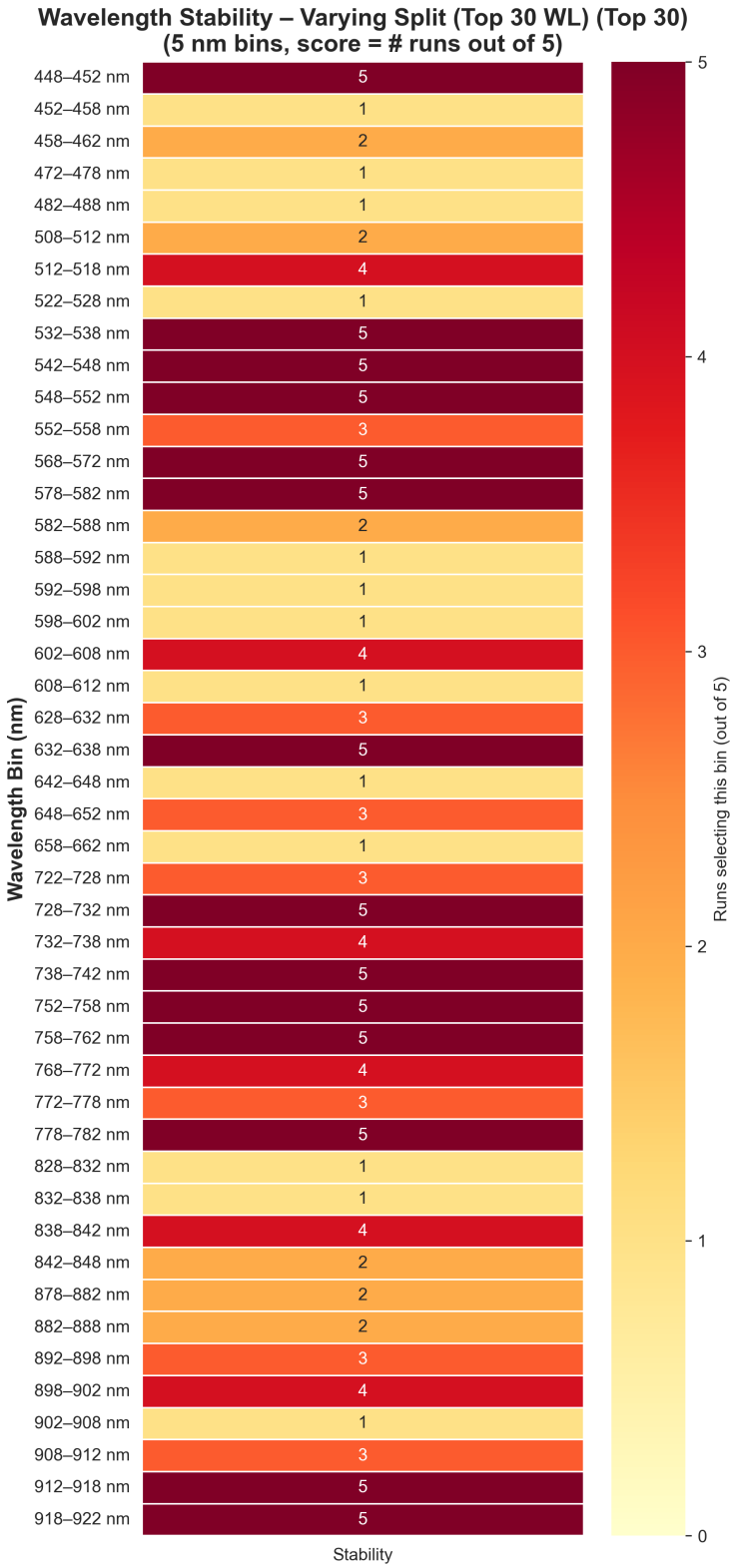


Figure 5.7: Stability heatmap for the top-30 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$). Each cell indicates how many seeds selected a wavelength within that 5 nm bin.

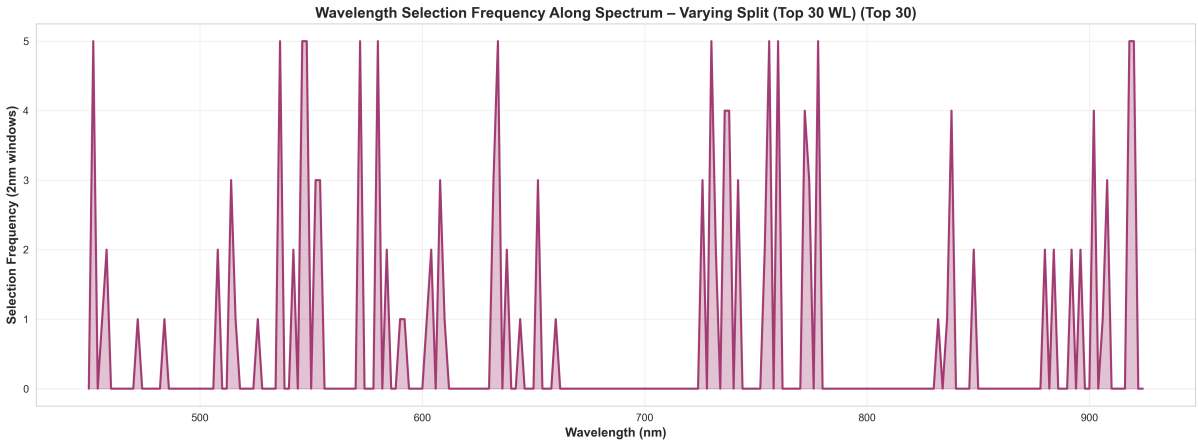


Figure 5.8: Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-30 subsets.

Wavelength Popularity and Stability – Top 11

The top-11 wavelength subsets (within 1% of peak PR–AUC$_{\text{CRACK}}$) exhibit lower cross-seed stability than the top-30 subset, a consequence of the stronger dimensionality reduction. Nevertheless, consistent spectral regions remain identifiable across seeds (Figures 5.9–5.12). Detailed overlap statistics are reported in Appendix A.6.

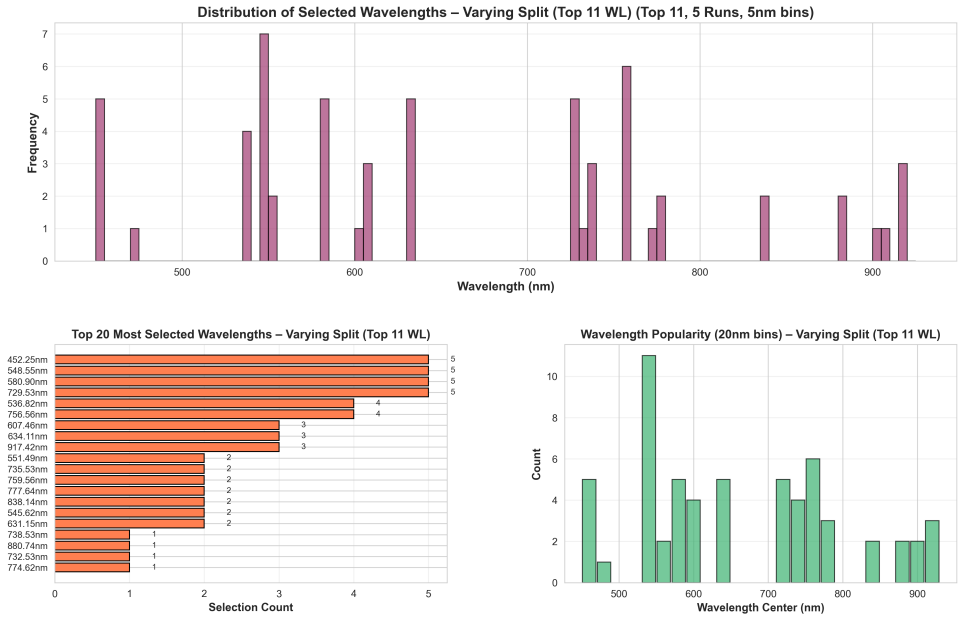


Figure 5.9: Distribution and popularity of the top-11 selected wavelengths across five varying-split seeds.

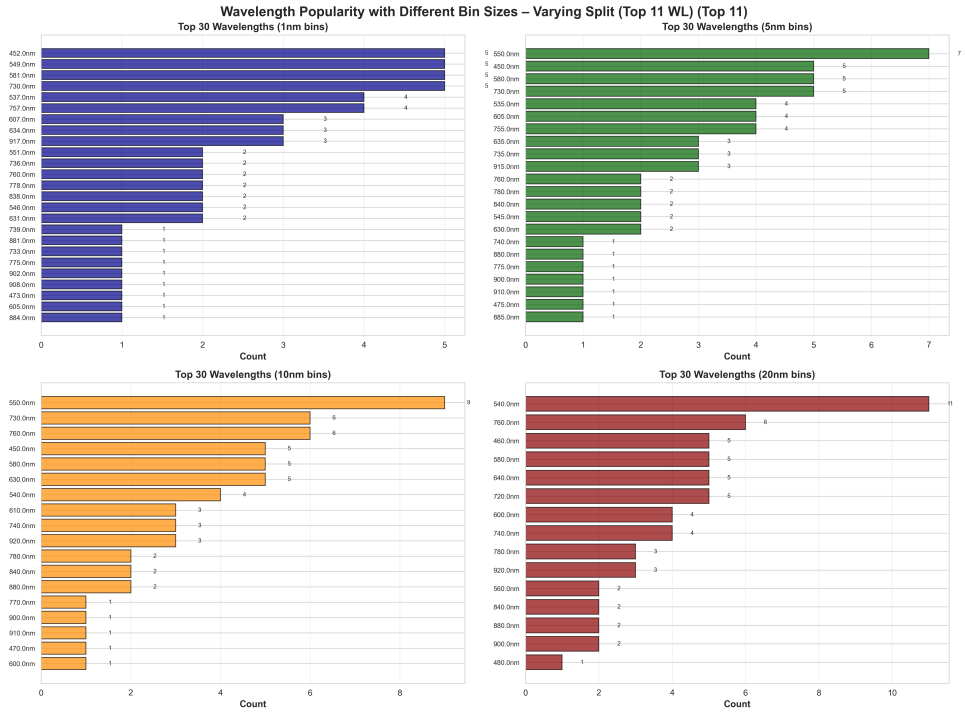


Figure 5.10: Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-11 subsets.

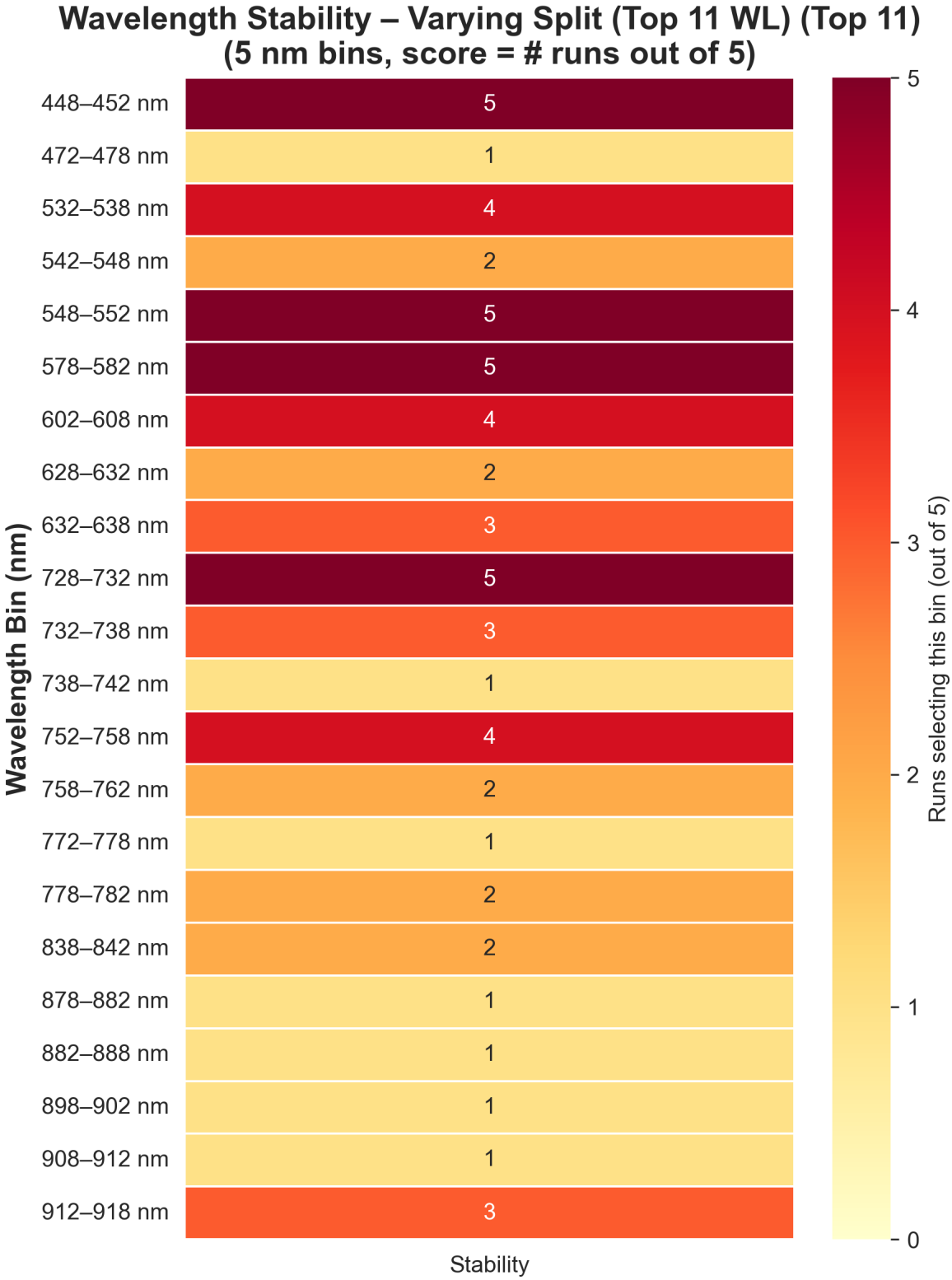


Figure 5.11: Stability heatmap for the top-11 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$; 0.5% drop setting). Each cell indicates how many seeds selected a wavelength within that 5 nm bin.

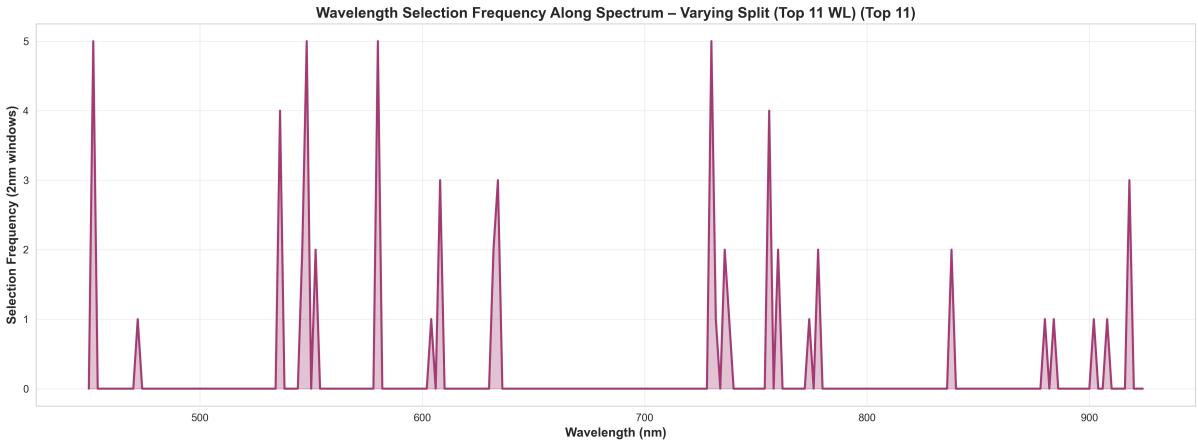


Figure 5.12: Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-11 subsets.

Appendix A

Supplementary Results

A.1 Cross-Validation Statistics and Full Precision Tables

This appendix contains full cross-validation results with complete precision and standard deviations for all experiments discussed in Chapter 5. The main text reports truncated mean values for clarity; unreduced values are provided below for reproducibility and detailed analysis.

A.1.1 Pixel-Level Classification: Binary and Multi-Class

Table A.1: Pixel-level classification metrics (unbalanced training set, binary). Values are reported as mean ± standard deviation across folds where available. Precision/recall and ROC–AUC refer to the *Cracked* class (positive class); F1 is the weighted F1-score.

Model	Accuracy	Precision (Cracked)	Recall (Cracked)	Weighted F1	ROC–AUC (Cracked)
PLS-DA	0.9927 ± 0.0168	0.9999	0.9682	0.9924	1.0000 ± 0.0000
Logistic Regression (L1)	0.9962 ± 0.0074	0.9890	0.9948	0.9963	0.9996 ± 0.0012
SVM (RBF)	0.9938 ± 0.0094	0.9943	0.9787	0.9938	0.9997 ± 0.0010
Random Forest	0.9968 ± 0.0055	0.9964	0.9904	0.9968	0.9990 ± 0.0035
XGBoost	0.9947 ± 0.0103	0.9953	0.9798	0.9946	1.0000 ± 0.0000

A.1.2 Autoencoder-Based Anomaly Detection

A.2 Regional VIS–NIR Reflectance Summaries

Cracked tissue shows lower reflectance than healthy tissue across the full VIS–NIR range, with the divergence most pronounced beyond 700 nm. Reflectance values were aggregated over two broad spectral regions (VIS and NIR); healthy tissue exceeds cracked tissue by 32% in the NIR (0.660 vs. 0.450) and 37% in the VIS (0.218 vs. 0.137).

Table A.2: Pixel-level classification metrics (**unbalanced** training set, multi-class). Values are reported as mean ± standard deviation across folds.

Label setup	Model	Accuracy	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	Macro-F1	ROC–AUC (Cracked)	PR–AUC (Cracked)
crack_regular_rest								
	PLS-DA	0.5087 ± 0.0044	0.6879 ± 0.0239	0.8825 ± 0.1174	0.7537 ± 0.2385	0.5819 ± 0.0199	0.9977 ± 0.0049	0.8630 ± 0.1242
	Logistic Regression (L1)	0.8537 ± 0.0021	0.8683 ± 0.0075	0.9654 ± 0.0185	0.9663 ± 0.0748	0.8010 ± 0.0049	0.9993 ± 0.0021	0.9839 ± 0.0116
	Random Forest	0.9175 ± 0.0056	0.9179 ± 0.0116	0.8942 ± 0.0644	0.9394 ± 0.0701	0.8904 ± 0.0106	0.9975 ± 0.0037	0.9622 ± 0.0495
	XGBoost	0.9212 ± 0.0040	0.9162 ± 0.0094	0.9804 ± 0.0118	0.9493 ± 0.0894	0.8832 ± 0.0074	0.9992 ± 0.0017	0.9902 ± 0.0160
	MLP (Small)	0.9175 ± 0.0059	0.9166 ± 0.0122	0.9530 ± 0.0272	0.9729 ± 0.0398	0.8759 ± 0.0176	0.9994 ± 0.0017	0.9911 ± 0.0122
crack_vs_rest								
	PLS-DA	0.5120 ± 0.0047	0.6940 ± 0.0189	0.8843 ± 0.1117	0.7623 ± 0.2303	0.5893 ± 0.0164	0.9979 ± 0.0045	0.8669 ± 0.1186
	Logistic Regression (L1)	0.8574 ± 0.0022	0.8704 ± 0.0076	0.9669 ± 0.0190	0.9651 ± 0.0759	0.8049 ± 0.0050	0.9993 ± 0.0022	0.9847 ± 0.0114
	Random Forest	0.9147 ± 0.0060	0.9159 ± 0.0126	0.8871 ± 0.0701	0.9368 ± 0.0751	0.8857 ± 0.0116	0.9975 ± 0.0041	0.9610 ± 0.0535
	XGBoost	0.9200 ± 0.0043	0.9154 ± 0.0098	0.9809 ± 0.0120	0.9508 ± 0.0888	0.8825 ± 0.0078	0.9992 ± 0.0018	0.9900 ± 0.0165
	MLP (Small)	0.9163 ± 0.0062	0.9158 ± 0.0129	0.9521 ± 0.0282	0.9736 ± 0.0390	0.8741 ± 0.0188	0.9994 ± 0.0017	0.9912 ± 0.0122
multi_class								
	PLS-DA	0.9670 ± 0.0211	0.9787 ± 0.0121	0.7204 ± 0.1749	0.9927 ± 0.0127	0.9013 ± 0.0709	0.9978 ± 0.0025	0.9677 ± 0.0372
	Logistic Regression (L1)	0.9972 ± 0.0031	0.9959 ± 0.0097	0.9675 ± 0.0225	0.9977 ± 0.0078	0.9920 ± 0.0085	0.9996 ± 0.0016	0.9971 ± 0.0079
	Random Forest	0.9965 ± 0.0030	0.9940 ± 0.0120	0.9675 ± 0.0234	0.9969 ± 0.0090	0.9906 ± 0.0100	0.9994 ± 0.0026	0.9970 ± 0.0077
	XGBoost	0.9978 ± 0.0021	0.9946 ± 0.0103	0.9812 ± 0.0149	0.9908 ± 0.0206	0.9923 ± 0.0074	0.9995 ± 0.0021	0.9977 ± 0.0069
	MLP (Small)	0.9905 ± 0.0249	0.9890 ± 0.0224	0.9244 ± 0.1648	0.9866 ± 0.0259	0.9699 ± 0.0704	0.9965 ± 0.0083	0.9898 ± 0.0247

Table A.3: Pixel-level autoencoder anomaly detection: binary setting (non-cracked as normal class). Values are reported as mean ± standard deviation across LOGO folds.

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.8969 ± 0.1646	0.8226 ± 0.1925	0.9971 ± 0.0016	0.9017 ± 0.1047	0.9814 ± 0.0327	0.9554 ± 0.0900	63.34 ± 2.80	0.043 ± 0.004
Autoencoder (128–64–32)	0.8597 ± 0.1813	0.7544 ± 0.2308	0.9974 ± 0.0015	0.8628 ± 0.1270	0.9673 ± 0.0469	0.9253 ± 0.1190	60.90 ± 1.48	0.057 ± 0.001
Autoencoder (64–32–8)	0.8923 ± 0.1648	0.8126 ± 0.1938	0.9978 ± 0.0013	0.8965 ± 0.1050	0.9799 ± 0.0334	0.9511 ± 0.0913	61.02 ± 1.46	0.041 ± 0.001

Table A.4: Pixel-level autoencoder anomaly detection: trained on CRACK pixels (multi-class setting). Values are mean across LOGO folds.

Architecture	Accuracy	Precision (non-CRACK)	Recall (non-CRACK)	F1 (non-CRACK)	ROC–AUC (non-CRACK)	PR–AUC (non-CRACK)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.9660	0.9995	0.9662	0.9826	0.9858	0.9999	26.96	0.148
Autoencoder (128–64–32)	0.9861	0.9991	0.9869	0.9930	0.9846	0.9999	27.12	0.207
Autoencoder (64–32–8)	0.9629	0.9994	0.9632	0.9809	0.9834	0.9999	26.20	0.155

Table A.5: Pixel-level autoencoder anomaly detection: trained on all non-CRACK classes (multi-class setting). Single group-wise train–test split.

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.9969	0.0096	0.1110	0.0177	0.4575	0.0083	1429.22	0.388
Autoencoder (128–64–32)	0.9986	0.0118	0.0126	0.0122	0.4513	0.0084	999.73	0.255
Autoencoder (64–32–8)	0.9984	0.0207	0.0316	0.0250	0.4542	0.0110	916.85	0.141

Table A.6: Summary statistics of reflectance values across spectral regions for healthy and cracked tissue.

Spectral Region	Class	Mean Reflectance	Std. Dev.	Median
VIS (400–700 nm)	Healthy	0.218	0.128	0.186
	Cracked	0.137	0.076	0.123
NIR (700–1000 nm)	Healthy	0.660	0.114	0.630
	Cracked	0.450	0.121	0.425

Table A.7: Regional spectral separability metrics between healthy and cracked tissue. Mean values are averaged across wavelengths within each range, while peak values indicate the maximum separability and the corresponding wavelength.

Metric	VIS Range	NIR Range	Full Spectrum
Mean $\lvert\Delta\mu\rvert$ (Abs. Mean Diff.)	0.0810	0.2101	0.1422
Max $\lvert\Delta\mu\rvert$ (nm)	0.2325 (708.6)	0.3704 (726.5)	0.3704 (726.5)
Mean Cohen's d	0.6330	1.6217	1.1014
Max Cohen's d (nm)	1.1910 (699.5)	2.6877 (756.6)	2.6877 (756.6)
Mean Fisher Score	0.2895	1.6590	0.9436
Max Fisher Score (nm)	0.8821 (694.5)	3.6311 (747.5)	3.6311 (747.5)

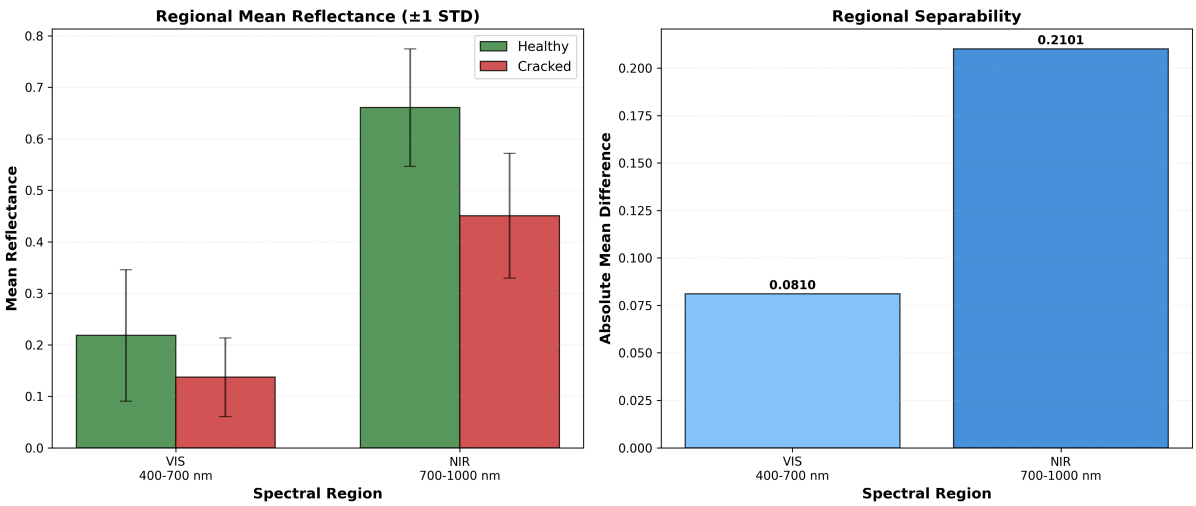


Figure A.1: Regional mean reflectance (±1 SD) and absolute mean class difference for the VIS and NIR ranges.

This appendix reports pixel-level classification and anomaly detection results obtained under class-balanced training, corresponding to the unbalanced experiments presented in the main Results chapter. Model rankings are consistent with those observed under unbalanced sampling.

A.3 Binary Pixel-Level Classification (Balanced)

Table A.8 reports binary pixel-level classification performance under balanced training.

Table A.8: Pixel-level classification metrics (balanced training set). Values are reported as mean ± standard deviation across folds where available. Precision/recall and ROC–AUC refer to the *Cracked* class (positive class); F1 is the weighted F1-score.

Model	Accuracy	Precision (Cracked)	Recall (Cracked)	Weighted F1	ROC–AUC (Cracked)
PLS-DA	0.9961 ± 0.0086	0.9998	0.9929	0.9961	1.0000 ± 0.0000
Logistic Regression (L1)	0.9930 ± 0.0148	0.9994	0.9861	0.9930	0.9999 ± 0.0002
SVM (RBF)	0.9876 ± 0.0202	0.9948	0.9771	0.9875	0.9998 ± 0.0007
Random Forest	0.9935 ± 0.0110	0.9970	0.9882	0.9934	0.9990 ± 0.0038
XGBoost	0.9934 ± 0.0107	0.9972	0.9882	0.9933	1.0000 ± 0.0001

A.4 Multi-Class Pixel-Level Classification (Balanced)

Table A.9 reports multi-class pixel-level classification performance under balanced training across all three labeling strategies.

Table A.9: Pixel-level classification metrics (**balanced** training set). Values are reported as mean ± standard deviation across folds.

Label setup	Model	Accuracy	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	Macro-F1	ROC–AUC (Cracked)	PR–AUC (Cracked)
crack_regular_rest								
	PLS-DA	0.4771 ± 0.0088	0.7061 ± 0.0107	0.8433 ± 0.0910	0.9451 ± 0.1069	0.5875 ± 0.0090	0.9982 ± 0.0046	0.9500 ± 0.0899
	Logistic Regression (L1)	0.8629 ± 0.0033	0.8653 ± 0.0073	0.9865 ± 0.0067	0.9671 ± 0.0739	0.8097 ± 0.0054	0.9994 ± 0.0018	0.9874 ± 0.0090
	Random Forest	0.9277 ± 0.0046	0.9199 ± 0.0110	0.9101 ± 0.0533	0.9194 ± 0.1124	0.9047 ± 0.0087	0.9975 ± 0.0035	0.9725 ± 0.0438
	XGBoost	0.9242 ± 0.0026	0.9111 ± 0.0084	0.9932 ± 0.0042	0.9437 ± 0.0954	0.8839 ± 0.0067	0.9993 ± 0.0016	0.9919 ± 0.0175
	MLP (Small)	0.9219 ± 0.0053	0.9140 ± 0.0125	0.9831 ± 0.0123	0.9832 ± 0.0223	0.8869 ± 0.0167	0.9982 ± 0.0057	0.9934 ± 0.0097
crack_vs_rest								
	PLS-DA	0.4782 ± 0.0090	0.7078 ± 0.0106	0.8445 ± 0.0904	0.9455 ± 0.1063	0.5889 ± 0.0093	0.9982 ± 0.0046	0.9507 ± 0.0886
	Logistic Regression (L1)	0.8642 ± 0.0032	0.8662 ± 0.0074	0.9866 ± 0.0067	0.9674 ± 0.0738	0.8109 ± 0.0055	0.9994 ± 0.0018	0.9876 ± 0.0089
	Random Forest	0.9261 ± 0.0049	0.9191 ± 0.0113	0.9092 ± 0.0546	0.9192 ± 0.1128	0.9038 ± 0.0091	0.9974 ± 0.0036	0.9722 ± 0.0444
	XGBoost	0.9238 ± 0.0027	0.9110 ± 0.0085	0.9931 ± 0.0042	0.9439 ± 0.0954	0.8836 ± 0.0068	0.9993 ± 0.0016	0.9919 ± 0.0176
	MLP (Small)	0.9221 ± 0.0052	0.9143 ± 0.0126	0.9830 ± 0.0124	0.9833 ± 0.0222	0.8871 ± 0.0165	0.9982 ± 0.0057	0.9934 ± 0.0097
multi_class								
	PLS-DA	0.9666 ± 0.0210	0.9786 ± 0.0120	0.7179 ± 0.1759	0.9926 ± 0.0129	0.9010 ± 0.0707	0.9978 ± 0.0025	0.9674 ± 0.0375
	Logistic Regression (L1)	0.9973 ± 0.0031	0.9960 ± 0.0097	0.9677 ± 0.0224	0.9977 ± 0.0078	0.9921 ± 0.0085	0.9996 ± 0.0016	0.9971 ± 0.0079
	Random Forest	0.9966 ± 0.0030	0.9941 ± 0.0120	0.9676 ± 0.0233	0.9969 ± 0.0090	0.9907 ± 0.0100	0.9994 ± 0.0026	0.9970 ± 0.0077
	XGBoost	0.9979 ± 0.0021	0.9947 ± 0.0103	0.9813 ± 0.0149	0.9908 ± 0.0206	0.9924 ± 0.0074	0.9995 ± 0.0021	0.9977 ± 0.0069
	MLP (Small)	0.9906 ± 0.0249	0.9891 ± 0.0224	0.9246 ± 0.1647	0.9866 ± 0.0259	0.9700 ± 0.0704	0.9965 ± 0.0083	0.9898 ± 0.0247

A.5 Autoencoder Anomaly Detection (Balanced)

A.5.1 One-Class Binary Formulation

Table A.10 reports one-class autoencoder performance trained on non-cracked pixels under balanced sampling.

Table A.10: Pixel-level anomaly detection performance using autoencoders trained on non-cracked samples (balanced sampling). Values are reported as mean ± standard deviation across folds. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class).

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.9343 ± 0.0989	0.9122 ± 0.1348	0.9906 ± 0.0340	0.9437 ± 0.0820	0.9823 ± 0.0457	0.9702 ± 0.0592	65.35 ± 1.90	0.008 ± 0.002
Autoencoder (128–64–32)	0.9083 ± 0.1332	0.8771 ± 0.1576	0.9982 ± 0.0058	0.9256 ± 0.1016	0.9919 ± 0.0178	0.9833 ± 0.0357	58.49 ± 3.87	0.007 ± 0.001
Autoencoder (64–32–8)	0.9353 ± 0.1027	0.9074 ± 0.1338	0.9977 ± 0.0100	0.9449 ± 0.0833	0.9858 ± 0.0344	0.9757 ± 0.0501	52.04 ± 2.44	0.006 ± 0.002

A.5.2 Crack-Only Training Formulation

Table A.11 reports multi-class autoencoder performance trained exclusively on cracked pixels under balanced sampling.

Table A.11: Pixel-level anomaly detection performance using autoencoders trained exclusively on *Cracked* pixels and evaluated in a multi-class setting (balanced sampling). Values are reported as mean ± standard deviation across LOGO folds. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class).

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.9660 ± 0.0106	0.9995 ± 0.0008	0.9662 ± 0.0108	0.9826 ± 0.0055	0.9858 ± 0.0281	0.9999 ± 0.0001	61.06 ± 7.56	0.045 ± 0.013
Autoencoder (128–64–32)	0.9861 ± 0.0052	0.9991 ± 0.0017	0.9869 ± 0.0051	0.9930 ± 0.0027	0.9846 ± 0.0274	0.9999 ± 0.0001	63.03 ± 8.09	0.056 ± 0.004
Autoencoder (64–32–8)	0.9629 ± 0.0149	0.9994 ± 0.0011	0.9632 ± 0.0150	0.9809 ± 0.0079	0.9834 ± 0.0298	0.9999 ± 0.0002	59.80 ± 8.55	0.041 ± 0.003

A.5.3 Non-Crack Training Formulation

Table A.12 reports multi-class autoencoder performance trained on all classes except cracked under balanced sampling.

Table A.12: Pixel-level anomaly detection performance using autoencoders trained on all semantic classes except *Cracked* and evaluated in a multi-class setting (balanced sampling). Values are reported for the held-out test split. Accuracy is reported overall, while Precision/Recall/F1/ROC–AUC/PR–AUC correspond to the *Cracked* class (positive class).

Architecture	Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	ROC–AUC (Cracked)	PR–AUC (Cracked)	Train Time [s]	Infer Time [s]
Autoencoder (64–32–16)	0.9791	0.0000	0.0000	0.0000	0.2472	0.0083	2567.17	0.381
Autoencoder (128–64–32)	0.9789	0.0000	0.0000	0.0000	0.2847	0.0084	2260.73	0.299
Autoencoder (64–32–8)	0.9781	0.0000	0.0000	0.0000	0.4513	0.0110	1853.00	0.227

A.6 Wavelength Selection: Generalization and Performance Analysis

This section provides supplementary metrics from the backward feature selection (BFS) analysis on the whole-image pipeline, including calibration-to-test generalization gaps and performance metrics across additional wavelength subsets.

A.6.1 BFS Thresholds (Full Precision)

Table A.13: Summary of BFS wavelength-count thresholds for CRACK metrics (full precision).

Metric	Max score	$n_{\max}$	0.5% drop: n	score	1.0% drop: n	score
CRACK PR–AUC	0.9948	30	11	0.9878	9	0.9790
CRACK F1	0.9854	159	105	0.9800	88	0.9751

A.6.2 Wavelength Results (Full Precision)

Table A.14: Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions (full precision).

Stage	# Wavelengths	Accuracy	Balanced Accuracy	Precision (Cracked)	Recall (Cracked)	F1 (Cracked)	MCC
Early	9	0.7833	0.7902	0.8800	0.6875	0.7719	0.5873
Late	9	0.8000	0.8054	0.8667	0.7647	0.8125	0.6054
Early	11	0.7500	0.7567	0.8400	0.6562	0.7368	0.5195
Late	11	0.8167	0.8111	0.8286	0.8529	0.8406	0.6254
Early	30	**0.8167**	**0.8214**	**0.8889**	**0.7500**	**0.8136**	**0.6447**
Late	30	**0.8333**	**0.8348**	**0.8750**	0.8235	0.8485	**0.6652**
Early	159	0.8000	0.8036	0.8571	**0.7500**	0.8000	0.6071
Late	159	**0.8333**	0.8258	0.8333	**0.8824**	**0.8571**	0.6591

A.6.3 Generalization Analysis

Calibration-to-test generalization gap. Table A.15 shows a consistent calibration-to-test gap across all wavelength subsets and both acquisition stages. Early-stage gaps range from 0.056 to 0.151; late-stage gaps are larger (0.212–0.291).

A.6.4 Wavelength Overlap Across Seeds

Tables A.16 and A.17 report the pairwise Jaccard similarity between the wavelength subsets selected by each seed for the top-30 and top-11 configurations, respectively.

Table A.15: Model Generalization Analysis: Comparison between Calibration and Test MCC for wavelength subsets.

Stage	# Features	MCC Calibration	MCC Test	**Gap (Δ)**
Early	9	0.7381	0.5873	0.1508
	30	0.7232	**0.6447**	**0.0785**
	159	0.6631	0.6071	0.0560
Late	9	0.8968	0.6054	0.2914
	30	0.8771	**0.6652**	**0.2119**
	159	0.8931	0.6591	0.2340

Table A.16: Pairwise Jaccard similarity of top-30 wavelengths across seeds (varying split). Mean Jaccard = 0.493.

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5
Seed 1	—	0.409	0.590	0.722	0.442
Seed 2	0.409	—	0.378	0.409	0.476
Seed 3	0.590	0.378	—	0.590	0.442
Seed 4	0.722	0.409	0.590	—	0.476
Seed 5	0.442	0.476	0.442	0.476	—

Table A.17: Pairwise Jaccard similarity of top-11 wavelengths across seeds (varying split). Mean Jaccard = 0.398.

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5
Seed 1	—	0.500	0.412	0.333	0.412
Seed 2	0.500	—	0.412	0.412	0.333
Seed 3	0.412	0.412	—	0.500	0.333
Seed 4	0.333	0.412	0.500	—	0.333
Seed 5	0.412	0.333	0.333	0.333	—

The top-30 subsets share 13 wavelengths across all five seeds (mean Jaccard = 0.493); the top-11 subsets share 4 wavelengths (mean Jaccard = 0.398). The cross-seed intersection at top-11 comprises 452.25, 548.55, 580.90, and 729.53 nm—all four also appear in the top-30 intersection, confirming consistent selection under both subset sizes.

Bibliography

Abade, A., et al. (2025). From spectrum to yield: Advances in crop photosynthesis with hyperspectral imaging. *Photosynthetica*, *63*(2). https://doi.org/10.32615/ps.2025.012

Akter, T., Faqeerzada, M. A., Kim, Y., Pahlawan, M. F. R., Aline, U., Kim, H., Kim, H., & Cho, B.-K. (2025). Hyperspectral imaging with multivariate analysis for detection of exterior flaws for quality evaluation of apples and pears. *Postharvest Biology and Technology*, *223*, 113453. https://doi.org/10.1016/j.postharvbio.2025.113453

Alchanatis, V., et al. (2022). *High throughput real-time monitoring and prediction of fruit cracking by utilising and upscaling sensing and digital data technologies* (tech. rep.) (Research and Innovation Action (RIA), HORIZON-CL6-2022-GOVERNANCE-01-11). Horizon Europe.

Attri, I., Awasthi, L. K., Sharma, T. P., & Rathee, P. (2023). A review of deep learning techniques used in agriculture. *Ecological Informatics*, *77*, 102217. https://doi.org/10.1016/j.ecoinf.2023.102217

Author, A., et al. (2022). Assessment of scarlotta seedless table grape maturation and quality. *Foods*, *11*(5), 663.

Author, B., et al. (2024). Impact of storage temperature on postharvest quality of scarlotta seedless. *Postharvest Biology and Technology*, *215*, 113013.

Ball, S. R. (2007). *Field guide to spectroscopy*. SPIE Press.

Benelli, A., Cevoli, C., Fabbri, A., & Ragni, L. (2020). Hyperspectral imaging and mini-video devices for non-destructive analysis of fruit quality: A review. *Biosystems Engineering*, *192*, 142–152. https://doi.org/10.1016/j.biosystemseng.2020.01.018

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), 3758.

Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of experimental botany*, *58*(4), 855–867.

Cao, Z., Sun, S., & Bao, X. (2025). A review of computer vision and deep learning applications in crop growth management. *Applied Sciences*, *15*(15), 8438.

Castellarin, S. D., & other. (2016). Characterization of major ripening events during softening in grape and their relationship with growth. *Journal of Experimental Botany*, *67*(3), 709–722.

Chang, B.-M., & Keller, M. (2021). Cuticle and skin cell walls have common and unique roles in grape berry splitting. *Horticulture Research*, *8*, 168. https://doi.org/10.1038/s41438-021-00602-2

Chang, B.-M., Zhang, Y., & Keller, M. (2019). Softening at the onset of grape ripening alters fruit rheological properties and decreases splitting resistance. *Planta*, *250*, 1293–1305. https://doi.org/10.1007/s00425-019-03226-y

Chen, Z., Hao, Y., Liu, Q., Liu, Y., Zhu, M., & Xiao, L. (2024). Deep learning for hyperspectral image classification: A critical evaluation via mutation testing. *Remote Sensing*, *16*(24), 4695. https://doi.org/10.3390/rs16244695

Dhanya, V., Subeesh, A., Kushwaha, N., Vishwakarma, D. K., Nagesh Kumar, T., Ritika, G., & Singh, A. (2022). Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture*, *6*, 211–229. https://doi.org/10.1016/j.aiia.2022.09.007

Dong, C., Yang, T., Liu, L., Wei, Z., Shi, C., & Gao, D. (2025). Early identification of apple bitter pit using hyperspectral imaging technology. *Applied Food Research*, *5*(2), 101166. https://doi.org/10.1016/j.afres.2025.101166

Emmons, C. L., & Scott, J. W. (1998). Ultrastructural and anatomical factors associated with resistance to cuticle cracking in tomato. *International Journal of Plant Sciences*, *159*, 14–22.

Fass, E., Shlomi, E., Ziv, C., Glikman, O., & Helman, D. (2025). Machine learning models based on hyperspectral imaging for pre-harvest tomato fruit quality monitoring. *Computers and Electronics in Agriculture*, *229*, 109788. https://doi.org/10.1016/j.compag.2024.109788

García-Vera, A., Polochè-Arango, F., & Gutiérrez-Bernal, A. (2024). Hyperspectral image analysis and machine learning techniques for crop disease detection and identification: A review. *Sustainability*, *16*(14), 6064. https://doi.org/10.3390/su16146064

Guirao, A., García-Pastor, M. E., Solana-Guilabert, A., Valverde, J. M., & Martínez-Romero, D. (2025). Enhancing calcium transport in table grapes using sorbitol: A sustainable strategy for promoting fruit quality. *Horticulturae*, *11*(11). https://doi.org/10.3390/horticulturae11111320

Haghbin, N., Bakhshipour, A., Zareiforoush, H., & Mousanejad, S. (2023). Non-destructive pre-symptomatic detection of gray mold infection in kiwifruit using hyperspectral data and chemometrics. *Plant Methods*, *19*(1), 53. https://doi.org/10.1186/s13007-023-01032-y

Hall, M. E., Loeb, G., Cadle-Davidson, L., et al. (2018). Grape sour rot: A four-way interaction involving the host, yeast, acetic acid bacteria, and insects. *Phytopathology*, *108*, 1429–1442.

Hasanzadeh, B., Abbaspour-Gilandeh, Y., Soltani-Nazarloo, A., Hernández-Hernández, M., Gallardo-Bernal, I., & Hernández-Hernández, J. L. (2022). Non-destructive detection of fruit quality parameters using hyperspectral imaging, multiple regression analysis and artificial intelligence. *Horticulturae*, *8*(7), 598. https://doi.org/10.3390/horticulturae8070598

Hong, D., Li, C., Yokoya, N., Zhang, B., Jia, X., Plaza, A., Gamba, P., Benediktsson, J. A., & Chanussot, J. (2025). Hyperspectral imaging. https://arxiv.org/abs/2508.08107

Kanwal, N., Kämper, W., Farrar, M. B., Tootoonchy, M., Lynch, C., Nichols, J., Wallace, H. M., Trueman, S. J., & Hosseini Bai, S. (2025). Rapid assessment of lychee and mango fruit quality using hyperspectral imaging. *LWT*, *224*, 117833. https://doi.org/10.1016/j.lwt.2025.117833

Karantzi, A. D., Kafkaletou, M., Tsaniklidis, G., Bai, J., Christopoulos, M. V., Fanourakis, D., & Tsantili, E. (2021). Preharvest foliar salicylic acid sprays reduce cracking of fig fruit at harvest. *Applied Sciences*, *11*(23). https://doi.org/10.3390/app112311374

Khadivi-Khub, A. (2014). Physiological and genetic factors influencing fruit cracking. *Acta Physiologiae Plantarum*, *36*, 1713–1727.

Khaled, A. Y., Ekramirad, N., Donohue, K. D., Villanueva, R. T., & Adedeji, A. A. (2023). Non-destructive hyperspectral imaging and machine learning-based predictive models for physicochemical quality attributes of apples during storage as affected by codling moth infestation. *Agriculture*, *13*(5), 1086. https://doi.org/10.3390/agriculture13051086

La Spada, P., Continella, A., Dominguez, E., Heredia, A., & Gentile, A. (2024). Improving cuticle thickness and quality traits in table grape cv. 'italia' using pre-harvest treatments. *Plants*, *13*(17). https://doi.org/10.3390/plants13172400

Lichter, A., Kaplunov, T., Zutahy, Y., et al. (2014). Effects of gibberellin on cracking and postharvest quality of the seeded table grape 'zainy'. *Acta Horticulturae*, *1079*, 265–271.

Liu, Y., Zhang, H., & Wang, J. (2023). Early detection of plant physiological stress using VIS–NIR hyperspectral imaging. *Sensors*, *23*. https://doi.org/10.3390/s23083915

Mehta, N. (2018). *Hyperspectral imaging cameras in rapid detection and potential diagnosis of agricultural crop threats* (tech. rep.). HinaLea Imaging. https://hinalea.ai/wp-content/uploads/2018/03/Agriculture-in-Hyperspectral-Imaging-Whitepaper.pdf

Min, D., Zhao, J., Bodner, G., Ali, M., Li, F., Zhang, X., & Rewald, B. (2023). Early decay detection in fruit by hyperspectral imaging–principles and application potential. *Food Control*, *152*, 109830. https://doi.org/10.1016/j.foodcont.2023.109830

Monteiro, H. S. A., Tecchio, M. A., Brito, S. d. N. S., Neto, F. J. D., Sánchez, C. A. P. C., Alonso, J. C., Feliciano, D. E. F., Maniero, C. R., Cunha, P. H. H., & Silva, M. d. S. (2025). Phenological development, thermal requirement, and quality of 'brs núbia' (vitis vinifera l. x vitis labrusca l.) grapes on different rootstocks. *Horticulturae*, *11*(5). https://doi.org/10.3390/horticulturae11050466

Ou, Y., Yan, J., Liang, Z., & Zhang, B. (2024). Hyperspectral imaging combined with deep learning for the early detection of strawberry leaf gray mold disease. *Agronomy*, *14*(11), 2694. https://doi.org/10.3390/agronomy14112694

Patrício, D. I., & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and electronics in agriculture*, *153*, 69–81.

Peschel, S., & Knoche, M. (2005). Characterization of microcracks in the cuticle of developing sweet cherry fruit. *Journal of the American Society for Horticultural Science*, *130*, 487–495.

Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, *222*, 109037. https://doi.org/10.1016/j.compag.2024.109037

Santos, M., Egea-Cortines, M., Gonçalves, B., & Matos, M. (2023). Molecular mechanisms involved in fruit cracking: A review. *Frontiers in Plant Science*, *14*, 1130857.

Seo, H.-J., Sawant, S. S., & Song, J. (2022). Fruit cracking in pears: Its cause and management—a review. *Agronomy*, *12*(10). https://doi.org/10.3390/agronomy12102437

Shaw, G., & Manolakis, D. (2002). *Hyperspectral imaging remote sensing*. John Wiley & Sons.

Sun, Q., Li, X., & Zhao, Y. (2024). VIS–NIR hyperspectral imaging for internal quality assessment of fruits. *Postharvest Biology and Technology*. https://doi.org/10.1016/j.postharvbio.2024.112095

Sun, T., Xu, R., Zhang, G.-x., Li, X.-y., Fan, S.-d., & Wang, X. (2020). Non-destructive determination of moisture content of single barley grain using hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *226*, 117623.

Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation—a review. *Information processing in agriculture*, *7*(1), 1–19.

Tian, P., Meng, Q., Wu, Z., Lin, J., Huang, X., Zhu, H., Zhou, X., Qiu, Z., Huang, Y., & Li, Y. (2023). Detection of mango soluble solid content using hyperspectral

imaging technology. *Infrared Physics & Technology*, *129*, 104576. https://doi.org/10.1016/j.infrared.2023.104576

Vignati, S., Tugnolo, A., Giovenzana, V., Pampuri, A., Casson, A., Guidetti, R., & Beghi, R. (2023). Hyperspectral imaging for fresh-cut fruit and vegetable quality assessment: Basic concepts and applications. *Applied Sciences*, *13*(17). https://doi.org/10.3390/app13179740

Wang, B., Yang, H., Zhang, S., & Li, L. (2023). Detection of defective features in cerasus humilis fruit based on hyperspectral imaging technology. *Applied Sciences*, *13*(5). https://doi.org/10.3390/app13053279

Wang, C., Li, X., Zhang, Z., Luo, X., Cai, J., & Wang, A. (2025). Nondestructive quality detection of characteristic fruits based on vis/nir spectroscopy: Principles, systems, and applications. *Agriculture*, *15*(20). https://doi.org/10.3390/agriculture15202167

Wang, T., & Zhao, L. (2024). Hyperspectral imaging for monitoring plant physiological status. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2024.1298743

Xu, J., Chen, L., Dong, J., Jiang, L., & Hong, L. (2025). Overview of fruit cracking in sweet cherry (prunus avium l.): Causes, testing methods, mitigation strategies, and research perspectives. *Frontiers in Sustainable Food Systems*, *9*, 1534778.

Yang, C., Guo, Z., Fernandes Barbin, D., Dai, Z., Watson, N., Povey, M., & Zou, X. (2025). Hyperspectral imaging and deep learning for quality and safety inspection of fruits and vegetables: A review. *Journal of Agricultural and Food Chemistry*, *73*(17), 10019–10035. https://doi.org/10.1021/acs.jafc.4c11492

Yu, J., Li, J., Hong, Z., Zeng, Q., Fu, Y., Deng, R., Xu, K., Huang, Z., & Zhu, M. (2025). Vvpip1;1 plays a role in grape berry cracking by regulating water uptake. *Horticultural Plant Journal*, *11*(4), 1506–1516. https://doi.org/10.1016/j.hpj.2024.07.010

Zhang, C., Wang, Q., He, H., Wu, Y., Shan, W., & Liu, H. (2025). Postharvest preservation strategies for table grapes: A comprehensive review from practical methods to future developments. *Plants*, *14*(16). https://doi.org/10.3390/plants14162462

Zhang, X., Li, Y., Chen, H., & Wang, Q. (2024). Application of hyperspectral imaging in fruit quality and safety assessment. *Frontiers in Plant Science*, *15*, 1432120. https://doi.org/10.3389/fpls.2024.1432120

Zhang, X., Vinatzer, B. A., & Li, S. (2024). Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease. *Scientific Reports*, *14*(1), 27666. https://doi.org/10.1038/s41598-024-78650-6

Zhang, Y., Chen, J., & Liu, Q. (2022). Physiological interpretation of VIS–NIR hyperspectral reflectance in vegetation. *Remote Sensing*, *14*. https://doi.org/10.3390/rs14215345

Zhang, Y., Li, X., Wang, M., Xu, T., Huang, K., Sun, Y., Yuan, Q., Lei, X., Qi, Y., & Lv, X. (2024). Early detection and lesion visualization of pear leaf anthracnose based on multi-source feature fusion of hyperspectral imaging. *Frontiers in Plant Science*, *15*, 1461855. https://doi.org/10.3389/fpls.2024.1461855

Zhang, Z., Cheng, H., Chen, M., Zhang, L., Cheng, Y., Geng, W., & Guan, J. (2024). Detection of pear quality using hyperspectral imaging technology and machine learning analysis. *Foods*, *13*(23), 3956. https://doi.org/10.3390/foods13233956

Zhu, H., Qin, S., Liang, S., Su, M., Wang, P., & He, Y. (2025). Hyperspectral imaging and machine learning for quality assessment of apples with different bagging types. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *343*, 126443. https://doi.org/10.1016/j.saa.2025.126443

Zwinkels, J. C. (2015). Spectrophotometry. In *Encyclopedia of color science and technology*. Springer. https://doi.org/10.1007/978-1-4419-8071-7_221