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 **9**

1.1 The Problem of Fruit Cracking . . . 9

1.2 Economic Impact . . . 9

1.3 Research Motivation . . . 9

1.3.1 Main Research Question . . . 11

1.4 Research Objectives and Hypotheses . . . 11

1.4.1 Research Objectives . . . 11

1.4.2 Contributions . . . 11

2 Literature Review **13**

2.1 The Challenge of Grape Cracking: Physiology and Economic Impact . . . 14

2.2 Biomechanical Mechanisms and Cultivar Susceptibility . . . 15

2.3 Bridging the Gap: The Need for Advanced Sensing in Complex Environments 16

2.4 Hyperspectral Imaging for Agricultural Applications . . . 17

2.5 Machine Learning Paradigms for Hyperspectral Data . . . 20

2.6 Related Work: From Visible Defect Detection to Early Warning . . . 21

2.7 Research Gap and Study Positioning . . . 22

3 Research Objectives and Hypotheses **23**

3.1 Research Objectives . . . 23

3.2 Research Hypotheses . . . 23

3.3 Summary of Main Contributions . . . 25

3.4 Thesis Structure . . . 25

4 Materials and Methods **26**

4.1 Data Acquisition . . . 27

4.1.1 Study Area and Experimental Design . . . 27

4.1.2 Multi-Modal Imaging Systems . . . 29

4.1.3 Data Collection Protocol . . . 31

4.2 Dataset Construction and Partitioning . . . 32

4.2.1 Interactive Pixel-Level Annotation via SAM2 . . . 32

4.2.2 Whole-Image Dataset Formulation 34
4.3 Modeling and Evaluation Methodology 35
4.3.1 Global Spectral Preprocessing 35
4.3.2 Data Partitioning and Leakage Prevention Strategies 35
4.3.3 Pixel-Level Classification Framework 36
4.3.4 Autoencoder-Based Anomaly Detection 37
4.3.5 Whole-Image Detection and Aggregation Pipeline 38
4.3.6 Wavelength Selection 39
4.3.7 Evaluation Protocols and Performance Metrics 41
4.3.8 Reproducibility and Implementation Details 41

5 Experimental Results **42**
5.1 Overview of Experimental Results 42
5.1.1 Evaluation Baselines 43
5.2 Spectral Separability and Informative Wavelengths 44
5.2.1 Mean Spectral Signatures of Healthy and Cracked Tissue 44
5.2.2 Spectral Separability Across Wavelength Regions 45
5.3 Pixel-Level Binary Classification 46
5.4 Pixel-Level Multi-Class Classification 47
5.5 Autoencoder-Based Anomaly Detection 48
5.6 Whole-Image Crack Detection 50
5.6.1 Supervised Detection Performance 50
5.6.2 Unsupervised Baseline Comparison 52
5.7 Wavelength Selection via Backward Feature Selection 53
5.7.1 BFS Trajectories and Subset Selection 53
5.7.2 End-to-End Wavelength-Selection Results 56
5.7.3 BFS Wavelength-Selection Stability Analysis 57

6 Discussion and Conclusions **59**
6.1 Physiological and Pre-processing Foundations 59
6.1.1 The Biological Basis of Spectral Signatures 59
6.1.2 Data Normalization and Pre-processing Restraint 59
6.2 Managing Field Clutter: The Role of Spatial Aggregation 61
6.2.1 Overcoming Spectral Ambiguity: The Dry Wood Challenge 61
6.2.2 Geometric Filtering and Structural Regularization 61
6.3 Dimensionality Reduction, Algorithm Selection, and Sensor Design 63
6.3.1 Feature Selection and Multispectral Sensor Design 63
6.3.2 Methodological Rationale: BFS and Evaluation Metrics 63
6.3.3 Classifier Selection and Spatial Validation 64

6.3.4 The Limits of Unsupervised Anomaly Detection 65
6.3.5 Methodological Challenges and Data Curation Bottlenecks 66
6.4 Conclusions and Future Work 67
6.4.1 Summary of Contributions 67
6.4.2 Future Research Directions 67

A Supplementary Results **69**
A.1 Cross-Validation Statistics and Full Precision Tables 69
A.1.1 Pixel-Level Classification: Binary and Multi-Class 69
A.1.2 Autoencoder-Based Anomaly Detection 69
A.2 Regional VIS–NIR Reflectance Summaries 69
A.3 Binary Pixel-Level Classification (Balanced) 72
A.4 Multi-Class Pixel-Level Classification (Balanced) 73
A.5 Autoencoder Anomaly Detection (Balanced) 74
A.5.1 One-Class Binary Formulation 74
A.5.2 Crack-Only Training Formulation 74
A.5.3 Non-Crack Training Formulation 75
A.6 Wavelength Selection: Generalization and Performance Analysis 76
A.6.1 BFS Thresholds (Full Precision) 76
A.6.2 Wavelength Results (Full Precision) 76
A.6.3 Generalization Analysis 76
A.6.4 Wavelength Overlap Across Seeds 76
A.7 Wavelength Stability: Supplementary Visualizations 78

List of Figures

2.1 The formula for reflectance Blackburn, 2007. 17
2.2 Spectral signatures of barley samples with varying moisture content, illustrating how physiological changes produce wavelength-specific shifts in reflectance Shaw and Manolakis, 2002. 18
2.3 Hyperspectral Imaging (HSI) data cube illustrating the spectral signature of a pixel. 18
2.4 Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.). 18

4.1 A general view of the Lachish vineyard (Row 2) illustrating the semi-arid landscape and vine rows. 27
4.2 (a) Vineyard map with randomized blocks and irrigation treatments; (b) example of a plastic-tagged grape cluster. 28
4.3 Hyperspectral imaging setup and data acquisition using the Specim IQ system under vineyard conditions. 29
4.4 Thermal imaging system and example thermal data acquired from grape clusters under vineyard conditions. 30
4.5 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, and (3) high-resolution RGB image of the same cluster. (4) Spectral signature extracted from a pixel within the segmented region, and (5) autoencoder reconstruction of the signature. 32
4.6 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. 34
4.7 Whole-image crack-detection pipeline from pixel-level CRACK probabilities to image-level classification via spatial aggregation. 39

5.1 Mean spectral signatures (± standard deviation) for healthy and cracked tissue across the VIS–NIR range. 44

5.2 Wavelength-wise spectral separability metrics between healthy and cracked tissue, including absolute mean difference, Cohen's d, and Fisher score. . 45
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 $n = 159$). 53
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. 54
5.5 Distribution and popularity of the top-30 selected wavelengths across five varying-split seeds. 57
5.6 Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-30 subsets. 58

A.1 Regional mean reflectance (±1 SD) and absolute mean class difference for the VIS and NIR ranges. 71
A.2 Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-30 subsets. 78
A.3 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. 79
A.4 Distribution and popularity of the top-11 selected wavelengths across five varying-split seeds. 80
A.5 Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-11 subsets. 80
A.6 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. 81
A.7 Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-11 subsets. 82

List of Tables

1.1 Estimated financial losses from fruit cracking in major crops.(Alchanatis et al., 2022). 10

4.1 Multi-class pixel-level dataset: exported pixels by class. 33

4.2 Exact dataset statistics used in the experiments. *Note: The negative class in the multi-class pixel dataset also includes the background classes detailed in Table 4.1. 34

5.1 Pixel-level classification metrics (unbalanced, binary setting). Truncated mean values. Full cross-validation statistics in Appendix A.1. 46

5.2 Pixel-level classification metrics (unbalanced, multi-class). Truncated mean values. Full cross-validation statistics in Appendix A.1. 47

5.3 Pixel-level autoencoder anomaly detection (binary setting). Truncated mean values. Full statistics in Appendix A.1. 48

5.4 Pixel-level autoencoder anomaly detection (crack-only training). Truncated mean values. Full statistics in Appendix A.1. 48

5.5 Pixel-level autoencoder anomaly detection (all non-CRACK training). Truncated mean values. Full statistics in Appendix A.1. 48

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. 50

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. 51

5.8 Full-image crack detection using autoencoder-based anomaly detection. F1 (Cracked) is the primary metric; the Cracked class is positive. 52

5.9 Summary of BFS wavelength-count thresholds for CRACK metrics (Full precision in Appendix A.6.1). 55

5.10 Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions. 56

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. 69
A.2 Pixel-level classification metrics (**unbalanced** training set, multi-class). Values are reported as mean ± standard deviation across folds. 70
A.3 Pixel-level autoencoder anomaly detection: binary setting (non-cracked as normal class). Values are reported as mean ± standard deviation across LOGO folds. 70
A.4 Pixel-level autoencoder anomaly detection: trained on CRACK pixels (multi-class setting). Values are mean across LOGO folds. 70
A.5 Pixel-level autoencoder anomaly detection: trained on all non-CRACK classes (multi-class setting). Single group-wise train–test split. 70
A.6 Summary statistics of reflectance values across spectral regions for healthy and cracked tissue. 70
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. 71
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. 72
A.9 Pixel-level classification metrics (**balanced** training set). Values are reported as mean ± standard deviation across folds. 73
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). 74
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). 74
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). 75

A.13 Summary of BFS wavelength-count thresholds for CRACK metrics (full precision). 76
A.14 Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions (full precision). 76
A.15 Model Generalization Analysis: Comparison between Calibration and Test MCC for wavelength subsets. 77
A.16 Pairwise Jaccard similarity of top-30 wavelengths across seeds (varying split). Mean Jaccard = 0.493. 77
A.17 Pairwise Jaccard similarity of top-11 wavelengths across seeds (varying split). Mean Jaccard = 0.398. 77

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 2.2 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 2.4, 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.

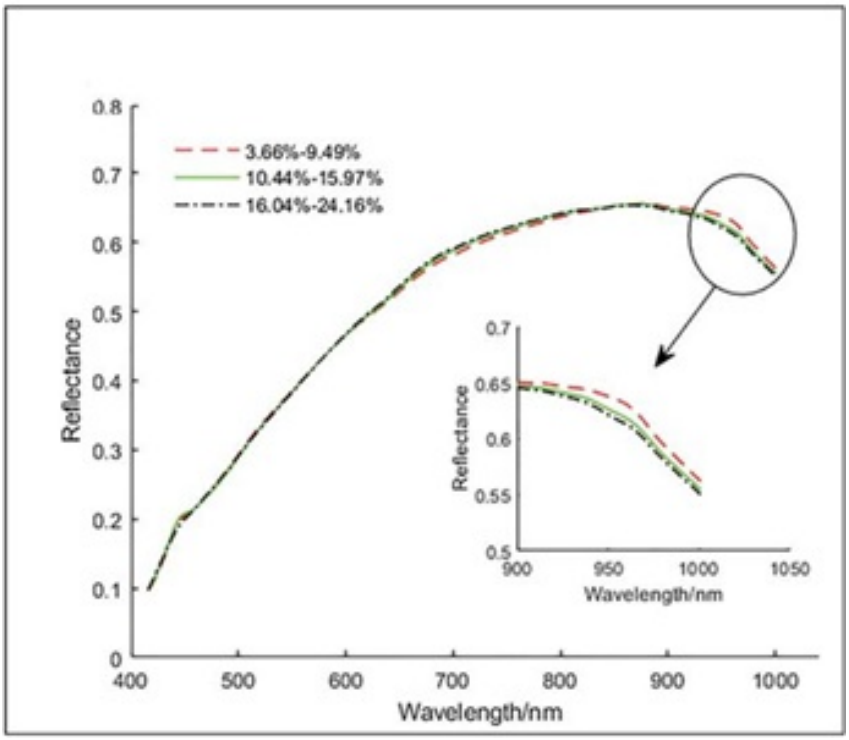


Figure 2.2: Spectral signatures of barley samples with varying moisture content, illustrating how physiological changes produce wavelength-specific shifts in reflectance Shaw and Manolakis, 2002.

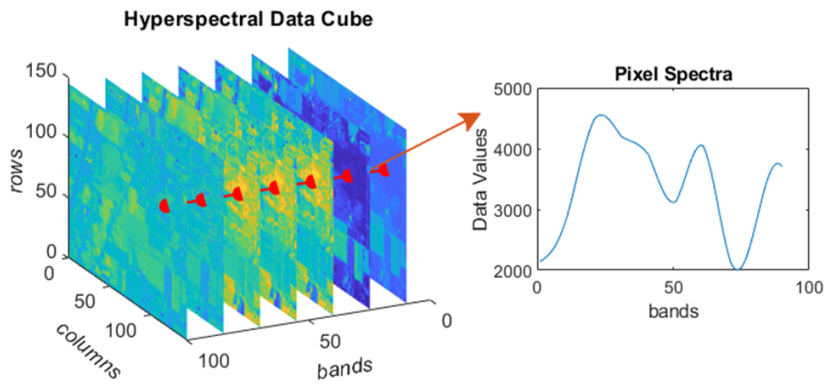


Figure 2.3: Hyperspectral Imaging (HSI) data cube illustrating the spectral signature of a pixel.

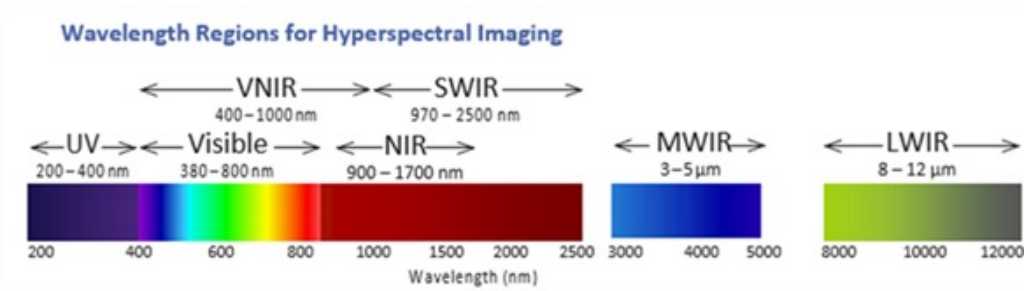


Figure 2.4: Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.).

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; 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.

3.3 Summary of Main Contributions

To address the objectives and hypotheses outlined above, this study developed an end-to-end hyperspectral imaging framework. As will be detailed in the subsequent chapters, the findings first confirm the physical distinguishability of cracks at the pixel level, where an optimized XGBoost classifier achieved near-perfect performance (PR–AUC $\approx$ 1.0). Crucially, the study successfully transitioned this capability to realistic, whole-image vineyard scenes containing complex background clutter. By employing strategic backward feature selection, the framework reduced the spectral dimensionality to an optimal subset of 30 wavelengths while achieving robust full-image detection performance. Specifically, the aggregated model yielded an F1-score of 0.81 under early-stage (pre-symptomatic) conditions and 0.84 under late-stage conditions. These results not only validate the research hypotheses but also demonstrate the operational feasibility of transitioning HSI into a proactive, field-ready agricultural early-warning tool.

3.4 Thesis Structure

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

This chapter details the experimental design, data acquisition procedures, and computational methodologies employed to detect grape cracking using hyperspectral imaging under field conditions. To address the research objectives, the methodology is organized into three progressive phases:

1. **Data Acquisition:** Describes the commercial vineyard experimental setup, the multi-modal sensor suite (HSI, RGB, and Thermal), and the field collection protocols.

2. **Dataset Construction:** Details the formulation of the pixel-level and whole-image datasets, including the use of an interactive annotation pipeline powered by foundation vision models (SAM2) for precise spectral extraction.

3. **Modeling and Evaluation Framework:** Outlines the end-to-end analytical pipeline, progressing from global spectral preprocessing and baseline pixel-level classification to robust whole-image spatial aggregation and strategic wavelength selection.

4.1 Data Acquisition

4.1.1 Study Area and Experimental Design

The study was conducted in the Lachish region of Israel, a semi-arid area characterized by hot, dry summers and mild winters, making it a central hub for Israeli viticulture. Figure 4.1 shows a view of the Lachish vineyard environment. The experiment took place in a commercial table grape vineyard (*Vitis vinifera* L., cv. 'Scarlotta'), operated in collaboration with the "Tali Grapes" brand. The vineyard was planted in 2018 and grafted onto 1103 Paulsen rootstock.

Figure 4.1: A general view of the Lachish vineyard (Row 2) illustrating the semi-arid landscape and vine rows.

The experimental layout followed a Randomized Complete Block Design (RCBD) comprising five replicates (Blocks A–E). Within each block, four distinct irrigation regimes were applied, yielding a total of 20 experimental plots (5 blocks $\times$ 4 treatments). The two panels in Figure 4.2 summarize the block structure and irrigation assignment and provide an example of a tagged grape cluster used for repeated imaging; together they establish the spatial provenance for the row-based calibration and testing sets used later in this study.

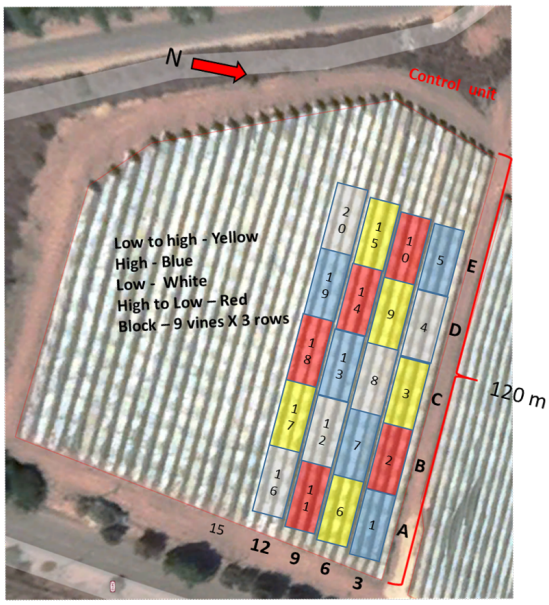
Each experimental plot consisted of three consecutive rows of nine vines. To minimize edge effects and water spillover between treatments, only vines from the middle row of each plot were monitored. From these middle rows, 12 representative grape clusters were tagged per plot, resulting in a total sample size of 240 monitored clusters. To ensure accurate repeated imaging throughout the season, these target clusters were marked

with durable, color-coded tags indicating their specific irrigation regime, and their vine locations were georeferenced using a GPS module.

The four irrigation regimes, initiated on April 29, 2024, and applied via drip irrigation, were:

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

The experimental campaign ran continuously until late September 2024. Alongside the imaging procedures detailed below, weekly physiological and yield measurements—including stem water potential, stomatal conductance, and soil moisture—were collected to provide agronomic context for the spectral analyses.



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

(b) Tagged grape cluster used for tracking and repeated imaging; tags indicate the specific irrigation regime.

Figure 4.2: (a) Vineyard map with randomized blocks and irrigation treatments; (b) example of a plastic-tagged grape cluster.

4.1.2 Multi-Modal Imaging Systems

A multi-modal sensor suite was deployed to capture complementary spectral, spatial, and thermal data. The three imaging systems were co-located and operated from a fixed working distance of approximately 1 m from the target grape clusters, ensuring consistent scale across all modalities.

Hyperspectral Imaging System (HSI)

Hyperspectral data were acquired using the **Specim IQ** (Specim, Spectral Imaging Ltd., Oulu, Finland), a portable snapshot camera operating in the Visible and Near-Infrared (VNIR) range (400–1000 nm). Figure 4.3 illustrates the field deployment of the system, highlighting the complex background and illumination conditions inherent to vineyard acquisition.

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

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

Figure 4.3: Hyperspectral imaging setup and data acquisition using the Specim IQ system under vineyard conditions.

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

- **Spatial Resolution:** The camera features a 512×512 pixel sensor. With a 31°×31° Field of View (FOV), the 1 m working distance provides a physical coverage of $\approx 0.55 \times 0.55$ m, yielding a theoretical spatial resolution of **1.08 mm/pixel**—sufficient for resolving individual berries and macroscopic cracks.

- **Radiometric Calibration:** A spectrally flat, diffuse white reference panel (Spectralon®, ≈ 99% reflectance) was placed within every scene to enable accurate conversion of raw radiance to relative reflectance, compensating for dynamic ambient illumination.

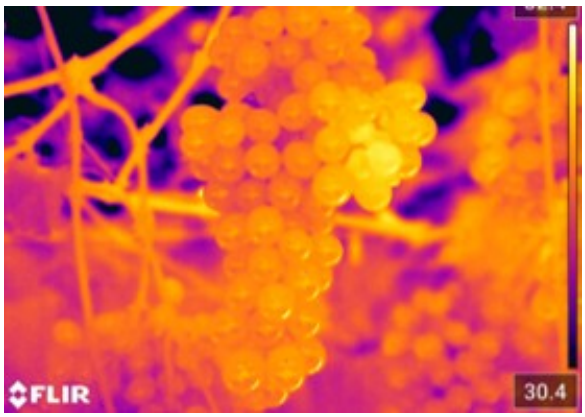
RGB Imaging System

High-resolution visual RGB Imaging System , a **Canon EOS R5** digital camera was utilized. Equipped with a 45-megapixel full-frame CMOS sensor (8192 × 4608 pixels) and a **35 mm** prime lens, the system provided a wide FOV (54.4° × 37.8°). At the standard 1 m working distance, this configuration achieved an extreme spatial resolution of approximately **0.13 mm/pixel**. This fine resolution is essential for identifying hairline fissures and cuticular alterations prior to their visibility in the HSI data.

Thermal Imaging System

Surface temperatures of the canopy and clusters were monitored using a **FLIR T1020** (Teledyne FLIR LLC, USA) high-definition thermal camera (Figure 4.4). This uncooled microbolometer operates in the Long-Wave Infrared (LWIR) range (7.5–14 μm) with a resolution of 1024 × 768 pixels. Equipped with a **34 mm** lens and boasting a high thermal sensitivity (NETD < 20 mK at 30°C), the system achieved a spatial resolution of **0.81 mm/pixel** at 1 m. This allowed for precise extraction of berry surface temperatures, minimizing mixed-pixel effects from surrounding foliage.

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



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

Figure 4.4: Thermal imaging system and example thermal data acquired from grape clusters under vineyard conditions.

The combined multi-modal imaging campaign generated a substantial data footprint. On average, a single HSI datacube required ≈ 200 MB, compared to 11 MB for RGB and 2.5 MB for thermal images. The entire season's dataset exceeded **575 GB**, necessitating robust computational infrastructure for downstream processing.

4.1.3 Data Collection Protocol

Data collection was conducted on dedicated field days, typically once per week from June through late September. To manage the different operational requirements of the sensors and to leverage specific environmental conditions, a structured diurnal sampling strategy was implemented:

- **Early Morning (RGB):** Image acquisition commenced between 07:00 and 08:00. All 240 target clusters were imaged first using the RGB system to capture baseline visual conditions under relatively uniform, soft morning lighting.
- **Mid-Morning (HSI):** Hyperspectral acquisition began at approximately 10:00 when solar illumination was more intense and stable. Because the HSI setup required manual re-positioning, white reference placement, and dedicated spectral focusing for each shot, a single scan took 2–3 minutes. Consequently, a representative subset of 120 clusters was imaged per session, alternating coverage across field days to encompass the entire experimental layout.
- **Midday (Thermal):** Thermal imaging was purposely scheduled for peak heat hours (starting around 12:00). Utilizing the high ambient temperatures of the semi-arid Negev region maximizes thermal contrast on the berry surface. Cracked or decaying tissues—characterized by exposed moisture or altered transpirational cooling—exhibit distinct thermal signatures compared to intact skin, making them significantly more detectable under peak thermal load.

All imaging systems were spatially co-aligned to the target clusters during acquisition. While this thesis focuses primarily on the analysis of the RGB and HSI datasets, the temporal and spatial synchronization of all three modalities provides a foundation for future integrative multi-sensor fusion.

4.2 Dataset Construction and Partitioning

Transforming raw hyperspectral datacubes into machine-learning-ready datasets required precise spatial annotation and rigorous partitioning. The dataset construction was divided into two levels: an extracted pixel-level dataset for spectral characterization and baseline modeling, and a whole-image dataset for evaluating spatial aggregation under field conditions.

4.2.1 Interactive Pixel-Level Annotation via SAM2

To enable accurate extraction of spectral signatures for both the binary (*Cracked* vs. *Healthy*) and multi-class problem formulations, a custom graphical user interface (GUI) was developed in Python. This tool integrated data visualization, interactive annotation, and automated segmentation into a single streamlined platform (Figure 4.5).

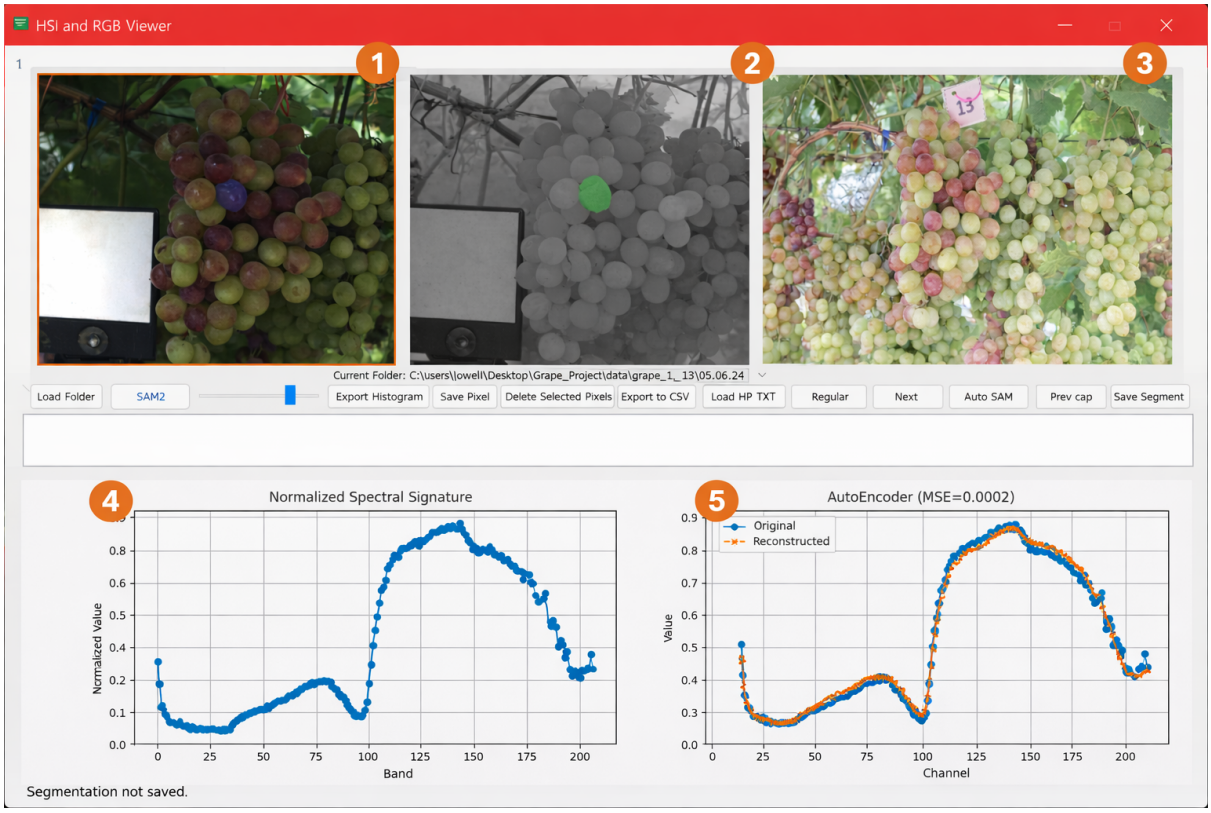


Figure 4.5: Graphical user interface of the HSI–RGB annotation tool. (1) Hyperspectral image visualized as an RGB composite with a SAM2-generated segmentation overlay, (2) wavelength-selectable grayscale HSI band, and (3) high-resolution RGB image of the same cluster. (4) Spectral signature extracted from a pixel within the segmented region, and (5) autoencoder reconstruction of the signature.

The annotation workflow leveraged foundation vision models to overcome the low spatial resolution of the HSI data. The procedure operated as follows:

1. **Interactive Seeding:** The user selected a seed pixel on a target object (e.g., a cracked berry, intact tissue, leaf, or background element) using either the high-resolution RGB image or the HSI composite.

2. **SAM2 Segmentation:** The seed pixel served as a point prompt for the **Segment Anything Model 2 (SAM2)** Ravi et al., 2024. SAM2 generated a precise binary mask of the object on the RGB image, which was subsequently spatially aligned and projected onto the hyperspectral datacube.

3. **Spectral Extraction:** The tool extracted the 204-band spectral vectors for all pixels within the projected mask. The user assigned a definitive class label to the region, and the grouped signatures were stored along with their parent cluster ID and acquisition date.

This process was initially used to build a binary dataset containing strictly healthy and cracked grape tissues. It was subsequently expanded to build a comprehensive 10-class multi-class dataset, incorporating diverse vineyard background elements (e.g., leaves, branches, soil, plastic covers, and metallic tripod parts) to train robust classifiers capable of rejecting non-grape clutter. The definitions and distributions of these classes are summarized in Table 4.1.

Class	**Pixels**	**Description**
BACKGROUND	378,194	Soil, ground, and mixed non-object pixels.
WHITE_REFERENCE	270,159	Calibration panel pixels.
TRIPOD	137,494	Camera tripod and imaging setup components.
LEAF	120,552	Green vine foliage.
BRANCH	78,654	Woody vine structures (stem, cane).
REGULAR	69,618	Intact grape tissue without visible defects.
PLASTIC	66,185	Artificial vineyard elements (covers, ties).
BURNT_PIXEL	36,676	Saturated or overexposed pixels.
CRACK	24,548	Grape tissue exhibiting visible cracking symptoms.
IRON	6,961	Metallic elements (wires, stakes).

Table 4.1: Multi-class pixel-level dataset: exported pixels by class.

4.2.2 Whole-Image Dataset Formulation

While pixel-level data are essential for training core spectral classifiers, operational evaluation requires inference on full spatial scenes. The whole-image dataset is structured at the *(grape_id, week_date)* level, with each record corresponding to a complete hyperspectral datacube of a cluster.

To rigorously evaluate the system's early-warning capabilities, the whole-image dataset was stratified into two distinct acquisition conditions based on the progression of the disorder at the time of measurement:

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

The exact sample distributions for these stages are detailed in Table 4.2. The spatial partitioning of these images to simulate a realistic early-warning scenario and prevent information leakage is detailed in Section 4.3.2.

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 4.6: 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.

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 4.2: Exact dataset statistics used in the experiments. *Note: The negative class in the multi-class pixel dataset also includes the background classes detailed in Table 4.1.

4.3 Modeling and Evaluation Methodology

This section outlines the computational frameworks used to analyze the hyperspectral data, progressing from fundamental spectral preprocessing to robust pixel-level classification and anomaly detection.

4.3.1 Global Spectral Preprocessing

To ensure methodological uniformity and prevent information leakage Thenkarail et al., 2016, all hyperspectral datacubes were subjected to a strictly fixed preprocessing pipeline prior to any model training or evaluation.

- **Spectral Range Restriction:** Raw signatures were cropped to the 450–925 nm range (159 contiguous bands). This step discarded extreme wavelengths (400–450 nm and 925–1000 nm) known to exhibit low signal-to-noise ratios and reflectance artifacts under outdoor field conditions Behmann et al., 2018.
- **Per-Spectrum Normalization:** Each remaining spectral vector was independently normalized using the *Standard Normal Variate* (SNV) transformation. By mean-centering and scaling each pixel exclusively by its own standard deviation, SNV mitigates illumination variability and scattering effects without relying on dataset-level statistics. This strictly per-pixel approach intrinsically prevents data leakage between training and testing partitions Kaltenborn et al., 2025; Mishra et al., 2017.

4.3.2 Data Partitioning and Leakage Prevention Strategies

A foundational principle of this study was the strict prevention of spatial and spectral information leakage. Because hyperspectral vineyard data exhibit strong spatial autocorrelation, improper data splitting would artificially inflate model performance. To ensure robust evaluation, distinct leakage prevention strategies were enforced at both the pixel and whole-image levels.

Pixel-Level Leakage Prevention: Leave-One-Group-Out (LOGO)

At the pixel level, contiguous pixels extracted from the same grape cluster share highly correlated spectral signatures. To prevent leakage, every extracted pixel was rigidly associated with its parent *cluster ID* and *acquisition date* (as defined in §4.2.1). This metadata was used to enforce a **Leave-One-Group-Out (LOGO)** cross-validation strategy. Under this protocol, all pixels originating from the same physical cluster were strictly assigned to either the training set or the test set in their entirety. For the multi-class

formulation, a domain-aware LOGO variant was implemented: grape-containing acquisitions were folded using standard LOGO, while non-grape background samples (e.g., soil, leaves) were allocated to a fixed 80/20 hold-out split to prevent semantic leakage.

Whole-Image Leakage Prevention: Spatial Row Partitioning

At the whole-image level, evaluating the early-warning framework required simulating a realistic deployment scenario on unseen vines. To prevent spatial autocorrelation between adjacent vines, a strict geographical partitioning scheme was enforced. All hyperspectral records acquired from vineyard **Row 1** were designated exclusively as the calibration and training set, while records from **Row 2** served as the held-out evaluation (test) set. Because the two rows are physically disjoint (separated by distance and foliage), this row-based split robustly evaluates the model's ability to generalize to entirely new regions of the vineyard.

Crucially, to simulate a true early-warning application, the underlying pixel-level models used in the whole-image pipeline were trained *exclusively* on pixels extracted from **Late-stage** images in Row 1. Early-stage records were entirely hidden from the pixel-level training phase. The whole-image aggregation hyperparameters were tuned strictly on Row 1, frozen, and then applied blindly to the Row 2 held-out set.

4.3.3 Pixel-Level Classification Framework

Supervised classification was applied at the pixel level to discriminate between cracked tissue, healthy tissue, and background elements based solely on their SNV-normalized spectral signatures (159 bands).

Problem Formulations: The classification was evaluated under two primary settings:

- **Binary:** Classifying pixels strictly as *Healthy* ($y = 0$) or *Cracked* ($y = 1$).
- **Multi-class:** Classifying pixels into the full 10-class taxonomy (incorporating background elements) to test robustness against vineyard clutter.

Algorithm Selection and Training: Six algorithms were evaluated: Partial Least Squares Discriminant Analysis (PLS-DA), L1-regularized Logistic Regression, Support Vector Machine (RBF kernel), Random Forest, a shallow Multi-Layer Perceptron (MLP), and **XGBoost** (Gradient-boosted decision trees). XGBoost was ultimately adopted as the baseline model across all primary comparisons due to its robust handling of high-dimensional, correlated spectral features. Models were evaluated under both naturally unbalanced and strictly balanced training regimes (achieved via segment-proportional undersampling to preserve within-class diversity).

4.3.4 Autoencoder-Based Anomaly Detection

To explore whether crack damage could be identified without explicit positive labels, a one-class anomaly detection framework was implemented using a fully connected, symmetric autoencoder.

The autoencoder (optimized via Adam, MSE loss) was trained exclusively on a designated "normal" class (e.g., healthy pixels). During inference, the anomaly score for a given spectrum x was defined as its reconstruction error, $e(x) = \frac{1}{D}\|x - \hat{x}\|_2^2$. An operating threshold τ was established at the 99th percentile of the training errors. For integration into the whole-image pipeline, reconstruction errors were mapped to probability-like crack scores via a sigmoid function:

$$p_{\text{CRACK}}(x) = \frac{1}{1 + \exp\left(\frac{e(x)-\tau}{\alpha}\right)} \tag{4.1}$$

where $\alpha \approx 0.3\tau$ served as a temperature parameter controlling the sharpness of the transition.

4.3.5 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 4.6, and consists of the following stages:

1. **Probability Map Generation:** The trained multi-class XGBoost model processes the full input datacube ($H \times W \times 159$), producing a dense crack probability map $P_{\text{CRACK}} \in [0, 1]^{H \times W}$.

2. **Morphological Integration:** To enforce spatial coherence, adjacent pixels with elevated crack probabilities are connected using morphological closing operations. This bridges tiny gaps in hairline micro-cracks and suppresses isolated noisy pixels, forming continuous regions of interest (BLOBs).

3. **Geometric Filtering:** The resulting connected components are filtered based on area constraints. BLOBs smaller than a defined minimum area are discarded as noise, while structurally valid regions are retained.

4. **Patch-Based Decision Logic:** The filtered probability map is divided into fixed-size spatial patches. A patch is classified as crack-positive if the density of crack-related pixels within its boundaries exceeds a calibrated threshold.

5. **Global Image-Level Inference:** The final image-level decision employs a dual-condition logic to maximize noise rejection. The entire cluster is flagged as *Cracked* only if both of the following conditions are met simultaneously: (1) at least one spatial patch satisfies the crack-positive criteria, **and** (2) the global proportion of crack-related pixels across the entire image exceeds a predefined overall density threshold. If either condition is not met, the image is classified as *Non-Cracked*.

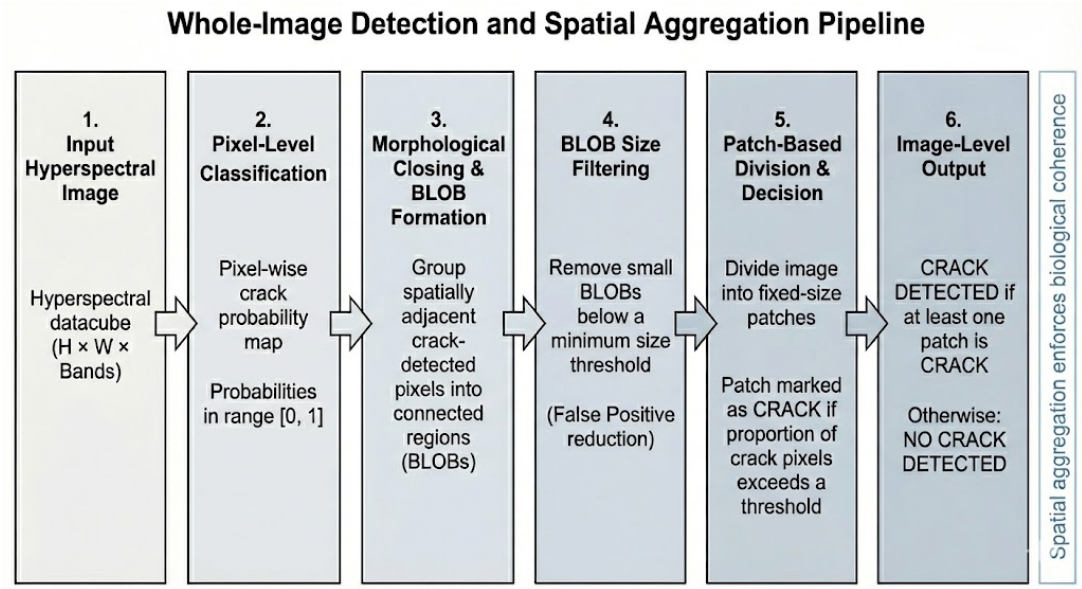


Figure 4.7: Whole-image crack-detection pipeline from pixel-level CRACK probabilities to image-level classification via spatial aggregation.

Calibration and Evaluation Stratification

The spatial hyperparameters for the aggregation stages (e.g., morphological kernel sizes, minimum BLOB area, and patch density thresholds) were optimized exclusively on the **Row 1** dataset using the **Optuna** hyperparameter optimization framework. By automating the search space exploration on the calibration set, this approach avoids manual tuning bias. Finally, to evaluate robust generalization across the timeline of the disorder, the frozen parameters were applied blindly to the **Row 2** held-out set and stratified by disease progression (**Early-stage** vs. **Late-stage**), directly addressing Hypothesis 3 and Hypothesis 4.

4.3.6 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 §4.3.3.

Spectral Preprocessing

Pixels were preprocessed as described in §4.3.1 (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 (§4.3.3).

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. The random seed supplied to the XGBoost model was kept fixed, while the seed controlling the grouped 80/20 train–test data partition was varied (split seeds 1–5). This approach evaluates the robustness of the wavelength selection process against variations in the data splitting.

4.3.7 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.

4.3.8 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.

Chapter 5

Experimental Results

5.1 Overview of Experimental Results

This chapter presents the experimental evaluation of the hyperspectral crack-detection framework. The findings are structured to systematically address the research hypotheses formulated in Chapter 3, progressing from fundamental spectral characterization to operational field-level detection. Specifically, the chapter is divided into four main sections:

1. Validation of the physical and spectral distinguishability between healthy and cracked tissues.
2. Pixel-level classification benchmarking.
3. Whole-image spatial aggregation and detection under early- and late-stage field conditions.
4. Wavelength selection analysis to determine the minimal spectral bands required for robust inference.

Presentation Note: For clarity and readability, the performance metrics reported in the main text and tables are presented as mean values truncated to three decimal places. Comprehensive cross-validation statistics, including standard deviations and fold-level data, are provided in Appendix A.1.

5.1.1 Evaluation Baselines

To ensure a rigorous and transparent evaluation, all experimental results are benchmarked against the following explicit baselines:

- **Pixel-Level Classification Baseline:** The reference model is **XGBoost** trained on the **full SNV-normalized hyperspectral signature** (450–925 nm, 159 contiguous bands), as defined in Section 4.3.3. 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 detailed in Appendix A.

- **Whole-Image Crack-Detection Baseline:** System-level detection is evaluated using the spatial aggregation pipeline described in Section 4.3.5. The baseline performance for this task relies on probability maps generated by the full-spectrum (159-band) XGBoost pixel-level model, establishing the upper bound for detection accuracy prior to any wavelength reduction. Early- and late-stage calibration and held-out test performances are summarized in Table 5.7.

- **Wavelength Selection (BFS) Reference:** The efficacy of Backward Feature Selection (BFS) is quantified by comparing the classification performance of reduced wavelength subsets directly against the full 159-band baseline. Elimination trajectories are illustrated in Figures 5.3 and 5.4, and whole-image test-set results at critical wavelength thresholds are reported in Table 5.10.

5.2 Spectral Separability and Informative Wavelengths

Prior to evaluating machine learning algorithms, a per-band spectral separability analysis was conducted to verify that healthy and cracked tissues exhibit physically distinguishable signatures, addressing Hypothesis 1. For this fundamental physical characterization, the relative reflectance spectra were analyzed *prior* to the application of per-pixel SNV normalization in order to preserve and quantify the absolute amplitude differences between the classes.

5.2.1 Mean Spectral Signatures of Healthy and Cracked Tissue

The mean spectral reflectance signatures ($\pm$ SD) for healthy and cracked grape tissues across the VIS–NIR range are presented in Figure 5.1. Cracked tissue consistently exhibits lower absolute reflectance across the full spectrum compared to intact healthy tissue. As visually evident in the plot, this divergence is exceptionally pronounced in the Near-Infrared (NIR) region, specifically between 700 nm and 850 nm.

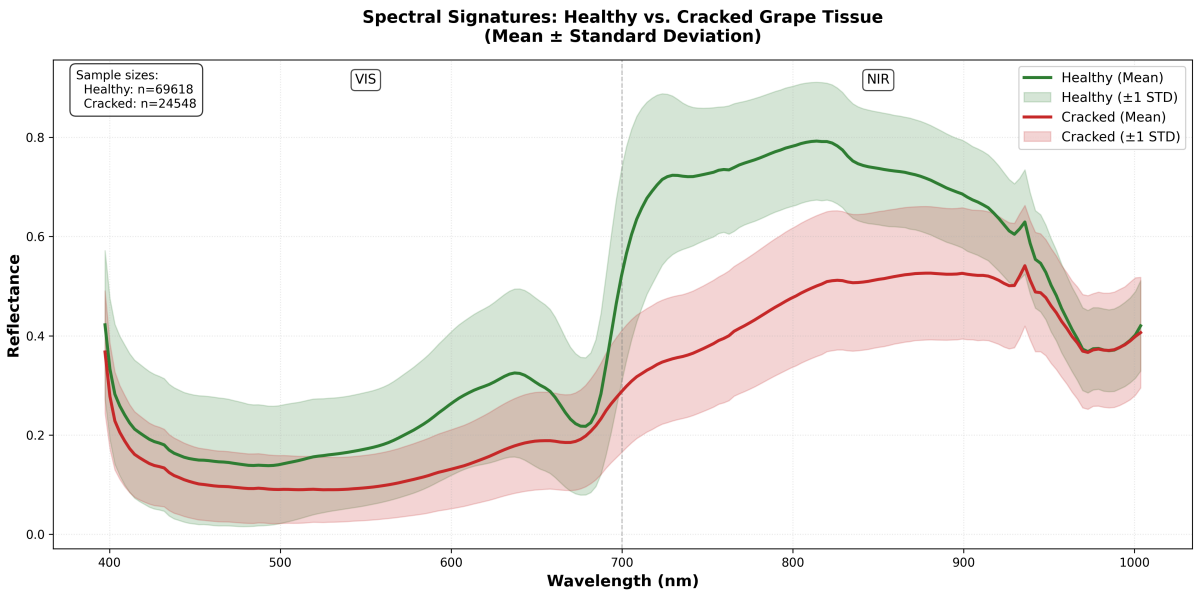


Figure 5.1: Mean spectral signatures ($\pm$ standard deviation) for healthy and cracked tissue across the VIS–NIR range.

5.2.2 Spectral Separability Across Wavelength Regions

To systematically quantify the discriminatory power of individual spectral bands, three complementary metrics were computed (Figure 5.2): **Absolute Mean Difference** ($|\Delta\mu|$) to capture the raw reflectance gap, **Cohen's** d **Cohen, 1988** to measure the standardized effect size, and the **Fisher Score Duda et al., 2001** to evaluate the ratio of inter-class separation to intra-class spread.

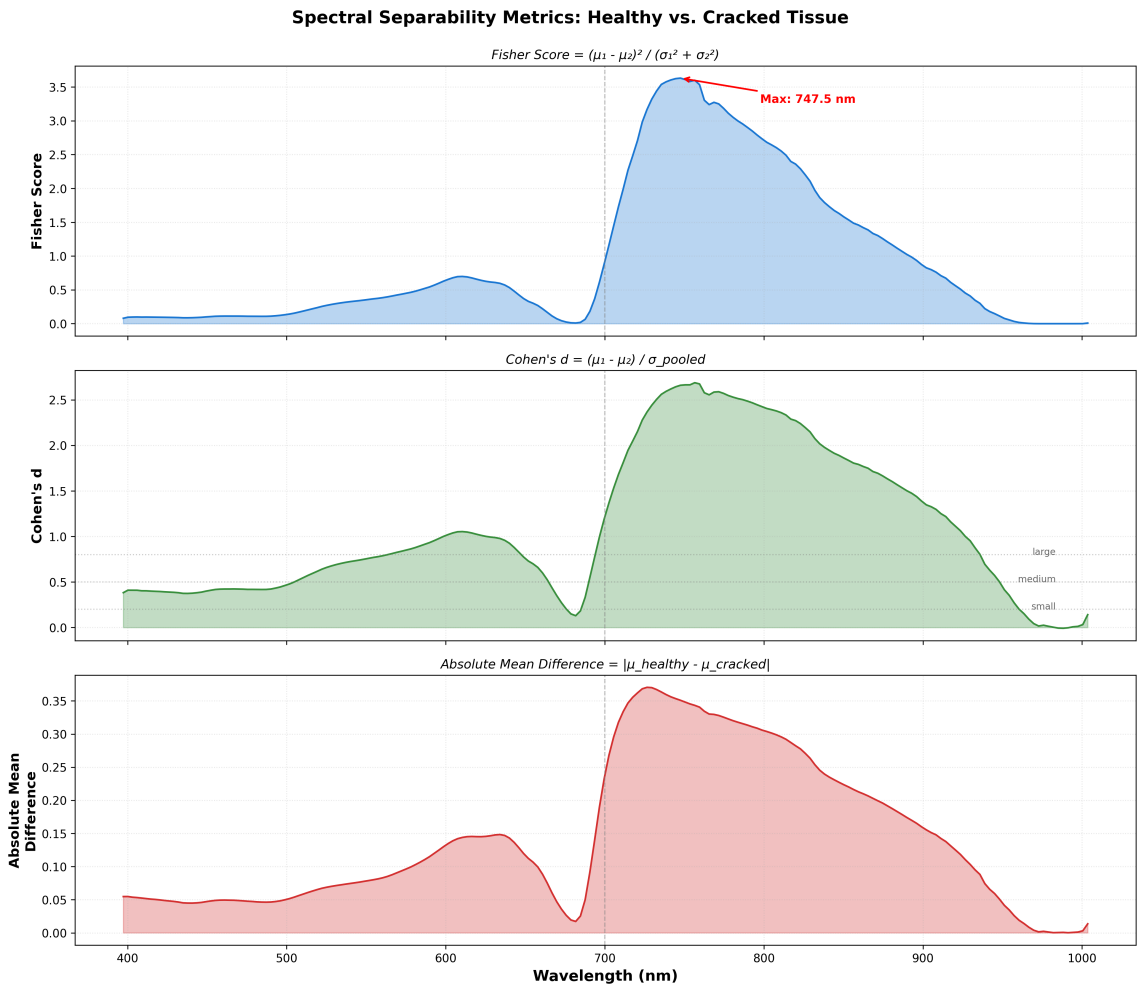


Figure 5.2: Wavelength-wise spectral separability metrics between healthy and cracked tissue, including absolute mean difference, Cohen's d, and Fisher score.

The separability metrics indicate that class separation is substantially higher in the NIR region, with peak discriminability consistently observed between 727–757 nm. As detailed in the regional summaries in Appendix A.2, healthy tissue reflectance exceeds that of cracked tissue by an average of 32% in the NIR and 37% in the VIS range. These clear physical differences confirm Hypothesis 1 and provide the basis for automated classification.

5.3 Pixel-Level Binary Classification

Supervised pixel-level classification addresses Hypothesis 2, following the LOGO cross-validation protocol described in Section 4.3.2. PR–AUC$_{\text{CRACK}}$ is the primary metric.

Table 5.1 reports performance across LOGO folds (Section 4.3.2; 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

While binary classification validates baseline spectral separability, operational field detection requires robustness against severe background clutter. To evaluate this, the detection task was expanded to a 10-class taxonomy encompassing diverse vineyard elements (Section 4.3.3). Performance was assessed across three labeling strategies: *crack_regular_rest*, *crack_vs_rest*, and *multi_class*. Balanced-training results are detailed 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_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 results demonstrate that supervised models maintain high discriminative power even in the presence of complex background materials. The *multi_class* formulation yielded the highest overall performance, with Logistic Regression, Random Forest, and XGBoost all exceeding a PR–$\text{AUC}_{\text{CRACK}}$ of 0.990 (Table 5.2). This confirms that crack signatures remain distinct from background clutter, justifying the use of the trained multi-class XGBoost model as the core engine for the whole-image aggregation pipeline.

5.5 Autoencoder-Based Anomaly Detection

To explore whether crack damage could be identified without the need for exhaustive pixel-level annotation, an unsupervised anomaly detection framework was evaluated (Section 4.3.4).

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

In the one-class binary setting, the autoencoder achieved PR–AUC$_{\text{CRACK}}$ ranging from 0.925 to 0.955 (Table 5.3). However, expanding

this to multi-class environments proved ineffective: training exclusively on CRACK pixels resulted in trivial reconstruction (Table 5.4), while training on all non-CRACK classes collapsed performance ($\text{PR–AUC}_{\text{CRACK}} \approx 0.01$, Table 5.5).

Overall, autoencoder-based formulations yielded substantially lower precision-recall performance at the pixel level compared to the supervised XGBoost baseline. Despite this performance gap, the best-performing binary autoencoder was carried forward to the whole-image detection stage to serve as an unsupervised system-level baseline. However, due to its superior discriminative power, the supervised multi-class XGBoost model was selected as the primary engine for the core evaluations and all subsequent wavelength selection (BFS) experiments.

5.6 Whole-Image Crack Detection

Having established the core predictive engine at the pixel level, the evaluation transitions to the primary operational objective: detecting cracking within full spatial scenes acquired under field conditions. This section specifically addresses Hypothesis 3 (robustness to background clutter) and Hypothesis 4 (early-stage detection capability). System performance was evaluated using the spatial aggregation pipeline (Section 4.3.5) and stratified by early- and late-stage acquisitions. The dataset composition for this phase is detailed in Table 5.6.

5.6.1 Supervised Detection Performance

The baseline for whole-image comparison is the full-spectrum XGBoost aggregation pipeline (159 bands). Table 5.7 presents the detection metrics on the spatially held-out Row 2 test set, compared against the Row 1 calibration set.

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

The results on the independent test set demonstrate robust system-level performance. Under late-stage conditions (advanced cracking), the supervised pipeline achieved a high $F1_{CRACK}$ score of 0.818. More importantly, under early-stage conditions–where cracks are minimal and visually indistinct–the system maintained strong predictive capability with an $F1_{CRACK}$ of 0.704. The expected performance drop between late and early stages reflects the physical difficulty of detecting incipient micro-cracks. Nevertheless, these results confirm both Hypothesis 3 and Hypothesis 4: the spatial aggregation logic successfully suppresses background noise, enabling reliable early-warning triggers on entirely unseen vines.

5.6.2 Unsupervised Baseline Comparison

To verify whether the supervised XGBoost engine was strictly necessary for whole-image inference, the best-performing unsupervised autoencoder was evaluated using the exact same spatial aggregation pipeline (Table 5.8).

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

The autoencoder-based pipeline yielded substantially lower performance across both stages on the test set (Late: $F1_{CRACK} = 0.698$; Early: $F1_{CRACK} = 0.448$). The sharp degradation in early-stage detection highlights that unsupervised reconstruction error is insufficiently sensitive to the subtle spectral shifts of incipient cracking. Consequently, the supervised XGBoost model is confirmed as the required foundation for the system, and is solely utilized in the subsequent wavelength selection experiments.

5.7 Wavelength Selection via Backward Feature Selection

Having demonstrated robust whole-image detection using the full 159-band spectrum, the final evaluation phase addresses Hypothesis 5: whether comparable performance can be maintained using a significantly reduced subset of spectral bands. To explore this, Backward Feature Selection (BFS) was applied to the pixel-level XGBoost model using the protocol described in Section 4.3.6, with the elimination trajectory guided by the PR–AUC$_{\text{CRACK}}$ metric.

5.7.1 BFS Trajectories and Subset Selection

The elimination trajectories for both the primary objective (PR–AUC$_{\text{CRACK}}$) and the corresponding F1 score are illustrated in Figure 5.3. The curves illustrate a high degree of spectral redundancy; performance remains remarkably stable even as the majority of the original 159 wavelengths are discarded. This suggests that the key physiological markers of grape cracking are tightly concentrated within specific spectral regions rather than distributed across the entire contiguous spectrum.

To systematically define candidate subsets for the whole-image pipeline, performance drop thresholds of 0.5% and 1.0% below the maximum PR–AUC score were established (Figure 5.4). As summarized in Table 5.9, the more conservative 0.5% tolerance yields a 30-wavelength subset, while the 1.0% tolerance isolates an extremely compact 11-wavelength subset.

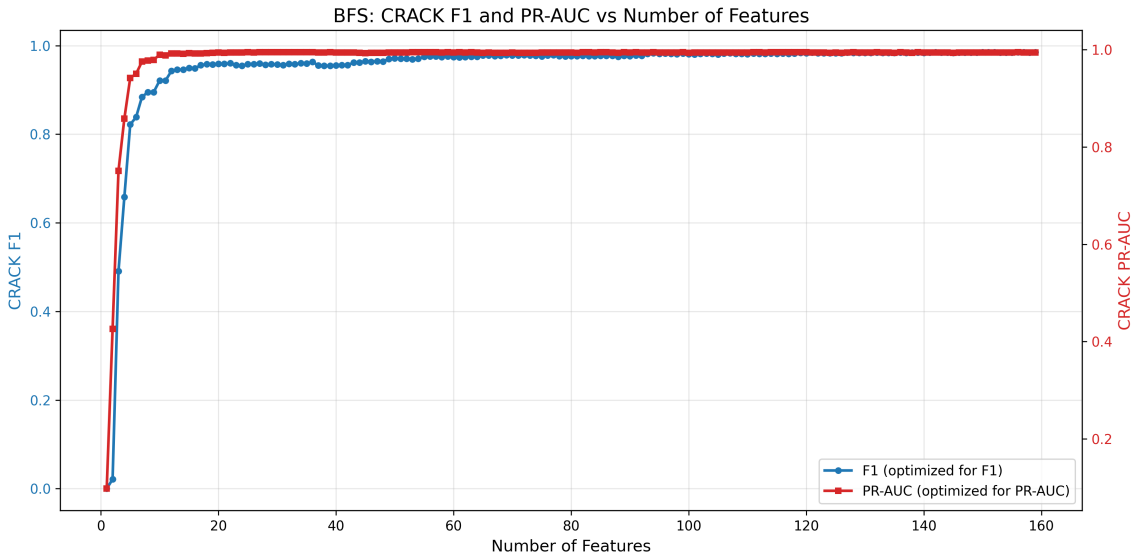
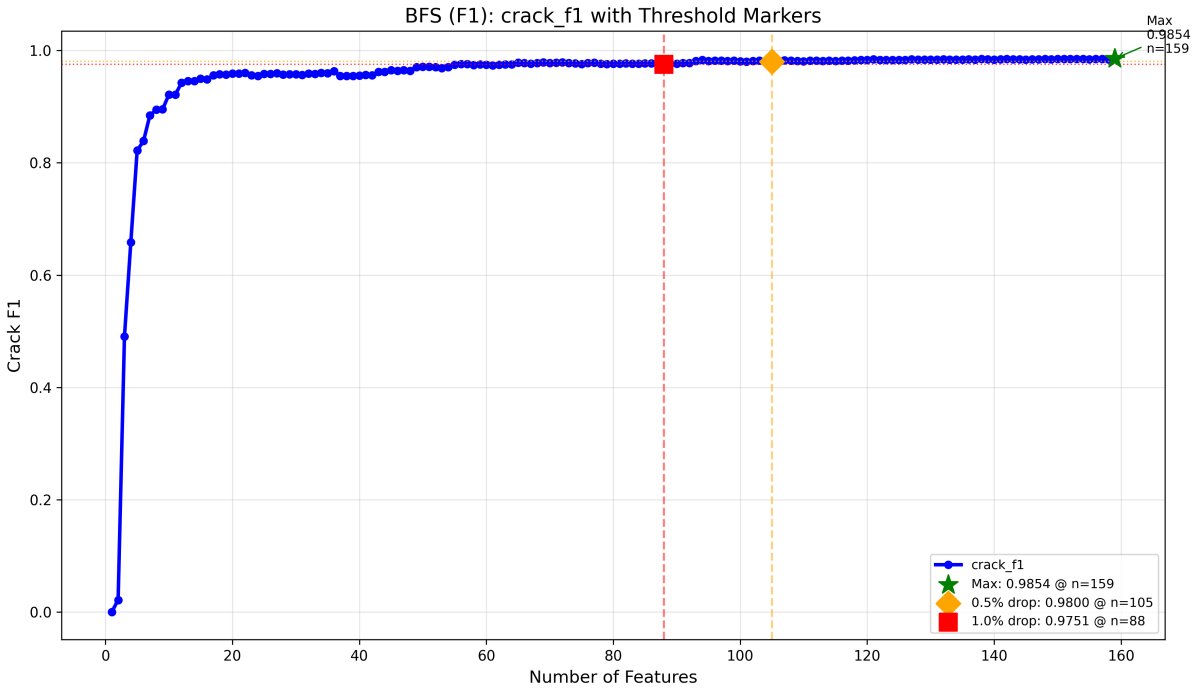
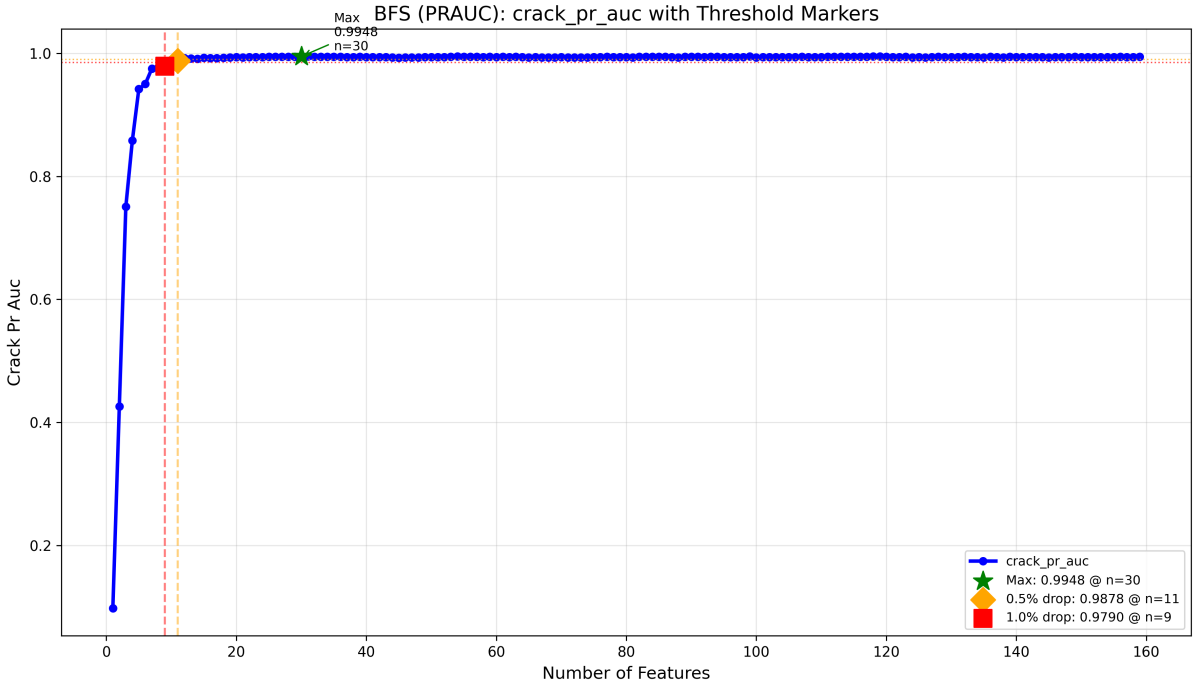


Figure 5.3: BFS trajectories for the whole-image pipeline: CRACK-class PR–AUC (primary objective) and CRACK-class F1 versus the number of retained wavelengths (n; full baseline $n = 159$).



(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.

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

5.7.2 End-to-End Wavelength-Selection Results

To validate the operational efficacy of these reduced subsets, the top-30 and top-11 configurations were integrated into the whole-image aggregation pipeline and evaluated on the held-out Row 2 test set. Table 5.10 compares their performance against the full 159-band baseline under both early- and late-stage conditions.

Table 5.10: Full-image crack-detection performance for different wavelength subsets under early- and late-stage conditions.

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 results strongly confirm Hypothesis 5. Notably, the top-30 subset achieves an $\mathrm{F1}_{\mathrm{CRACK}}$ of 0.813 under early-stage conditions, effectively matching and even slightly exceeding the 159-band baseline ($F1$=0.800). This slight improvement suggests that removing noisy or redundant spectral bands can actually enhance model generalization. Furthermore, even the extremely compact top-11 subset maintains robust late-stage detection ($F1_{CRACK} = 0.840$), demonstrating that the core discriminative information is preserved despite a 93% reduction in spectral dimensionality.

5.7.3 BFS Wavelength-Selection Stability Analysis

Beyond raw detection performance, the physical reliability of the selected bands was evaluated through a cross-seed stability analysis. By varying the train-test data split across five random seeds, the spatial and spectral consistency of the selected subsets was quantified.

Wavelength Popularity and Stability

The top-30 wavelength subsets demonstrated strong cross-seed consistency, predominantly converging on specific spectral regions in the near-infrared (NIR) portion of the spectrum (Figures 5.5 and 5.6). This physical consistency is maintained even under coarser binning schemes, indicating that the selection algorithm successfully targets robust tissue characteristics rather than overfitting to dataset-specific noise. Detailed binning and heatmap visualizations for the top-30 subsets are provided in Appendix A.7.

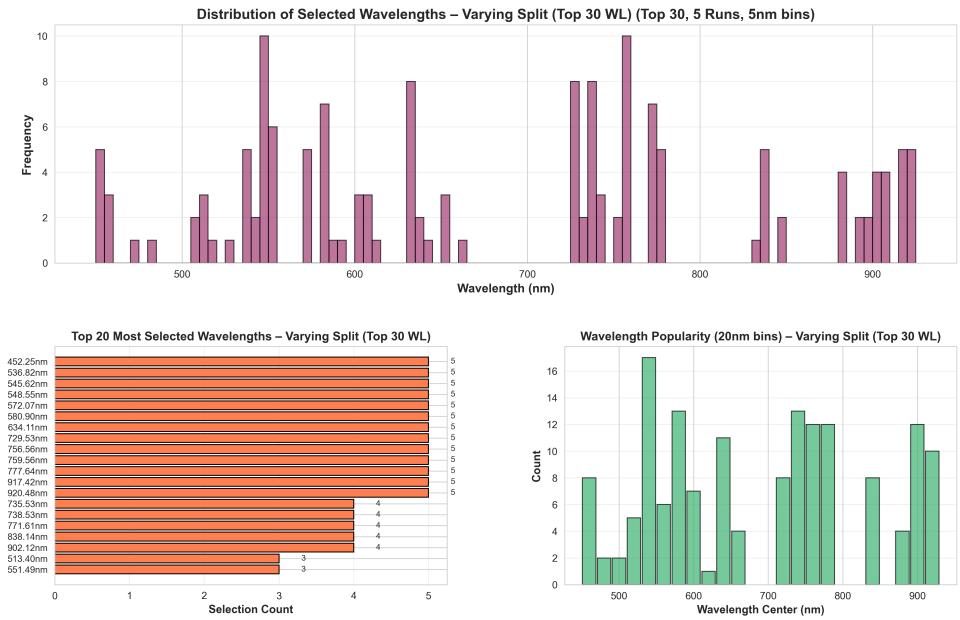


Figure 5.5: Distribution and popularity of the top-30 selected wavelengths across five varying-split seeds.

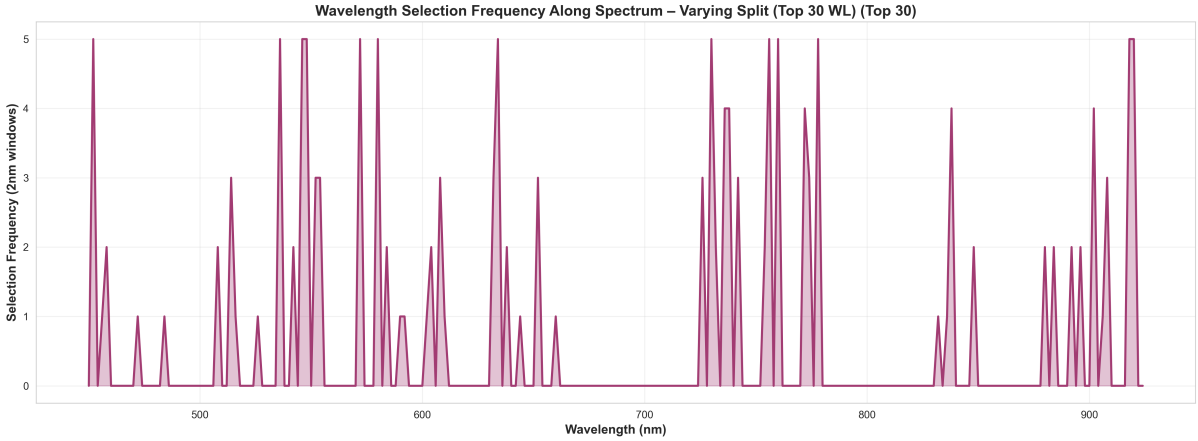


Figure 5.6: Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-30 subsets.

As expected with aggressive dimensionality reduction, the top-11 subsets exhibited more variance across seeds. This occurs because multiple highly correlated adjacent bands often act as interchangeable proxies for the same underlying physical trait. Nevertheless, consistent spectral regions remain clearly identifiable, supporting the feasibility of targeted multispectral sensor design. (Detailed popularity, binning, heatmap, and frequency spectrum visualizations for the top-11 subsets are provided in Appendix A.7).
Overall, the BFS procedure revealed a highly consistent selection pattern, heavily concentrated in the NIR region (700–850 nm). Specific anchors, such as 730 nm and 750 nm, were repeatedly identified as primary contributors to model performance, perfectly aligning with the peak spectral separability metrics established earlier in this chapter. Even under the extreme constraints of the 11-wavelength subset, the selection reliably converged on key markers like 452 nm (visible blue) and 729 nm (NIR-edge) across diverse data splits. While these findings definitively validate the statistical importance of these specific bands for robust crack detection, a deeper interpretation of their physiological relevance and biomechanical implications is reserved for the Discussion (Chapter **??**).

Chapter 6

Discussion and Conclusions

6.1 Physiological and Pre-processing Foundations

6.1.1 The Biological Basis of Spectral Signatures

The initial finding of this study—that cracked tissues exhibit significantly lower reflectance across both the VIS and NIR regions compared to healthy tissues—is strongly rooted in the biomechanical and physiological consequences of fruit cracking. In the VIS region (400–700 nm), the depressed reflectance of cracked tissue is primarily driven by the rupture of the waxy cuticle and the subsequent oxidation of sub-epidermal cells. Once this protective barrier is breached, exposed phenolic compounds undergo rapid oxidative browning Santos et al., 2023, which substantially increases light absorption. Furthermore, the loss of the smooth cuticle layer eliminates specular reflectance, causing the cracked area to appear darker. In the NIR region (700–850 nm), where the spectral separability is most pronounced, the governing factors are cellular structure and water content Thenkarail et al., 2016. Healthy mesocarp tissue acts as a strong diffuse reflector due to internal scattering at air–cell interfaces and high turgor pressure. Cracking induces a localized collapse of this cellular architecture; the rapid loss of turgor and leakage of fluids destroy the scattering interfaces, resulting in a dramatic drop in NIR reflectance Chang et al., 2019; Yu et al., 2025.

6.1.2 Data Normalization and Pre-processing Restraint

While these physiological changes establish a strong theoretical basis for detection, capturing them in uncontrolled vineyard conditions required specific pre-processing interventions to mitigate both sensor-level and field-induced noise. First, the usable spectral range was deliberately restricted to 450–925 nm (yielding 159 contiguous bands). The extreme boundaries of the camera's full spectral range suffer from low signal-to-noise ratios (SNR) and high variance—a known hardware limitation in VIS-NIR hyperspec-

tral sensors that is corroborated by prior studies Behmann et al., 2018; Thenkarail et al., 2016. Truncating these noisy extremes prevented the classifiers from fitting to instrumental artifacts. Second, the extraction of tissue segments (ROIs) for the pixel-level models was explicitly designed to exclude the outer boundary pixels using a morphological erosion strategy. This spatial buffer was critical to prevent the inclusion of *mixed pixels* Bioucas-Dias et al., 2013, where spectral information from the grape tissue blends with background elements or adjacent berries, corrupting the pure class signatures. Third, the highly variable illumination in the field—exacerbated by canopy shading and diverse berry acquisition angles—necessitated integration times that often deviated significantly from the camera's optimal recommendations. To prevent the models from overfitting to these lighting artifacts (e.g., underexposure or shadowing) rather than the underlying physiology, Standard Normal Variate (SNV) transformation was applied. By normalizing each pixel's spectrum relative to its own mean and variance, SNV effectively decoupled the absolute illumination intensity from the spectral shape. This ensured that the classifiers learned the true structural and chemical manifestations of the cracks, compensating for instances of over- or under-exposure within the same class.

It is also worth noting the deliberate restraint exercised in the preprocessing pipeline. While early exploratory experiments tested spatial smoothing techniques (such as moving averages across neighboring pixels) to reduce localized noise, these were ultimately discarded. Applying aggressive smoothing in tandem with SNV normalization artificially inflates spatial collinearity and distances the dataset from its raw physical measurements. The adopted pipeline intentionally prioritized minimal, physically justified transformations over heavy mathematical manipulation, ensuring that the spectral signatures remained faithful to the underlying tissue physiology.

6.2 Managing Field Clutter: The Role of Spatial Aggregation

6.2.1 Overcoming Spectral Ambiguity: The Dry Wood Challenge

The transition from a controlled pixel-level environment to whole-image field scenes exposed the severe limitations of binary classification in complex agricultural settings. Initial attempts to classify pixels using a simple binary approach (crack versus non-crack) resulted in unacceptable levels of random noise. This is mathematically predictable: grouping the infinite optical heterogeneity of the vineyard background into a single "non-crack" class creates an ill-defined, highly unstable decision boundary. To impose order on this heterogeneity, the problem was reformulated into a 10-class scheme, explicitly modeling the most prevalent background elements (e.g., leaves, soil, wood, and plastic markers). This multi-class mapping significantly stabilized the feature space and reduced random pixel-level noise, yet localized false positives persisted in the full-image predictions.

Empirical observation of the whole-image classification maps revealed that the residual false positives were not entirely random; they were systematically driven by specific background materials, primarily dry woody branches (and occasionally plastic markers). Physiologically, this spectral confusion is highly logical. As established in Section 6.1, cracked tissue is characterized by depressed reflectance due to water loss and tissue oxidation. Dry wood shares these exact traits—it is devoid of cellular water and composed of brown, oxidized lignin—making its spectral signature in the VIS and NIR regions deceptively similar to that of a severe crack.

6.2.2 Geometric Filtering and Structural Regularization

To overcome this inherent spectral ambiguity, the detection pipeline was fundamentally shifted to incorporate spatial and geometric context, leveraging a key visual distinction: true berry cracks manifest as dense, concentrated "clouds" or blobs of pixels on the berry surface, whereas noise from branches typically appears as elongated lines, and other artifacts appear as scattered, isolated pixels.

A multi-stage spatial aggregation strategy was implemented to exploit this geometric difference. First, morphological filtering operations were applied to the pixel-level probability maps. This step effectively consolidated the dense pixel clouds into cohesive blobs (closing) while eliminating the isolated "salt-and-pepper" noise points generated by random artifacts. Subsequently, a patch-based density evaluation was introduced. By dividing the image into spatial patches and analyzing the local crack probability, the

system applied a strict density criterion: a region was only flagged as a true defect if the percentage of crack pixels within a patch exceeded a calibrated threshold.

It is important to note that this spatial filtering was intentionally kept minimalist, avoiding more complex geometric descriptors such as circularity, eccentricity, or linearity. While advanced shape analysis could theoretically separate cracks from branches with higher precision on the calibration set, optimizing multiple morphological thresholds forces the pipeline to memorize the specific artifact shapes present in the calibration row. This excessive parameterization leads to severe post-processing overfitting, destroying the model's ability to generalize to new data. By restricting the spatial logic to fundamental density and blob-extraction, the system acts as a structural regularizer—sacrificing marginal gains in calibration performance to preserve robustness when deployed on unseen vines.

Ultimately, this patch-density logic served as a robust structural filter. It successfully discarded the elongated, low-density pixel clusters associated with thin branches, ensuring that only concentrated, blob-like formations triggered a detection. While exceptionally thick dry branches occasionally satisfied this density threshold—remaining the primary source of whole-image classification errors—the spatial aggregation pipeline proved essential. It demonstrated that robust field-level detection cannot rely on spectral signatures alone, but must fuse pixel-level spectral predictions with macroscopic geometric rules.

Developing this spatial pipeline exposed a significant computational optimization challenge. Initial exploratory iterations relied on brute-force grid searches across an 8-dimensional hyperparameter space, encompassing the baseline pixel-probability threshold, minimum and maximum blob areas, circularity and solidity constraints, morphological kernel sizes, patch dimensions, and the global crack-percentage criterion. It rapidly became evident that brute-force exploration of this massive combinatorial space was computationally prohibitive and yielded sub-optimal local minima. Consequently, the experimental design transitioned to Bayesian optimization via Optuna. By intelligently modeling the hyperparameter landscape, Optuna efficiently converged on a robust operating point that maximized early-stage precision while maintaining spatial coherence. The resulting pipeline achieved substantially higher precision (0.864 on the early-stage test set) than a naive pixel-level approach would.

6.3 Dimensionality Reduction, Algorithm Selection, and Sensor Design

6.3.1 Feature Selection and Multispectral Sensor Design

A central objective of this research was to determine whether the complex, high-dimensional data produced by a hyperspectral camera could be distilled into a practical, cost-effective agricultural tool. The Backward Feature Selection (BFS) analysis provided a definitive and positive answer. The results demonstrated that the original 159-band spectrum contains massive redundancy, and that performance can be maintained using a significantly truncated subset of wavelengths.

The stability analysis across multiple random seeds revealed that the 30-wavelength configuration represents the optimal "sweet spot" for operational deployment. While the extreme 11-wavelength subset successfully detected cracks, its feature selection variance across seeds indicated instability due to the over-constraint of the feature space. Conversely, the full 159-wavelength cube introduced noise and risk of overfitting. The 30-wavelength subset achieved the highest early-stage independent test MCC (0.644) by balancing sufficient spectral resolution to separate cracks from background clutter, while discarding noisy, uninformative bands. This finding has immense practical implications for precision viticulture. In contrast to fragile and expensive full-range hyperspectral line-scan cameras, the robust spectral markers identified here—anchored by universally stable features closely aligned with key physiological traits, such as the Soret band (452 nm), green reflection (548 nm), anthocyanin absorption (580 nm), and the red-edge inflection (729 nm)—pave the way for manufacturing a dedicated, low-cost multispectral sensor. Utilizing 10 to 30 fixed optical filters targeting these exact physiological features, such a device could be easily mounted on UAVs or vineyard tractors for high-throughput screening.

6.3.2 Methodological Rationale: BFS and Evaluation Metrics

The methodological choice of Backward Feature Selection (BFS) over alternative dimensionality reduction strategies was fundamentally driven by the physical constraints of sensor design. Standard reduction techniques, such as Principal Component Analysis (PCA), were explicitly avoided. While PCA effectively compresses variance into a lower-dimensional space, it generates synthetic features (linear combinations of all original wavelengths). A PCA-based model cannot be translated into a physical, low-cost multispectral camera, as the hardware would still require the full 159-band hyperspectral sensor to compute the principal components. BFS, by contrast, isolates the exact, unaltered physical wavelengths required for manufacturing physical optical filters. Further-

more, BFS circumvents the pitfalls of Forward Feature Selection (FFS)—which struggles with the extreme multicollinearity of adjacent spectral bands—by starting with the full spectral context and meticulously pruning only the redundant features.

Furthermore, while heuristic search methods like GA can optimize high-dimensional spaces, their stochastic nature (relying on random mutation and crossover) produces non-deterministic subsets that fluctuate wildly between runs, hindering physiological interpretation. BFS, conversely, provides a deterministic elimination sequence and a strict feature ranking. This traceability is what enabled the identification of the performance "sweet spot" across the elimination continuum, ensuring that the selected subsets are not random mathematical artifacts, but stable, physically meaningful markers.

The explicit choice to optimize the BFS algorithm based on the Precision-Recall Area Under the Curve (PR–AUC) rather than the standard Receiver Operating Characteristic (ROC–AUC) was dictated by the extreme class imbalance inherent to the problem. In whole-image field scenes, cracked pixels typically constitute less than 0.001% of the dataset. Under such severe imbalance, ROC–AUC yields overly optimistic and misleading scores, as it heavily rewards the correct classification of the massive, trivial "background" majority. PR–AUC, conversely, isolates the performance on the rare positive class, ensuring that the selected wavelengths genuinely maximize crack detection rather than merely excelling at background rejection.

6.3.3 Classifier Selection and Spatial Validation

Within the supervised paradigm, the superiority of tree-based ensembles (XGBoost, Random Forest) over linear models (PLS-DA, Logistic Regression) underscored the highly non-linear and collinear nature of the multi-class spectral boundaries. However, the ultimate selection of XGBoost as the primary classifier was driven as much by operational scalability as by accuracy. During preliminary trials, Support Vector Machines (SVM) achieved competitive pixel-level accuracy. Yet, their inference complexity ($O(n^2)$ to $O(n^3)$) proved prohibitive for whole-image deployment. A single field scene contains over 260,000 pixels (512×512)—exceeding the total size of the extracted pixel-level training dataset. SVM inference on a single full image was computationally unfeasible for high-throughput screening. Tree-based architectures, alongside shallow Multi-Layer Perceptrons (MLPs), provided the necessary inference speed to make full-image field deployment practical, highlighting the critical transition from laboratory-scale accuracy to operational field constraints.

Crucially, this pixel-level robustness was validated using a strict Leave-One-Group-Out (LOGO) cross-validation strategy, where each group corresponded to a distinct physical grape cluster. In agricultural computer vision, standard random K-fold splits frequently

suffer from spatial data leakage, as adjacent pixels from the same physical object fall into both training and validation sets, artificially inflating performance. By enforcing a cluster-level separation during cross-validation, the reported pixel-level metrics represent true generalization to unseen biological samples rather than mere spatial memorization.

6.3.4 The Limits of Unsupervised Anomaly Detection

The algorithmic comparison between the supervised XGBoost model and the unsupervised Autoencoder architectures provided profound insights into the spectral structure of the dataset. Two distinct anomaly-detection paradigms were tested: training the autoencoder on all non-crack classes, and training it exclusively on the crack class. Training on the diverse non-crack background failed catastrophically. Because the "normal" background encompasses an enormous spectral variety (leaves, wood, soil, plastic, shadows), the autoencoder was forced to learn a highly generalized, expansive latent space. Consequently, its generalization capacity became entirely too broad; it successfully reconstructed unseen crack signatures simply because they fell within the interpolated boundaries of the diverse background manifold. As a result, the Mean Squared Error (MSE) for cracks rarely exceeded the predefined anomaly threshold.

Conversely, training the autoencoder exclusively on the crack class yielded significantly better pixel-level discrimination. By learning only the narrow, specific spectral manifold of a crack, the model became highly adept at reconstructing cracks while failing to reconstruct (producing a high MSE for) any other vineyard material. However, while analytically elegant, this approach proved brittle. It overfits to the exact crack signatures present in the calibration set, reducing its ability to generalize to novel defect variations in unseen vines. Ultimately, the supervised XGBoost succeeded precisely because it was forced to explicitly learn the decision boundaries separating the crack from a well-defined 10-class background, rather than relying on reconstruction failures.

It is also critical to acknowledge how the SNV transformation interacted with these models. While SNV was mathematically necessary to eliminate illumination disparities (as discussed in Section 6.1), it inherently discards absolute reflectance intensity (albedo), preserving only the spectral shape. This normalization severely handicapped the autoencoder. Forced to reconstruct only the normalized spectral geometry, the autoencoder reconstructed background elements that had drastically different physical brightness but incidentally similar spectral shapes to a crack with low MSE, effectively blinding the anomaly detector. This loss of absolute intensity information also exacerbated XGBoost's confusion between cracks and bright background artifacts (like plastic or reflective wood). Future pipelines might benefit from a hybrid approach, passing both the SNV-normalized spectrum and the absolute brightness vector to the classifiers.

6.3.5 Methodological Challenges and Data Curation Bottlenecks

Beyond the algorithmic findings, this study highlights a critical, often-underreported bottleneck in applied hyperspectral research: the massive data-engineering effort required to translate unstructured, high-dimensional field imagery into structured tabular data suitable for machine learning. In the context of an individual research project, this data curation phase constitutes the most labor-intensive and Sisyphean aspect of the pipeline.

The first major hurdle was the manual ground-truth annotation. Labeling required a demanding dual-phase approach: physically scouting and tagging cracked clusters in the vineyard under harsh field conditions, followed by painstaking post-processing to isolate and label the exact cracked berries within the visually complex, overlapping canopy of the digital scenes.

The second major challenge was cross-modal alignment and spatial processing. Operating on both RGB and Hyperspectral (HSI) modalities required flawlessly mapping coordinates between sensors with different resolutions and fields of view. Relying solely on automated scripts to segment an RGB image and crop the corresponding HSI region is highly prone to hidden geometric errors, such as unintended 90-degree rotations, mirroring, or scaling mismatches. Consequently, custom Graphical User Interfaces (GUIs) had to be developed from scratch for nearly every preprocessing and filtering task.

These custom visual tools were not merely conveniences; they were absolute methodological necessities. Visual verification was the only way to guarantee that the extracted hyperspectral segments aligned perfectly with the physical berries, and that the morphological filters were behaving as mathematically intended before feature extraction. This rigorous, manual visual-validation loop was the invisible foundation of the study, ensuring that the downstream classifiers learned from accurate, biologically meaningful data rather than processing artifacts.

Early in the research phase, fully automated unsupervised clustering algorithms (e.g., K-Means) were explored to bypass this manual annotation bottleneck. The objective was to automatically isolate cracked pixels as a distinct unsupervised cluster based on their spectral distance. However, this approach encountered a fundamental mathematical barrier: extreme class imbalance. Cracked tissue represents an infinitesimal fraction of the total field of view (often less than 0.001% of the scene's pixels). In high-dimensional hyperspectral space, unsupervised algorithms are overwhelmingly dominated by the variance of the majority classes (canopy, soil, and healthy berries). The rare crack signatures were consistently swallowed by broader clusters, making it methodologically impossible to define stable distance thresholds or reliably isolate the sub-population without prior labels. This empirical limitation confirmed that for highly imbalanced, real-world agricultural scenes, labor-intensive supervised annotation remains a mandatory prerequisite.

6.4 Conclusions and Future Work

6.4.1 Summary of Contributions

This study successfully demonstrated that visible–near-infrared (VIS-NIR) spectral imaging can detect table grape cracking, even under early-stage conditions prior to severe visual degradation. The physical mechanism driving this detection is deeply rooted in the biomechanics of the crack: the destruction of the waxy cuticle and subsequent cellular water loss produce a highly distinct, depressed reflectance signature, particularly in the NIR plateau and the red-edge inflection.

Moving from laboratory pixel-level analysis to uncontrolled whole-image field deployment exposed significant challenges related to background clutter and illumination variance. These challenges were systematically mitigated through a novel pipeline that combined a multi-class supervised learning approach (XGBoost) with geometric spatial aggregation (patch-density blob filtering). This spatial logic proved that while spectral data is necessary for identifying the chemical and structural changes of a crack, spatial context is mandatory for suppressing the noise of the agricultural environment. Furthermore, the successful reduction of the required spectral bands from 159 to a stable subset of 30 wavelengths proves the feasibility of transitioning this technology from a scientific research tool to an affordable, tractor-mountable multispectral sensor.

6.4.2 Future Research Directions

Future research should focus on four primary avenues to bridge the gap to commercial deployment. First, cross-cultivar validation is essential; the spectral markers identified for 'Scarlotta Seedless' must be tested against grape varieties with different cuticle thicknesses and anthocyanin profiles. Second, the development of a hybrid preprocessing method that retains absolute albedo information alongside normalized spectral shapes could significantly reduce false positives caused by dry wood and plastic artifacts. Third, the hardware implementation of the selected 30-wavelength configuration into a physical, real-time multispectral camera must be tested in dynamic vineyard conditions.

Finally, a highly promising direction involves deep-learning-based spatial filtering to refine the whole-image detection pipeline. Preliminary investigations during this study explored a multi-stage cascade: utilizing the spectral model to detect suspected defect blobs, extracting their centroids, applying the Segment Anything Model 2 (SAM2) to isolate the local structure, and feeding the corresponding RGB bounding box into a Convolutional Neural Network (CNN) to classify the object as 'grape' or 'non-grape' (e.g., dry wood). While the CNN achieved excellent accuracy on isolated calibration segments—testing various inputs such as tight bounding boxes, 5-pixel padded boxes, and fully masked

background segments—it suffered a severe performance drop when deployed on whole-image field scenes. This discrepancy highlights a critical domain shift: spatial models trained on cleanly extracted ROIs struggle with the overlapping geometries, shadows, and contextual noise of a full canopy. Nevertheless, the core concept of a spectral-spatial cascade—using the hyperspectral pixel model as a highly sensitive primary detector, and an RGB-based CNN as a geometric secondary filter—remains conceptually sound. Future iterations of this architecture, potentially utilizing end-to-end training, could effectively eliminate the remaining false positives and yield a near-perfect early-warning detection system.

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)

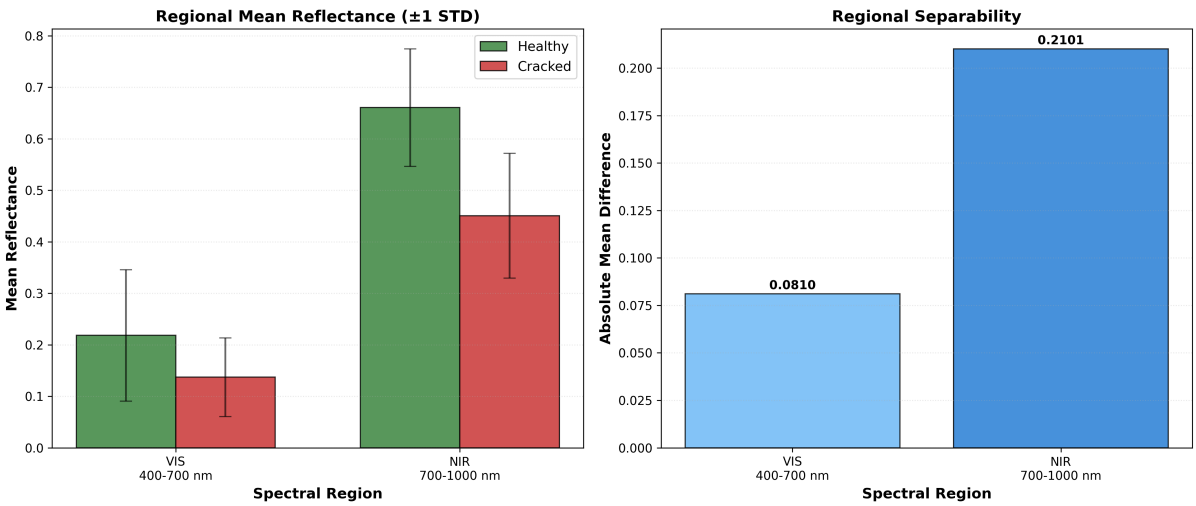


Figure A.1: Regional mean reflectance (±1 SD) and absolute mean class difference for the VIS and NIR ranges.

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.

A.7 Wavelength Stability: Supplementary Visualizations

This section contains the secondary wavelength popularity and stability figures removed from the main text for brevity. Figures A.2 and A.3 correspond to the top-30 BFS subsets; Figures A.4–A.7 correspond to the top-11 BFS subsets.

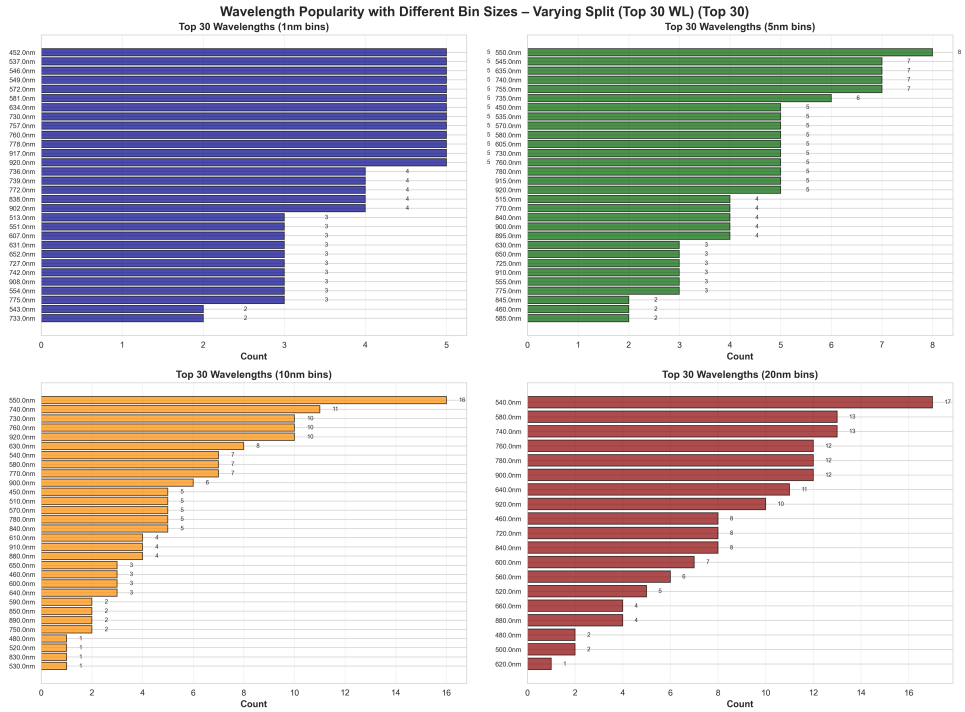


Figure A.2: Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-30 subsets.

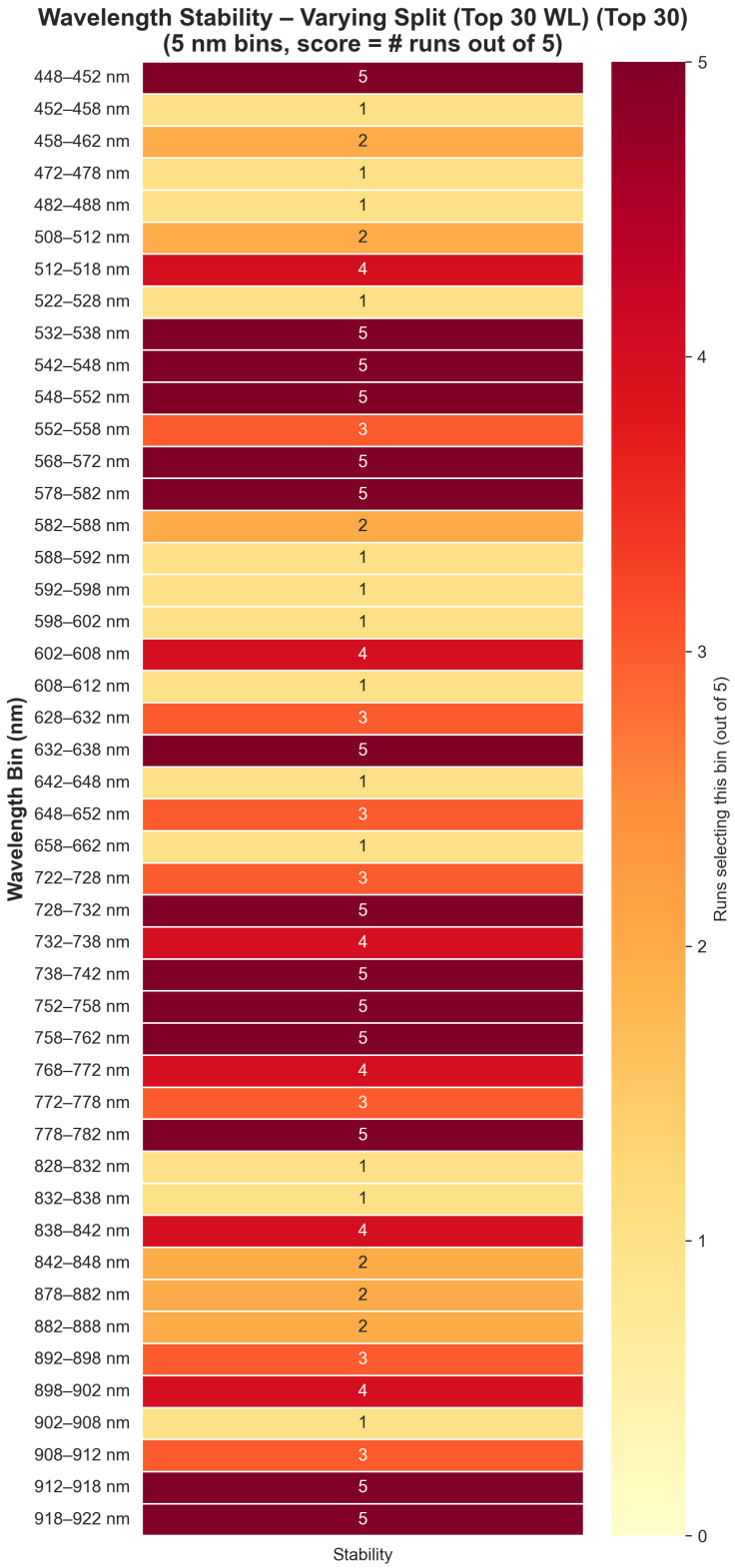


Figure A.3: Stability heatmap for the top-30 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$). Each cell indicates how many seeds selected a wavelength within that 5 nm bin.

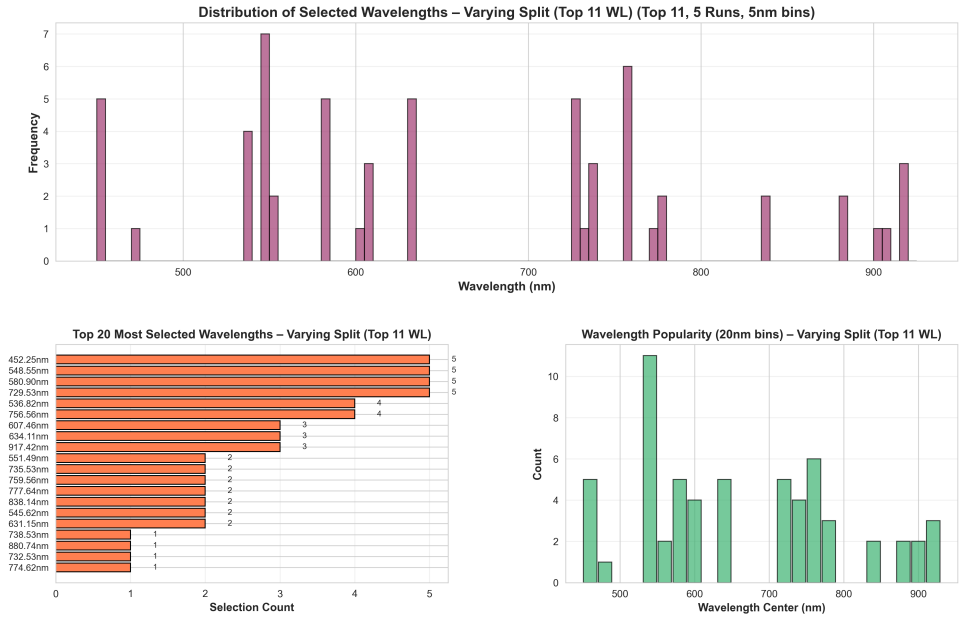


Figure A.4: Distribution and popularity of the top-11 selected wavelengths across five varying-split seeds.

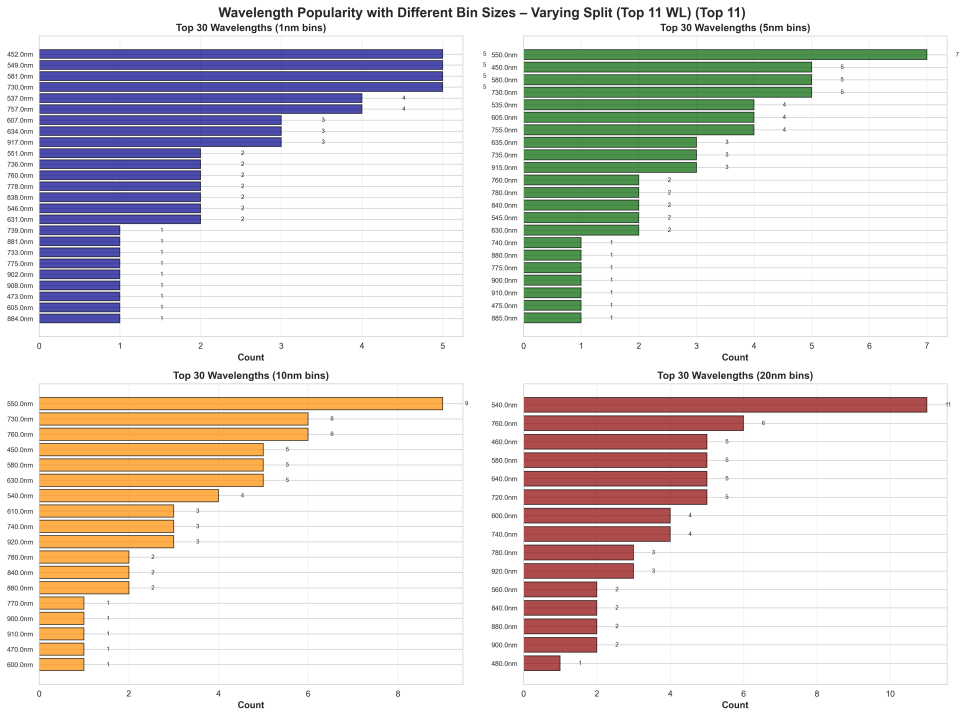


Figure A.5: Wavelength popularity at different bin sizes (1, 5, 10, 20 nm) for the top-11 subsets.

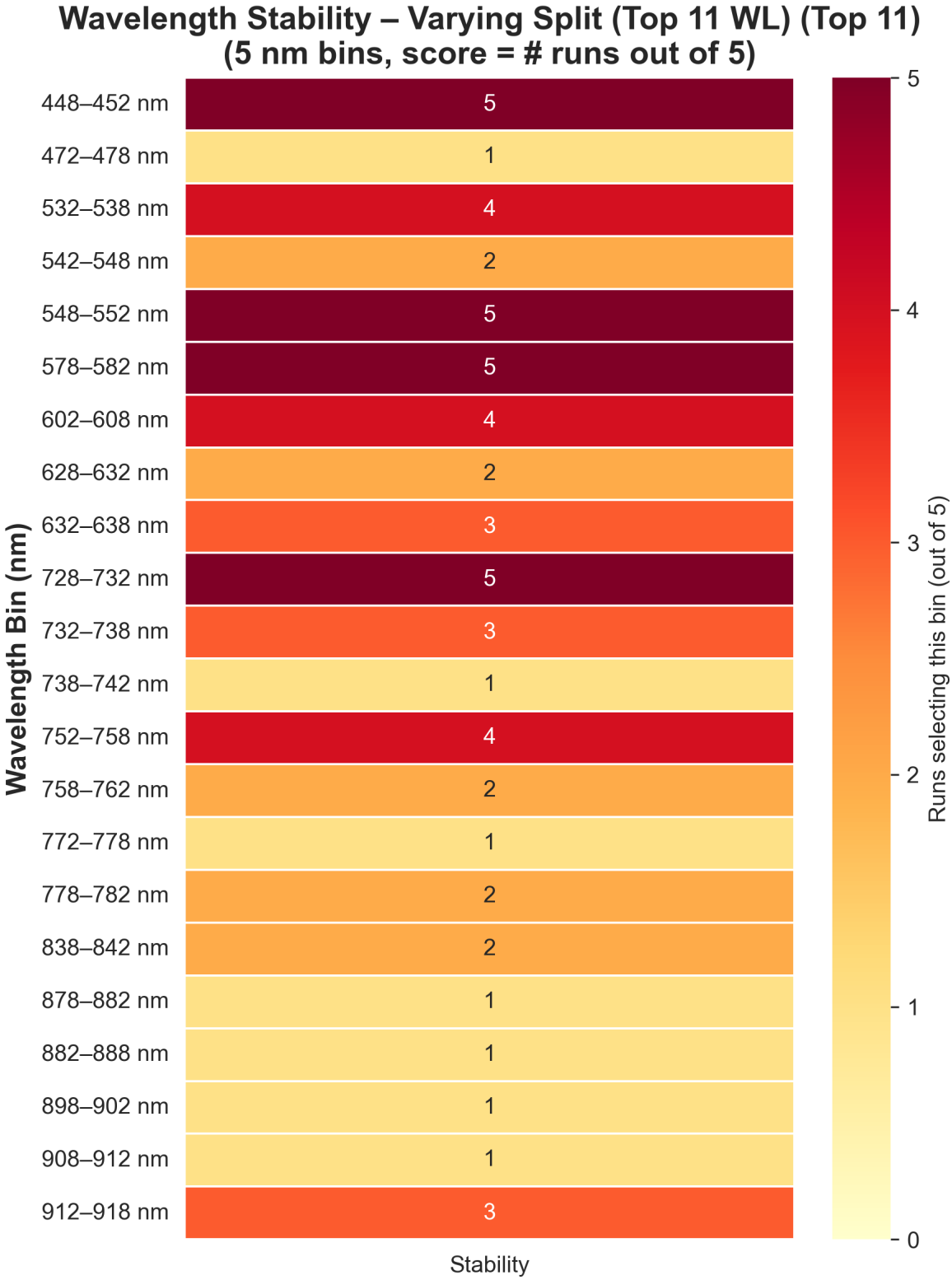


Figure A.6: Stability heatmap for the top-11 BFS subsets across five seeds (objective: PR–AUC$_{\text{CRACK}}$; 0.5% drop setting). Each cell indicates how many seeds selected a wavelength within that 5 nm bin.

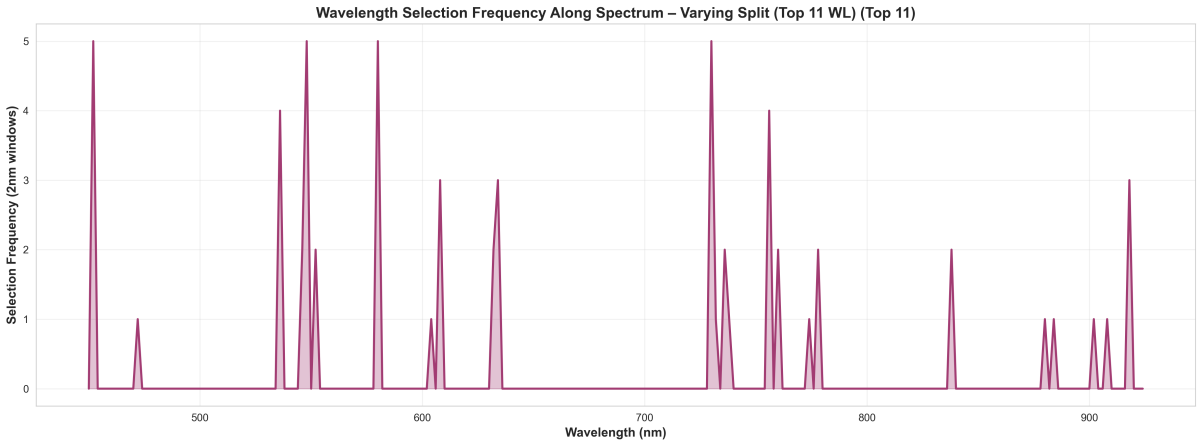


Figure A.7: Wavelength selection frequency along the spectrum (2 nm sliding window) for the top-11 subsets.

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.

Behmann, J., Acebron, K., Emin, D., Bennertz, S., Matsubara, S., Thomas, S., Bohnenkamp, D., Kuska, M. T., Jussila, J., Salo, H., et al. (2018). Specim iq: Evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors*, *18*(2), 441.

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.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., et al. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, *1*(2), 6–36.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Lawrence Erlbaum Associates.

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

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd). Wiley-Interscience.

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

Kaltenborn, J., et al. (2025). Data splitting to avoid information leakage with datasail. *Scientific Reports*, *15*(1), 1–15.

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

Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., & Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosystems Engineering*, *164*, 49–67.

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

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al. (2024). Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*.

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

Thenkarail, P. S., Lyon, J. G., & Huete, A. (2016). *Hyperspectral remote sensing of vegetation*. CRC Press.

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