

Early Detection and Prediction of Cracking in Grapes using Hyperspectral Imaging and Machine Learning

By Yovel Gani

Supervised by:

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

August 2025

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

Abstract

Grape cracking is a severe physiological disorder that leads to significant economic losses in the viticulture industry by reducing marketable yield and increasing susceptibility to secondary infections like sour rot. Early and non-destructive detection of cracking potential is critical for implementing preventative measures. This research explores the feasibility of using hyperspectral imaging (HSI) in the 400-1000 nm range, coupled with machine learning, for the identification and prediction of cracks in table grapes.

The study followed a multi-stage approach: initial identification, prediction modeling, and comparative analysis of spectral signatures. For the identification task, a pixel-level dataset was meticulously created by manually selecting and labeling pixels from healthy and cracked berries. An XGBoost classifier trained on these spectral signatures achieved outstanding performance, with an accuracy of approximately 97% and an AUC of 0.99, demonstrating that a distinct spectral difference exists between intact and cracked grape tissue.

However, applying this pixel-level model to whole-image segmentation proved challenging. The model produced noisy results, often misclassifying background elements such as plastic ties or dry twigs, which exhibited spectral signatures similar to those of decay associated with cracks.

To overcome this limitation and move towards prediction, the research shifted to analyzing the temporal evolution of mean spectral signatures of entire grape clusters under different irrigation regimes. This approach revealed significant and consistent trends over time, particularly in the Near-Infrared (NIR) region (700-900 nm). These findings suggest that changes in the average spectral signature can serve as a robust indicator of physiological stress and cracking susceptibility, potentially offering a reliable method for early warning. The results demonstrate that while pixel-level identification is highly accurate under controlled conditions, a macro-level analysis of mean spectral signatures provides a more practical and promising avenue for predicting grape cracking in complex field environments.

Contents

1	Introduction	6
1.1	The Problem of Fruit Cracking	6
1.2	Economic Impact	6
1.3	Research Motivation and Objectives	6
1.4	Hypotheses	7
1.5	Thesis Structure	8
2	Literature Review	9
2.1	Hyperspectral Imaging	9
2.1.1	Hyperspectral imaging across the electromagnetic spectrum	10
2.1.2	Reflectance	11
2.1.3	Spectral signature	12
2.2	Hyperspectral imaging in agriculture	12
2.2.1	Non-destructive Detection of Fruit Defects	12
2.2.2	Machine Learning for Hyperspectral Data Analysis	12
2.2.3	Physiological Mechanisms of Grape Cracking	13
2.2.4	Research Gap	13
3	Materials and Methods	14
3.1	Study Area and Experimental Design	14
3.2	Imaging Systems	14
3.3	Data Collection Protocol	14
3.3.1	Dataset Preparation for Pixel-Level Classification	15
3.3.2	Data Analysis and Modeling	17
3.3.3	Automated Signature Extraction for Temporal Analysis	17
3.4	Model Evaluation Strategy	18
3.4.1	Validation: Leave-One-Group-Out Cross-Validation	18
4	Results	20
4.1	Spectral Signature Analysis and Identification of Most Informative Wave- lengths	20

4.1.1	Mean Spectral Signature Analysis	20
4.1.2	Consensus-Based Feature Selection	21
4.2	Pixel-Level Classification Model Benchmark	23
4.2.1	Performance on the Unbalanced Dataset	23
4.2.2	Performance on the Balanced Dataset	24
4.2.3	Anomaly Detection Benchmark	24
4.2.4	Wavelength Sensitivity Analysis with LOGO Cross-Validation . .	25
4.2.5	One-Class Training with Autoencoder (LOGO)	29
4.3	Challenges in Whole-Image Classification	32
4.4	Temporal Trends in Mean Spectral Signatures	33
5	Discussion	34
6	Conclusions and Future Work	35
6.1	Conclusions	35
6.2	Future Work	35

List of Figures

2.1	Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.).	10
2.2	The formula for reflectance [?].	11
2.3	Hyperspectral signature of barley grains with varying moisture content [?].	12
3.1	The custom-built graphical user interface (GUI) for data annotation and analysis. The tool displays the Canon RGB image (right), the HSI camera's RGB representation (left), and a single HSI band (center). The bottom panel shows the spectral signature plots for a selected pixel, in this case from a healthy berry, resulting in a low AutoEncoder reconstruction error ($MSE = 0.0001$).	16
4.1	Mean spectral signatures of pixels sampled from 'Cracked' (Orange) versus 'Regular' (Blue) grape tissue. The shaded areas represent the standard deviation. The largest separation is evident in the NIR region (700-900 nm).	21
4.2	Mean spectral signatures with the top-ranked, high-consensus wavelengths highlighted by vertical dashed lines. The selected features are clearly concentrated in the NIR region, coinciding with the area of greatest spectral separation between cracked and regular tissue.	23
4.3	Visual validation of the one-class AutoEncoder for anomaly detection. The tool demonstrates the model's ability to differentiate between (a) normal and (b) anomalous spectral signatures in real-time based on the reconstruction error (MSE).	25
4.4	Performance vs. k for LDA and Logistic Regression (SFS, LOGO).	26
4.5	Performance (ROC AUC) vs. k for LDA and Logistic Regression (SFS, LOGO). The curve saturates near $k \approx 10-12$	27
4.6	Marginal gain (Δ ROC AUC) vs. k . Most of the improvement arrives with the first few wavelengths.	28
4.7	Cumulative Δ ROC AUC relative to $k=1$. Gains plateau around $k \approx 10-12$.	29
4.8	Fold-level variability (faint) and mean curves (bold) for Logistic Regression at $k=10$. Both ROC AUC and AP are very high, indicating excellent separability under cluster-level generalization.	30

4.9	Feature selection stability across k (Top-40 wavelengths by fold-frequency; SFS+LOGO). Dominant bands reside in the NIR (700–900 nm).	31
4.10	Full feature-frequency heatmap across all wavelengths.	32
4.11	Micro-averaged ROC curve for the one-class Autoencoder (LOGO). . . .	33
4.12	Confusion matrices on concatenated OOF predictions at the two operating points.	33

List of Tables

1.1	Estimated financial losses from fruit cracking in major crops.	7
4.1	The 19 most informative wavelengths, selected concurrently by all three feature selection methods (Fisher Score, Mutual Information, and Random Forest), forming the high-consensus feature set for model development. .	22
4.2	Model Performance on the Unbalanced Dataset.	24
4.3	Model Performance on the Balanced Dataset.	24
4.4	Mean performance across LOGO folds using SFS for selected k	26
4.5	Autoencoder performance on concatenated OOF predictions (LOGO) at two operating points.	31

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 and Objectives

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.

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.

Main Research Question: Can hyperspectral imaging and machine learning models be used to identify existing grape cracks and predict their future occurrence?

Specific 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. **Prediction:** To investigate methods for predicting cracking potential by analyzing temporal changes in the spectral characteristics of grapes over the ripening period.
3. **Comparative Analysis:** To compare the mean spectral signatures of grape clusters under different agricultural treatments (e.g., irrigation regimes) to identify spectral indicators of stress related to cracking.

1.4 Hypotheses

1. The spectral signatures of cracked grape tissue are significantly different from those of healthy tissue, allowing for accurate classification.
2. Changes in the mean spectral signature of a grape berry over time, particularly in the NIR region, can serve as an early indicator of physiological stress and increased cracking risk.
3. Different irrigation regimes will induce distinct and measurable changes in the temporal spectral signatures of grape clusters.

1.5 Thesis Structure

This thesis is organized as follows: Chapter 2 provides a review of the relevant literature. Chapter 3 details the materials and methods used for data collection and analysis. Chapter 4 presents the results of the classification and temporal analysis. Chapter 5 discusses the interpretation and implications of these results. Finally, Chapter 6 concludes the study and suggests directions for future research.

Chapter 2

Literature Review

2.1 Hyperspectral Imaging

The spectrum of light recorded by cameras varies significantly across different imaging technologies. An RGB camera records light reflected by an object using three bands of visible radiation: blue, green, and red (Benelli et al., 2020). Multispectral imaging typically includes between 3 to 15 distinct bands and has a spectral resolution greater than 10 nm. Alternatively, commercially available hyperspectral imaging (HSI) systems offer a more precise resolution (less than 5 nm) and encompass tens to hundreds of contiguous spectral bands, capturing a broader and more detailed spectrum (Mehta et al., 2018). It captures broad segments of the electromagnetic spectrum, from the ultraviolet (UV), around 200 nm, through the visible (VIS) light range from 380 nm to 800 nm, and extends into the near-infrared (NIR) up to 1700 nm, as illustrated in Figure 2.2. Additionally, it often includes the short-wave infrared (SWIR) from 970 nm to 2500 nm (Ball, 2007). This capability allows HSI to perform detailed material analysis, enabling detection of both surface and subsurface characteristics that are not visible through conventional imaging (Mehta et al., 2018). The VNIR spectrum, ranging from 400 to 1000 nm, is highly effective for assessing the health and composition of various materials. By providing detailed spectral data, this range facilitates the identification of variations in surface coatings and compositions, while the NIR and SWIR ranges offer deeper insights into the underlying chemical and physical structures (Zwinkels, 2015).

2.1.1 Hyperspectral imaging across the electromagnetic spectrum

The spectrum of light recorded by cameras varies significantly. An RGB camera records light using three bands of visible radiation. Multispectral imaging typically includes between 3 to 15 distinct bands. Alternatively, hyperspectral imaging (HSI) systems offer a more precise resolution (less than 5 nm) and encompass tens to hundreds of contiguous spectral bands. HSI captures broad segments of the electromagnetic spectrum, from the ultraviolet (UV), through the visible (VIS) and near-infrared (NIR), and often includes the short-wave infrared (SWIR). This capability allows HSI to perform detailed material analysis not visible through conventional imaging.



Figure 2.1: Segments of the electromagnetic spectrum commonly used in hyperspectral imaging (Middleton Spectral Vision, n.d.).

2.1.2 Reflectance

Reflectance is defined as the ratio between the radiation reflected by a surface and the radiation received by that surface. The ratio is affected by the object's chemical properties, the micro-topography of its surface, and the angle of the light source. Reflectance is the primary value captured at each pixel for every wavelength in a hyperspectral image.

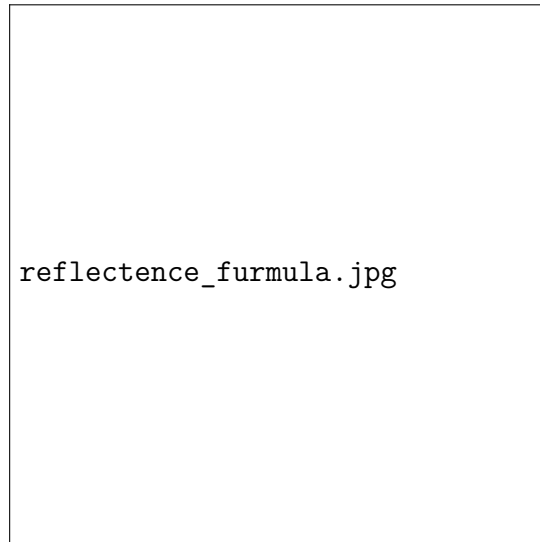


Figure 2.2: The formula for reflectance [?].

2.1.3 Spectral signature

As a certain material is exposed to a light source, it reflects particular portions of electromagnetic energy according to its chemical composition. This reaction is called the spectral signature of a material. Analyzing spectral responses allows for the identification and characterization of different materials [?].



Figure 2.3: Hyperspectral signature of barley grains with varying moisture content [?].

2.2 Hyperspectral imaging in agriculture

2.2.1 Non-destructive Detection of Fruit Defects

... (Review of studies on detecting bruises, diseases, and ripeness in various fruits like pomegranates, tomatoes, etc., using HSI and other imaging modalities) ...

2.2.2 Machine Learning for Hyperspectral Data Analysis

... (Discussion of common ML algorithms used for HSI, including traditional methods like SVM, KNN, and advanced deep learning models like 1D/3D-CNNs) ...

2.2.3 Physiological Mechanisms of Grape Cracking

... (Summary of the biological and environmental factors leading to cracking) ...

2.2.4 Research Gap

... (Conclude by highlighting that while HSI has been used to detect existing defects, its application for predicting a future physiological disorder like cracking based on temporal spectral trends is an underexplored area) ...

Chapter 3

Materials and Methods

3.1 Study Area and Experimental Design

The study was conducted in a commercial vineyard in the Lachish region of Southern Israel. The experiment involved 240 grape clusters selected across four vineyard rows. These clusters were subjected to five different irrigation treatments, with 12 clusters randomly assigned to each treatment within each row. Data collection was performed twice weekly from July to mid-August.

3.2 Imaging Systems

Three primary imaging systems were used:

1. **Hyperspectral Camera (Specim IQ):** Captured hyperspectral cubes in the 400-1000 nm range with a spectral resolution of 7 nm and spatial resolution of 512x512 pixels.
2. **RGB Camera (Canon R5):** Provided high-resolution visual context images.
3. **Thermal Camera (FLIR T1020):** Acquired thermal data to monitor canopy temperature.

3.3 Data Collection Protocol

A rigorous daily protocol was followed, starting at 05:30 at the Volcani Institute for equipment preparation, followed by 6 hours of imaging in the vineyard (07:00-13:00). A white reference panel was imaged periodically for hyperspectral data calibration. Each cluster's position was meticulously documented to enable temporal analysis.

3.3.1 Dataset Preparation for Pixel-Level Classification

To facilitate an efficient and accurate workflow for creating the pixel-level dataset, a custom graphical user interface (GUI) was developed using Python with the `PyQt5` library. This tool integrated data visualization, interactive annotation, and machine learning model inference into a single platform (see Figure 3.1). The workflow, which was used to generate the final dataset, proceeded as follows:

1. **Image Loading and Visualization:** The tool loaded and displayed a grape cluster’s hyperspectral (HSI) data cube alongside its corresponding high-resolution RGB context images (both from the HSI camera and a separate Canon camera). A slider mechanism allowed for interactive exploration of the HSI cube by cycling through its 204 spectral bands.
2. **Interactive Pixel Sampling:** The user manually selected a single “seed pixel” by clicking on a region of interest—either a clearly cracked berry or a clearly healthy berry—within any of the displayed images.
3. **Automated Segmentation with SAM2:** The selected seed pixel was passed as a point prompt to the **Segment Anything Model 2 (SAM2)**, which was integrated directly into the tool via its Python API. SAM2 then performed instance segmentation, generating a precise binary mask of the entire berry corresponding to the selected point. This semi-automated approach significantly accelerated the tedious process of manual segmentation.
4. **Data Extraction and Labeling:** With the berry segmented, the tool extracted the full spectral signatures for all pixels contained within the mask. The user then assigned a final label—either ‘Cracked’ or ‘Regular’—to the entire segment. Upon confirmation, the labeled spectral data was saved for inclusion in the training dataset.

This process was repeated for 50 cracked and 50 healthy berries, resulting in a dataset of approximately 125,000 labeled pixels ($\sim 45,000$ cracked, $\sim 80,000$ regular), forming the basis for the pixel-level classification models.

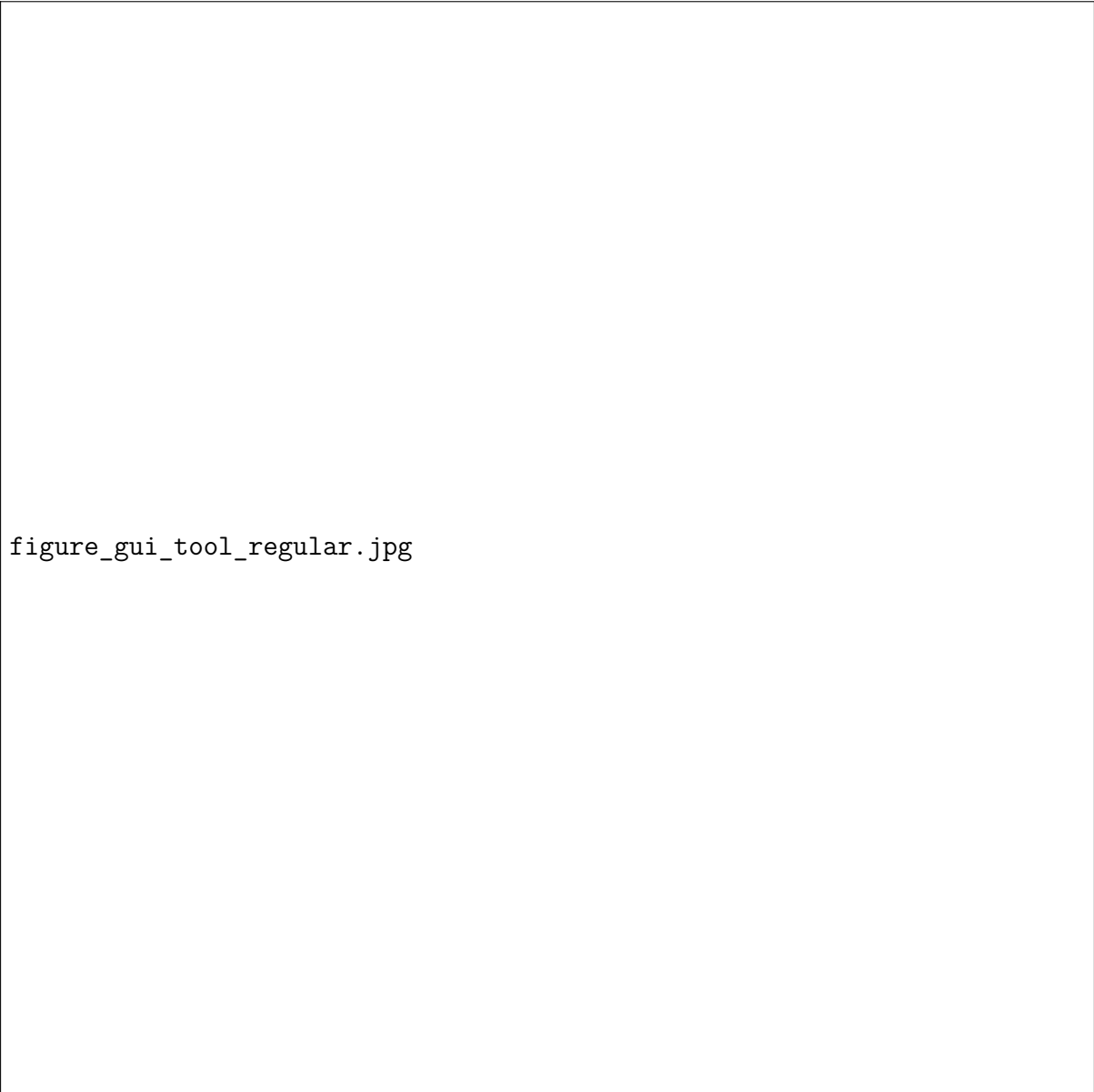


Figure 3.1: The custom-built graphical user interface (GUI) for data annotation and analysis. The tool displays the Canon RGB image (right), the HSI camera’s RGB representation (left), and a single HSI band (center). The bottom panel shows the spectral signature plots for a selected pixel, in this case from a healthy berry, resulting in a low AutoEncoder reconstruction error ($\text{MSE} = 0.0001$).

3.3.2 Data Analysis and Modeling

Feature Selection

To identify the most discriminative wavelengths for crack detection, three feature selection methods were employed:

- **Fisher Score:** To find wavelengths with the largest separation between class means relative to their variance.
- **Mutual Information:** To identify wavelengths that carry the most information about the class label.
- **Random Forest Feature Importance:** To rank wavelengths based on their contribution to the accuracy of a Random Forest classifier.

XGBoost Classification

For the pixel-level identification task, an XGBoost classifier pipeline was implemented:

1. Data Filtering: The spectral range was narrowed to 180 bands (484 nm to 933 nm).
2. Scaling: Per-sample min-max scaling was applied to normalize each spectral vector.
3. Data Splitting: A stratified 80/20 train/test split was used.
4. Model Training: An `XGBoostClassifier` was trained on the training set.
5. Evaluation: The model was evaluated on the test set using accuracy, confusion matrix, ROC curve, and Precision-Recall (PR) curve.

3.3.3 Automated Signature Extraction for Temporal Analysis

To investigate the prediction objective and analyze temporal changes in spectral characteristics across different irrigation regimes, a separate, fully automated data extraction pipeline was developed. This approach was designed to circumvent the labor-intensive process of manual segmentation and to capture a holistic spectral representation of the entire grape cluster over time. The methodology focused on analyzing the central region of each hyperspectral image, where the main body of the grape cluster was consistently located.

The process was implemented via a Python script that systematically processed all collected images. The core logic of the script is outlined below:

Methodological Steps:

1. **Batch Image Loading:** The script iterates through the main data directory, identifying all subfolders corresponding to different clusters and capture dates. For each entry, it utilizes a helper function, similar to the `load_hsi_image_from_folder` function found in the project’s backend, to load the hyperspectral (HSI) data cube into memory.
2. **Defining a Central Region of Interest (ROI):** For each HSI data cube (with dimensions $512 \times 512 \times 204$), a square Region of Interest (ROI) was defined at the geometric center of the image. A typical ROI size was set to 50×50 pixels. This size was chosen as a conservative estimate to ensure that the extracted pixels primarily belonged to the grape cluster, minimizing the inclusion of background elements like leaves or the sky, regardless of the cluster’s specific position in the frame.
3. **Mean Signature Calculation:** All pixel vectors within the defined central ROI were extracted. The script then computed the element-wise mean of these vectors, resulting in a single, average spectral signature of 204 bands for that specific cluster at that point in time. This averaging process effectively filters out spatial noise and provides a robust representation of the cluster’s overall spectral state.
4. **Data Aggregation and Grouping:** The calculated mean signature was stored along with its corresponding metadata, including the cluster ID, capture date, and the specific irrigation treatment group it belonged to. This process was repeated for every image collected throughout the season.
5. **Final Analysis:** After processing all images, the aggregated data was grouped by irrigation treatment and capture date. For each group at each time point, the final mean spectral signature and its standard deviation were calculated. This allowed for the direct comparison of temporal trends between the different irrigation regimes, as presented in Chapter 4.

This automated approach enabled the efficient processing of thousands of images and formed the foundation of the temporal analysis aimed at predicting cracking susceptibility based on physiological stress indicators.

3.4 Model Evaluation Strategy

3.4.1 Validation: Leave-One-Group-Out Cross-Validation

To robustly evaluate model generalization, a Leave-One-Group-Out (LOGO) cross-validation strategy was implemented. This method was chosen because the spectral pixel data is

naturally grouped by the grape cluster from which it was sampled. The LOGO process operates as follows:

- The dataset is split into N groups, where N is the number of unique grape clusters.
- In each iteration, one entire group (cluster) is held out as the test set.
- The model is trained on the remaining $N-1$ groups.
- The trained model is then evaluated on the held-out test group.
- This process is repeated N times, with each group serving as the test set exactly once.

The final performance metrics reported in Chapter 4 are the mean and standard deviation of the scores calculated across all N iterations.

Chapter 4

Results


4.1 Spectral Signature Analysis and Identification of Most Informative Wavelengths

The foundational step in developing a robust classification model was to conduct an in-depth exploratory analysis of the hyperspectral data. The primary objective was to determine whether a consistent and statistically significant spectral difference exists between pixels corresponding to healthy ('Regular') grape tissue and those corresponding to cracked tissue. This analysis serves not only to validate the feasibility of spectral classification but also to guide the subsequent, more granular process of feature selection.

4.1.1 Mean Spectral Signature Analysis

To visualize the fundamental differences between the two classes, the mean spectral signatures for all pixels labeled as 'Cracked' and 'Regular' were computed across the entire dataset. Figure 4.1 plots these mean signatures, with the shaded areas representing the standard deviation, providing a sense of the variance within each class.

The graph clearly illustrates that while the two signatures are nearly identical below 500 nm, a significant and consistent separation emerges and grows throughout the visible and Near-Infrared (NIR) spectrum. The most substantial difference between healthy and cracked tissue is visibly concentrated in the **700 nm to 900 nm range**. This wide gap indicates that the pixels' reflectance values in this specific region are highly sensitive to the presence of cracks (or rott), confirming that sufficient spectral information exists for classification.



avg_std_signatures.png

Figure 4.1: Mean spectral signatures of pixels sampled from 'Cracked' (Orange) versus 'Regular' (Blue) grape tissue. The shaded areas represent the standard deviation. The largest separation is evident in the NIR region (700-900 nm).

4.1.2 Consensus-Based Feature Selection

While the visual analysis confirms the potential for classification, using the entire spectrum as input for a model is inefficient and can lead to overfitting due to the high dimensionality of the data. Therefore, a formal feature selection process was employed to identify the most potent subset of wavelengths. To ensure the robustness of the selected features, three distinct methods were used: **Fisher Score**, **Mutual Information**, and **Random Forest Feature Importance**.

The analysis revealed a remarkable consensus among the methods, confirming the initial observation. All three algorithms consistently highlighted the Near-Infrared (NIR)

region, specifically the **700 nm to 900 nm range**, as the most informative for detecting cracks. Out of the top 30 wavelengths ranked by each method, **19 key wavelengths** were unanimously identified as significant by all three algorithms, and a further 11 were selected by at least two methods. This high level of agreement provides strong validation for the selected features.

The 19 high-consensus wavelengths, listed in Table 4.1, form the final feature set for model development. Figure 4.2 visually reinforces this result, plotting the mean signatures and highlighting the locations of these top-ranked features, which align perfectly with the region of maximum spectral separation.

Table 4.1: The 19 most informative wavelengths, selected concurrently by all three feature selection methods (Fisher Score, Mutual Information, and Random Forest), forming the high-consensus feature set for model development.

High-Consensus Wavelengths (nm)				
720.5	723.5	726.5	729.5	732.5
735.5	738.5	741.5	744.5	747.5
750.5	753.5	756.6	759.6	762.6
768.6	780.6	786.7	801.8	

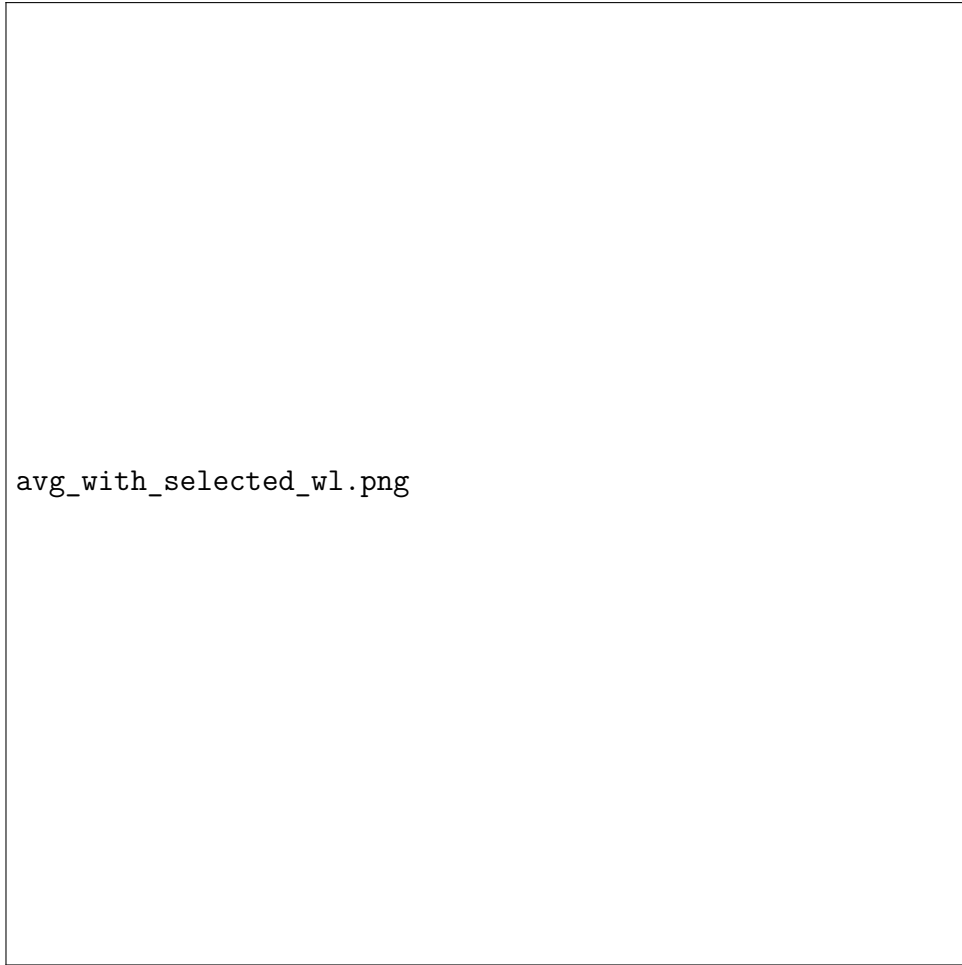


Figure 4.2: Mean spectral signatures with the top-ranked, high-consensus wavelengths highlighted by vertical dashed lines. The selected features are clearly concentrated in the NIR region, coinciding with the area of greatest spectral separation between cracked and regular tissue.

4.2 Pixel-Level Classification Model Benchmark

To identify the most effective method for distinguishing between cracked and healthy grape tissue at the pixel level, a comprehensive benchmark of nine different classification models was conducted. The evaluation was performed on two versions of the dataset: the original, naturally imbalanced dataset, and a balanced version created through under-sampling of the majority class. A robust Leave-One-Group-Out (LOGO) cross-validation strategy was employed to ensure the models’ ability to generalize to unseen grape clusters.

4.2.1 Performance on the Unbalanced Dataset

The models were first evaluated on the full, unbalanced dataset. Simpler linear models demonstrated top-tier performance. Logistic Regression achieved the highest mean accuracy (97.13%), while the Support Vector Classifier (SVC) yielded the highest Mean ROC AUC (0.9943). The Linear Discriminant Analysis (LDA) model provided the highest

precision for the 'Cracked' class (90.55%). Full results are in Table 4.2.

Table 4.2: Model Performance on the Unbalanced Dataset.

Model Name	Mean Acc.	Mean ROC AUC	Mean PR AUC	Mean F1	Prec. (Cracked)	Recall (Cracked)	Train Time (s)
XGBoost	0.9525	0.9845	0.8983	0.9634	0.8845	0.8454	0.84
RandomForest	0.9491	0.9849	0.8984	0.9620	0.8868	0.8425	47.16
GradientBoosting	0.9362	0.9850	0.8973	0.9525	0.8863	0.8345	660.83
MLP (CPU)	0.9394	0.9843	0.8953	0.9529	0.8578	0.8464	44.60
LDA	0.9624	0.9931	0.9025	0.9733	0.9055	0.8510	1.15
QDA	0.9568	0.9763	0.8915	0.9661	0.8988	0.8560	1.51
LogisticRegression	0.9713	0.9931	0.9023	0.9783	0.8991	0.8719	3.04
DecisionTree	0.9200	0.9387	0.8374	0.9383	0.8514	0.8245	13.03
SVC	0.9663	0.9943	0.9026	0.9712	0.8775	0.8616	403.41

4.2.2 Performance on the Balanced Dataset

To assess model behavior without class imbalance, the benchmark was repeated on a balanced dataset. After balancing, the linear models continued to excel. LDA emerged with the highest mean accuracy (96.72%). This indicates that the underlying spectral signatures are robust. The results are presented in Table 4.3.

Table 4.3: Model Performance on the Balanced Dataset.

Model Name	Mean Acc.	Mean ROC AUC	Mean PR AUC	Mean F1	Prec. (Cracked)	Recall (Cracked)	Train Time (s)
XGBoost	0.9389	0.9859	0.9023	0.9522	0.8909	0.8493	0.70
RandomForest	0.9383	0.9844	0.9018	0.9530	0.8939	0.8482	29.43
GradientBoosting	0.9347	0.9851	0.9019	0.9490	0.8925	0.8524	356.38
MLP	0.9401	0.9810	0.8983	0.9508	0.8842	0.8551	26.78
LDA	0.9672	0.9919	0.9048	0.9743	0.9030	0.8669	0.58
QDA	0.9459	0.9758	0.8964	0.9558	0.8985	0.8584	0.75
LogisticRegression	0.9654	0.9929	0.9051	0.9728	0.9003	0.8753	1.73
DecisionTree	0.9052	0.9396	0.8591	0.9254	0.8735	0.8224	6.43
SVC	0.9600	0.9936	0.9048	0.9654	0.8881	0.8647	175.60

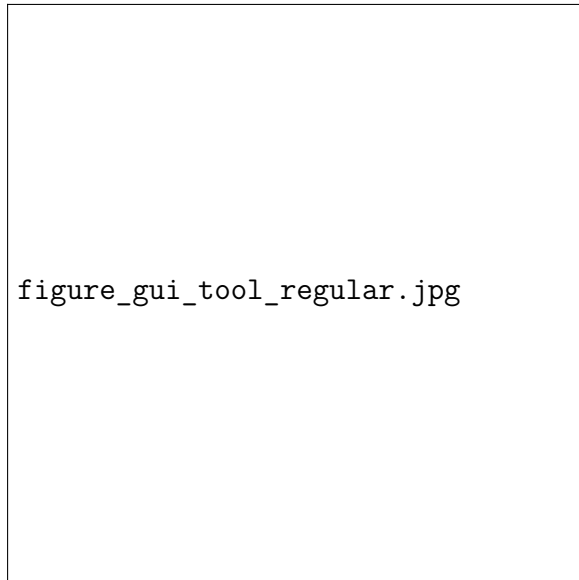
4.2.3 Anomaly Detection Benchmark

In addition to benchmarking conventional supervised classifiers, we also evaluated a one-class anomaly detection framework tailored for hyperspectral grape crack detection. Specifically, we implemented an autoencoder trained exclusively on pixels from one class (either *Cracked* or *Regular*), and evaluated its ability to detect anomalies in unseen clusters. Training was performed with a strict Leave-One-Group-Out (LOGO) cross-validation strategy, where entire grape clusters were left out in each fold, thereby preventing information leakage across clusters.

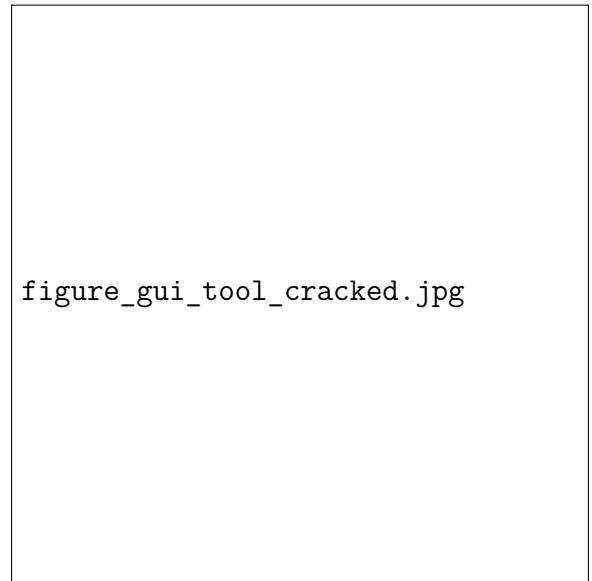
This design ensured that the models did not simply memorize spectral characteristics of individual berries, but instead demonstrated generalization to entirely new grape clusters. Early stopping and adaptive learning rate scheduling were applied to stabilize training and avoid overfitting. The anomaly detection results (ROC curves, anomaly score distributions, and per-sample predictions) provided complementary validation of

the discriminative power of spectral signatures, and confirmed that robust separation between cracked and healthy tissue can be achieved even under one-class training scenarios.

The custom-built UI tool was also instrumental in the exploratory analysis phase of this benchmark. It was enhanced to provide real-time inference using the trained autoencoder. For any given image, the tool could automatically generate segments using SAM2 and then calculate the mean squared reconstruction error (MSE) for the spectral signature of each segment. This allowed for an interactive workflow where segments could be sorted by their anomaly score (MSE), enabling rapid visual validation of the model’s ability to distinguish between healthy tissue (low MSE) and potential cracks or other anomalies (high MSE).



(a) A pixel from a healthy berry. The AutoEncoder, trained exclusively on healthy tissue, reconstructs the signature with high fidelity, resulting in a very low MSE of 0.0001.



(b) A pixel from a cracked berry. Because the model has never seen cracked tissue, it struggles to reconstruct the anomalous signature, resulting in a significantly higher MSE of 0.0027 and a visible divergence between the original and reconstructed signals.

Figure 4.3: Visual validation of the one-class AutoEncoder for anomaly detection. The tool demonstrates the model’s ability to differentiate between (a) normal and (b) anomalous spectral signatures in real-time based on the reconstruction error (MSE).

4.2.4 Wavelength Sensitivity Analysis with LOGO Cross-Validation

To evaluate how the number of selected wavelengths (k) affects model performance, a dedicated wavelength sensitivity analysis was conducted. The primary aim of this experiment was to determine the smallest subset of spectral features sufficient for robust classification of *Cracked* versus *Regular* grape tissue, while ensuring that the evaluation strictly generalizes across unseen grape clusters.

A rigorous *Leave-One-Group-Out (LOGO)* cross-validation strategy was employed, in which entire grape clusters were held out as test sets in each fold. To prevent biased evaluation, folds in which either the training or test partition contained only a single class were discarded. This ensured that the reported results reflect true generalization rather than memorization of cluster-specific properties.

Two linear classifiers, **Logistic Regression** and **Linear Discriminant Analysis (LDA)**, were benchmarked. For feature selection, we employed **Sequential Forward Selection (SFS)**, where wavelengths were added one by one to maximize the ROC AUC score within a nested cross-validation on the training folds. This approach allows identification of an informative yet compact subset of wavelengths.

The analysis was repeated for $k = 1 \dots 20$ selected wavelengths. For each configuration, the following metrics were computed and averaged across all valid folds: *Accuracy*, *ROC AUC*, *PR AUC* (average precision), *Precision*, *Recall*, *F1-score*, and training time. In addition to reporting mean values, fold-level variability was examined through box-plots, highlighting the stability of the models as k increased. Representative results are shown in Table 4.4, while Figure 4.4 illustrates the performance curves.

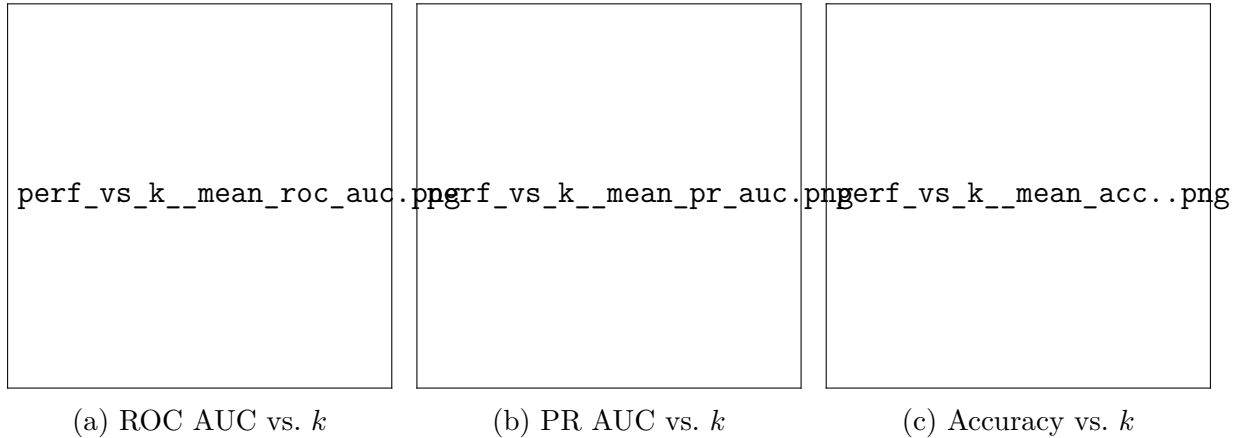


Figure 4.4: Performance vs. k for LDA and Logistic Regression (SFS, LOGO).

Table 4.4: Mean performance across LOGO folds using SFS for selected k .

Model	k	Mean Acc.	Mean ROC AUC	Mean PR AUC	Mean F1	Recall (Cracked)
LDA	5	0.977	0.991	0.985	0.955	0.943
LDA	10	0.978	0.991	0.987	0.959	0.941
LDA	15	0.977	0.991	0.987	0.958	0.940
LDA	20	0.977	0.990	0.986	0.958	0.940
Logistic Regression	5	0.968	0.990	0.983	0.943	0.942
Logistic Regression	10	0.973	0.992	0.988	0.952	0.954
Logistic Regression	15	0.972	0.993	0.988	0.950	0.956
Logistic Regression	20	0.974	0.994	0.990	0.952	0.955

Summary and trends. Both models reach near-saturated ROC/PR AUC around $k \approx 10$ –12. Logistic Regression exhibits a slight edge in ROC/PR AUC and recall as k

increases, whereas LDA remains highly competitive already with fewer wavelengths. In other words, most of the separability between *Cracked* and *Regular* pixels can be captured by a compact subset of features.

Figure 4.5 focuses on ROC AUC versus k ; the curve rises sharply for small k and then plateaus, indicating diminishing returns beyond roughly ten wavelengths. To expose the effect of each additional feature, Figure 4.6 reports the *marginal gain* (Δ ROC AUC from $k-1$ to k), and Figure 4.7 shows the *cumulative* Δ relative to $k=1$. Together, these plots confirm that the improvement is front-loaded: the first few wavelengths deliver most of the gain.

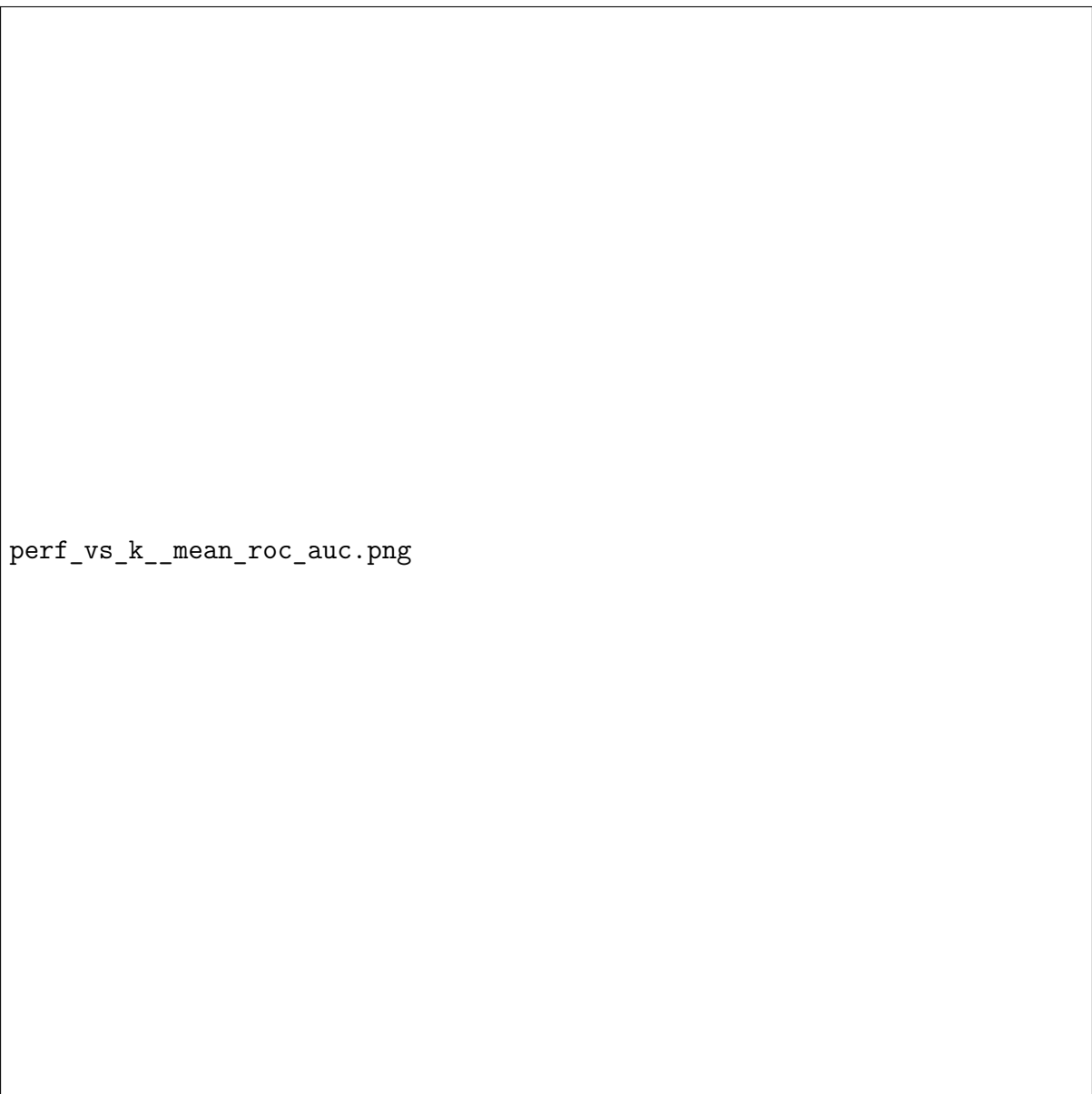



Figure 4.5: Performance (ROC AUC) vs. k for LDA and Logistic Regression (SFS, LOGO). The curve saturates near $k \approx 10$ –12.

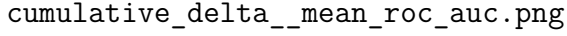


marginal_gain__mean_roc_auc.png

Figure 4.6: Marginal gain (Δ ROC AUC) vs. k . Most of the improvement arrives with the first few wavelengths.

Fold-averaged curves at a representative operating point ($k = 10$). To visualize the resulting discrimination in probability space, we plot mean ROC and Precision–Recall (PR) curves at $k=10$ (averaged over LOGO folds). The thin lines represent individual folds and the bold line their mean.

Selection stability and spectral locality. Finally, we examined the stability of SFS across folds and k values. The Top-40 frequency heatmap in Figure 4.9 reveals consistent emphasis on the NIR region (approximately 700–900 nm), aligning with the spectral-signature analysis in Section 4.1 and supporting the hypothesis that crack-related contrasts are most pronounced in NIR.



cumulative_delta__mean_roc_auc.png

Figure 4.7: Cumulative Δ ROC AUC relative to $k=1$. Gains plateau around $k \approx 10$ –12.

Practical implication. Reliable pixel-level classification is attainable with as few as 8–12 carefully selected wavelengths, predominantly in the NIR. This suggests that compact, application-specific multispectral sensors are feasible for early crack detection in grapes, with minimal loss relative to using the full spectrum.

4.2.5 One-Class Training with Autoencoder (LOGO)

In many practical settings only *normal* (Regular) examples are readily available, while the goal is to detect *anomalies* (Cracked) without fully supervised training. We therefore used a **one-class Autoencoder (AE)**: the model is trained *only* on pixels from the Regular class to reconstruct the spectral vector. At inference, the anomaly score is the reconstruction error (MSE); higher values indicate higher likelihood of anomaly.

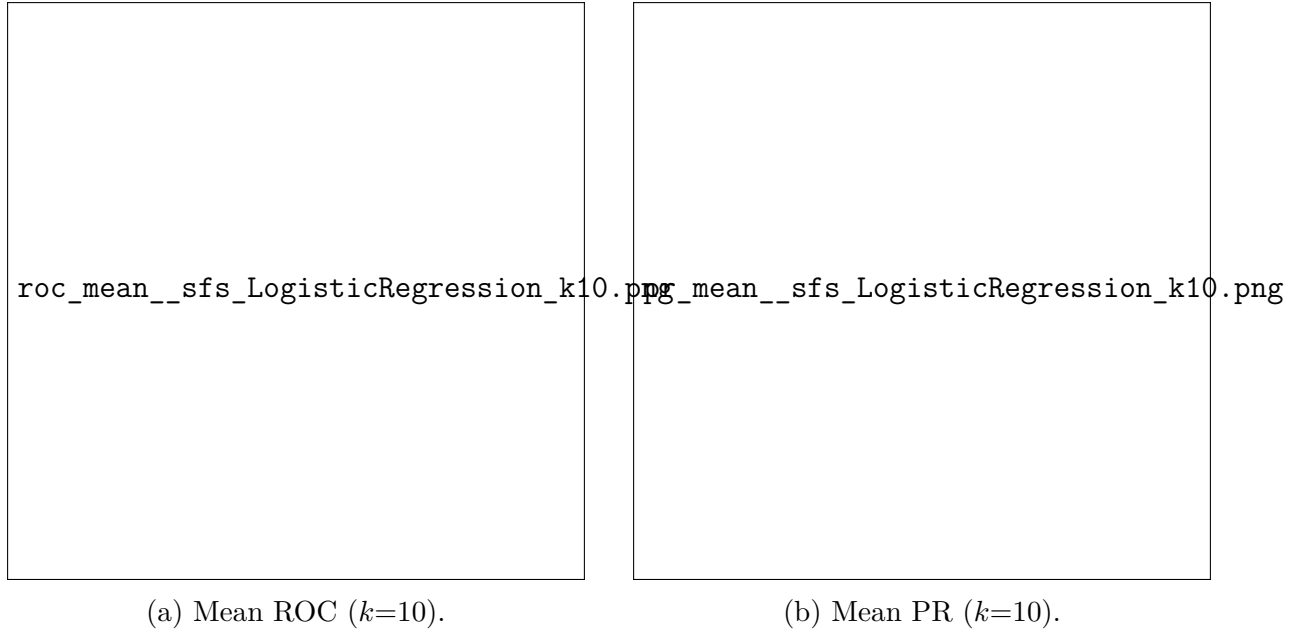


Figure 4.8: Fold-level variability (faint) and mean curves (bold) for Logistic Regression at $k=10$. Both ROC AUC and AP are very high, indicating excellent separability under cluster-level generalization.

Validation protocol. We employed **Leave-One-Group-Out (LOGO)** cross-validation over grape clusters. In each fold, one cluster was held out for testing and the AE was trained *only* on Regular pixels from the training clusters. A **StandardScaler** was fitted on the Regular training data of the fold and applied to both train and test. Folds with single-class test sets or with no Regular samples for training were discarded. For every fold we computed ROC-AUC and AP; then we concatenated all out-of-fold (OOF) predictions to select thresholds and to report *micro*-averaged curves.

Threshold selection. We report two operating points:

1. **Target-FPR (5%):** choose the threshold that yields a 5% false positive rate on the OOF predictions (conservative, fewer false alarms).
2. **F1-max:** sweep all thresholds returned by `precision_recall_curve` and pick the one that maximizes the F1-score on the OOF predictions (balances precision/recall).

Results. Aggregating across folds we obtained:

Mean ROC-AUC = 0.913 ± 0.128 , Micro ROC-AUC = 0.916, Mean AP = 0.845 ± 0.234 , Micro AP

The two operating points are summarized in Table 4.5. The micro-averaged ROC curve is shown in Figure 4.11, and the confusion matrices for both thresholds are shown in Figure 4.12.


feature_frequency_heatmap_topN.png

Figure 4.9: Feature selection stability across k (Top-40 wavelengths by fold-frequency; SFS+LOGO). Dominant bands reside in the NIR (700–900 nm).

Table 4.5: Autoencoder performance on concatenated OOF predictions (LOGO) at two operating points.

Operating point	Threshold	Accuracy	Precision	Recall	F1
Target-FPR = 5%	0.0199	0.852	0.853	0.635	0.728
F1-max	0.0119	0.834	0.692	0.850	0.762

Discussion. The AE provides good separability between Regular and Cracked pixels (ROC-AUC ≈ 0.91) despite training on a single class only. The operating point controls the behavior: *Target-FPR=5%* yields higher precision (0.853) at the cost of recall (0.635), suitable for conservative screening; *F1-max* increases recall to 0.85 and improves F1 to



feature_frequency_heatmap_full.pdf

Figure 4.10: Full feature-frequency heatmap across all wavelengths.

0.762 at the expense of precision. The across-fold standard deviations indicate non-negligible cluster variability—expected for a one-class approach under changing spectral profiles. While supervised models (Section 4.2.4) perform better, these results support one-class AE as a practical *pre-screening* tool when anomaly labels are scarce or expensive.

4.3 Challenges in Whole-Image Classification

Despite high pixel-level accuracy, applying the trained XGBoost model to classify every pixel in a full image resulted in noisy outputs. The model frequently misclassified non-grape materials, such as plastic ties and dry twigs, as 'Cracked', as they share similar spectral properties with cracked grape tissue.

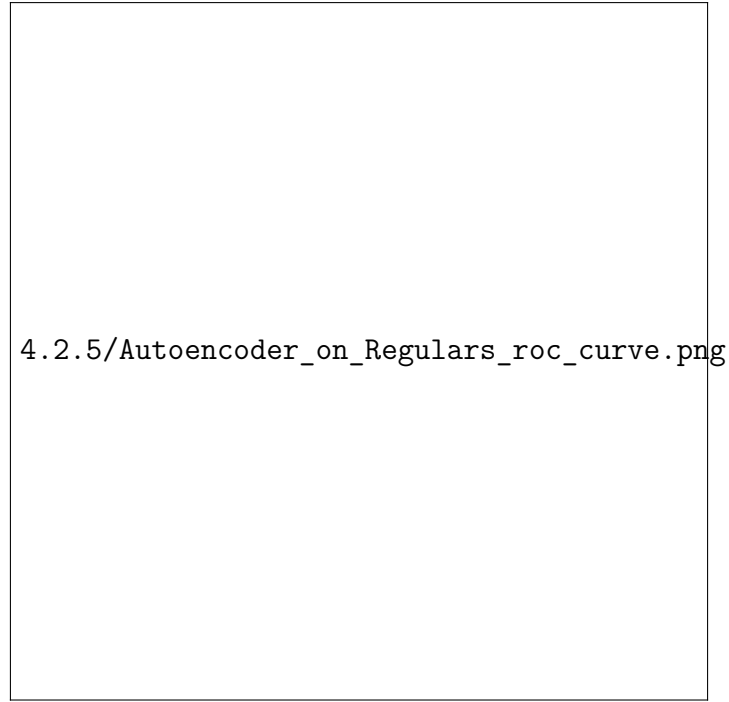


Figure 4.11: Micro-averaged ROC curve for the one-class Autoencoder (LOGO).

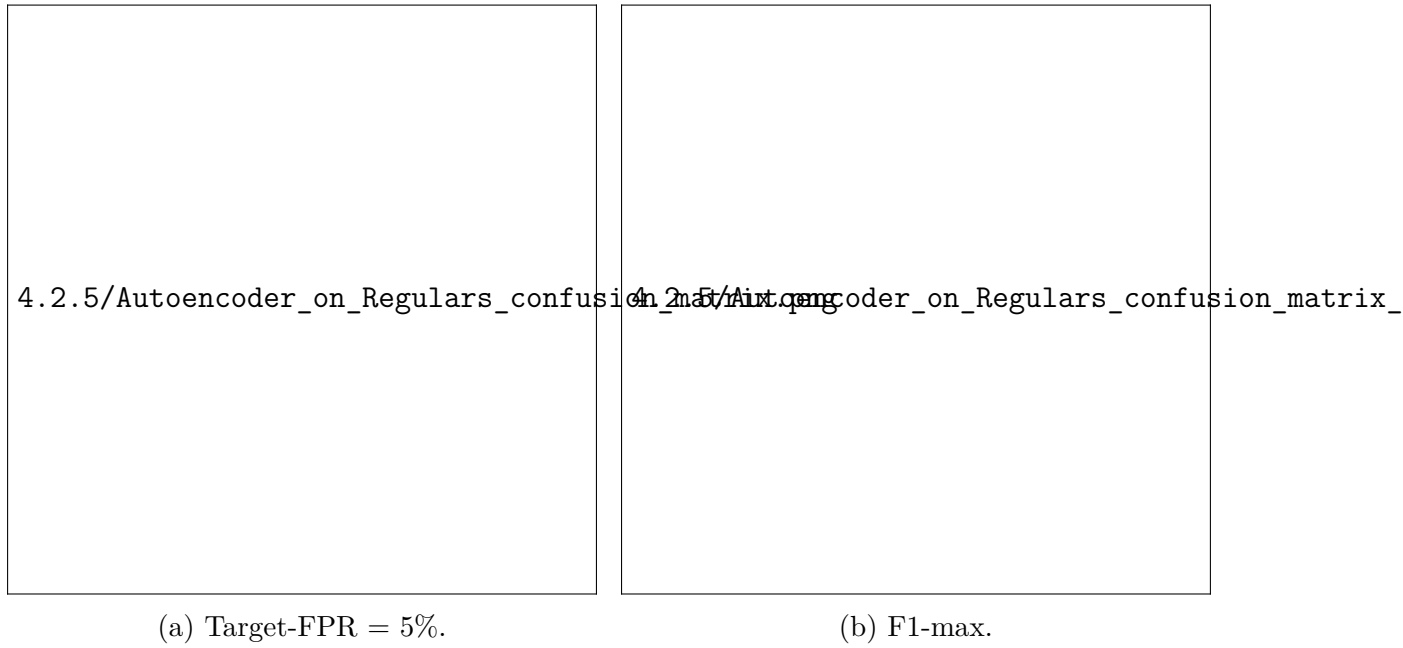


Figure 4.12: Confusion matrices on concatenated OOF predictions at the two operating points.

4.4 Temporal Trends in Mean Spectral Signatures

Analysis of the mean spectral signature of entire clusters over time yielded much more stable results. A clear divergence was observed between the different irrigation treatments as the season progressed. The most significant separation between the groups occurred in the NIR region (approx. 700-900 nm).

Chapter 5

Discussion

The results of this study present a dual narrative. On one hand, the spectral signal of cracked grape tissue is fundamentally distinct from that of healthy tissue, enabling remarkably high classification accuracy at a granular, pixel level. An accuracy of 97% with an AUC of 0.99 confirms the first hypothesis and establishes HSI as a technically viable tool for crack identification.

However, the transition from a controlled, pixel-based analysis to a practical, whole-image application exposed a critical limitation. The "noisy" results from the whole-image classification underscore the challenge of semantic segmentation in complex agricultural scenes. The model's confusion of cracks with background elements highlights that a simple pixel-wise classifier is insufficient for field deployment. A more robust system would require a preliminary segmentation step.

This challenge led to the most promising finding of the research: the value of analyzing mean spectral signatures over time. This approach effectively averages out pixel-level noise and reveals underlying physiological trends. The clear separation of irrigation treatments, especially in the 700-900 nm range, is highly significant. This spectral region is sensitive to changes in water content, cell structure, and chlorophyll concentration. The observed divergence strongly suggests that HSI is not just detecting damage but is monitoring the level of water stress in the grapes—a known precursor to cracking. This shifts the paradigm from simple detection to proactive prediction. This finding directly supports the second and third hypotheses.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This research successfully demonstrated that:

1. Hyperspectral imaging in the 484-933 nm range can differentiate between cracked and healthy grape tissue with extremely high accuracy (97%) at the pixel level.
2. Direct application of pixel-level classifiers to entire field images is unreliable due to spectral confusion with background materials.
3. Monitoring the temporal evolution of a grape cluster's mean spectral signature, particularly in the 700-900 nm NIR region, is a robust and promising method for assessing physiological stress related to irrigation and predicting cracking susceptibility.

6.2 Future Work

Based on these conclusions, several avenues for future research are recommended:

- **Advanced Segmentation:** Develop a deep learning-based segmentation model (e.g., a U-Net or Mask R-CNN) to isolate grape clusters from the background before classification.
- **Spatio-Spectral Models:** Explore 3D-Convolutional Neural Networks (3D-CNNs) that can simultaneously learn from both spatial and spectral information.
- **Physiological Validation:** Conduct further studies to establish a direct quantitative link between the observed spectral trends and key physiological indicators of stress.

- **In-Field System Development:** Work towards integrating these models into a real-time, in-field decision support system to help growers optimize irrigation and harvesting strategies.