



UNIVERSITY OF CAPE TOWN

MSC ADVANCED ANALYTICS

MASTER'S DISSERTATION

---

# Autonomous Anomaly Detection of Individual Tree Crown Delineations

---

*Author:*  
Edward Baleni

*Student Number:*  
BLNEDW003

Supervisor(s)  
Patrick Marais  
Sebnem Er

July 23, 2026

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

## Abstract

Individual tree crown detection and delineation (ITCD) algorithms extract the boundaries of individual tree crowns from images. State-of-the-art machine learning techniques segment crowns using drones and multispectral sensors. Farmers use this emerging technology to help improve crop yields, reduce resource inputs, and improve management approaches, ultimately enhancing agricultural sustainability. However, ITCD methods often fail to delineate all tree boundaries accurately. Poorly estimated delineations significantly limit the efficiency of monitoring and management strategies within orchards. This work proposes a new processing pipeline to detect three commonly occurring anomalous delineations: i) under-segmentation, ii) over-segmentation, and iii) false positives. A hand-crafted feature set comprising coarse shape descriptors, Haralick features, and spectral indices are extracted from orchard imagery to accomplish this. Furthermore, a novel approach was developed based on local Geary’s  $c$  statistic in a multivariate context (GBOD), leveraging physical neighbourhoods to identify contextual outlierness. This method was evaluated — using average precision and AUC-ROC performance measures — against unsupervised anomaly detection algorithms, including isolation forest (IForest), angle-based outlier detection (ABOD), empirical-cumulative-distribution-based outlier detection (ECOD), local outlier factor (LOF) and PCA-based anomaly detection (PCA). Finally, a meta-learner is trained to automate anomaly detection in new orchards. Results show that local outlier detection methods effectively capture over-segmentation. In contrast, global outlier detection methods better capture under-segmentation and false positives, and fare well in general cases where multiple anomaly types are present. Overall, ABOD, IForest and PCA are the top models for detecting poor delineations (average AUC scores: 0.97, 0.964, and 0.962, respectively). GBOD gave less consistent results (average AUC score: 0.943). Nevertheless, it identified problematic regions while attributing outlierness, a capability none of the top three possess. This work successfully demonstrates automatic model selection and anomaly detection for out-of-sample delineations, a notoriously problematic task due to lacking of reference data.

# Contents

|                                                               |            |
|---------------------------------------------------------------|------------|
| <b>List of Figures</b>                                        | <b>iii</b> |
| <b>List of Tables</b>                                         | <b>vii</b> |
| <b>1 Introduction</b>                                         | <b>1</b>   |
| 1.1 Aims and Objectives . . . . .                             | 2          |
| 1.2 Thesis Structure . . . . .                                | 2          |
| <b>2 Background</b>                                           | <b>4</b>   |
| 2.1 Precision Agriculture and Remote Sensing . . . . .        | 4          |
| 2.2 Individual Tree Crown Detection and Delineation . . . . . | 6          |
| 2.2.1 Difficulties in ITCD . . . . .                          | 7          |
| 2.3 Related Work on Anomalous Segmentations . . . . .         | 8          |
| 2.3.1 Anomaly Detection of ITCD Algorithms . . . . .          | 9          |
| 2.4 Feature Engineering . . . . .                             | 9          |
| 2.4.1 Digital Surface Model . . . . .                         | 9          |
| 2.4.2 Spectral Features . . . . .                             | 10         |
| 2.4.3 Shape Descriptors . . . . .                             | 11         |
| 2.4.4 Texture Features . . . . .                              | 13         |
| 2.5 Overview . . . . .                                        | 14         |
| <b>3 Data and Feature Engineering</b>                         | <b>15</b>  |
| 3.1 Study Area and Materials . . . . .                        | 15         |
| 3.2 Feature Engineering . . . . .                             | 19         |
| 3.3 Summary . . . . .                                         | 22         |
| <b>4 Anomaly Detection</b>                                    | <b>23</b>  |
| 4.1 Background . . . . .                                      | 23         |
| 4.1.1 Types of Anomalies . . . . .                            | 23         |
| 4.1.2 Learning Paradigms . . . . .                            | 24         |
| 4.1.3 Anomaly Scoring . . . . .                               | 24         |
| 4.2 Fundamental and State-of-the-Art Algorithms . . . . .     | 25         |
| 4.2.1 Diagram-Based Methods . . . . .                         | 26         |

|          |                                                            |           |
|----------|------------------------------------------------------------|-----------|
| 4.2.2    | Model/Probabilistic-Based Methods . . . . .                | 26        |
| 4.2.3    | Proximity-Based Methods . . . . .                          | 27        |
| 4.2.4    | Cluster-Based Methods . . . . .                            | 28        |
| 4.2.5    | Learning-Based Methods . . . . .                           | 29        |
| 4.2.6    | Spectral/Reconstruction-Based Methods . . . . .            | 30        |
| 4.2.7    | Graph-Based Methods . . . . .                              | 31        |
| 4.2.8    | Spatial Anomaly Detection . . . . .                        | 31        |
| 4.3      | Summary . . . . .                                          | 31        |
| <b>5</b> | <b>Proposed Algorithm</b>                                  | <b>33</b> |
| 5.1      | Formulation of Multivariate LISA . . . . .                 | 33        |
| 5.2      | Algorithm of Geary-Based Outlier Detection (GBOD). . . . . | 35        |
| 5.3      | Interpretability . . . . .                                 | 36        |
| 5.4      | Constraints . . . . .                                      | 36        |
| 5.5      | Summary . . . . .                                          | 37        |
| <b>6</b> | <b>Experimental Study</b>                                  | <b>38</b> |
| 6.1      | Dataset, Train/Test Split . . . . .                        | 39        |
| 6.2      | Exploratory data analysis . . . . .                        | 41        |
| 6.2.1    | Ablation Study . . . . .                                   | 42        |
| 6.3      | Baseline Algorithms . . . . .                              | 42        |
| 6.4      | Evaluation Metrics . . . . .                               | 43        |
| 6.5      | Hyperparameter Tuning . . . . .                            | 44        |
| 6.6      | Model Comparison . . . . .                                 | 45        |
| 6.7      | Generalisation Tests, Out-of-Sample . . . . .              | 46        |
| 6.8      | Environment Setup . . . . .                                | 46        |
| 6.9      | Summary . . . . .                                          | 46        |
| <b>7</b> | <b>Exploratory Results and Feature Reduction</b>           | <b>47</b> |
| 7.1      | Exploratory Data Analysis . . . . .                        | 47        |
| 7.1.1    | Profiles . . . . .                                         | 48        |
| 7.1.2    | Univariate Analysis . . . . .                              | 50        |
| 7.1.3    | Bivariate Analysis . . . . .                               | 54        |
| 7.2      | Initial Feature Reduction . . . . .                        | 58        |
| 7.3      | Ablation Study . . . . .                                   | 58        |
| 7.4      | Summary . . . . .                                          | 60        |
| <b>8</b> | <b>Experimental Results</b>                                | <b>61</b> |
| 8.1      | Model Evaluation . . . . .                                 | 61        |
| 8.1.1    | Model Comparison . . . . .                                 | 64        |
| 8.1.2    | Model Selection . . . . .                                  | 65        |
| 8.1.3    | Runtime Efficiency . . . . .                               | 66        |
| 8.2      | Generalisation . . . . .                                   | 67        |
| 8.2.1    | Out-Of-Sample Results . . . . .                            | 67        |

|          |                                        |            |
|----------|----------------------------------------|------------|
| 8.3      | A Closer Look at GBOD . . . . .        | 69         |
| 8.3.1    | Attribution . . . . .                  | 69         |
| 8.4      | Discussion . . . . .                   | 71         |
| <b>9</b> | <b>Conclusion</b>                      | <b>73</b>  |
| 9.1      | Key Findings . . . . .                 | 73         |
| 9.2      | Future Work . . . . .                  | 74         |
| 9.3      | Concluding Remarks . . . . .           | 75         |
| <b>A</b> | <b>Exploratory Data Analysis</b>       | <b>87</b>  |
| A.1      | Visual Analysis . . . . .              | 87         |
| A.2      | DSM Profile Plots . . . . .            | 89         |
| A.3      | Spectral Profile Plots . . . . .       | 91         |
| A.4      | Univariate Analysis . . . . .          | 102        |
| A.5      | Texture Boxplot Outliers . . . . .     | 103        |
| A.6      | Bivariate Analysis . . . . .           | 107        |
| <b>B</b> | <b>Additional Experimental Results</b> | <b>110</b> |
| B.1      | Model Evaluation . . . . .             | 110        |
| B.2      | Out-of-sample . . . . .                | 116        |
| B.3      | A Look at GBOD . . . . .               | 117        |

# List of Figures

|     |                                                                                                                                                                                                                           |    |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Illustrating anomaly classifications in the context of this study: A) false positive. B) Over-segmentation. C) Under-segmentation. . . . .                                                                                | 2  |
| 2.1 | Depicting the common results observed in ITCD algorithms (Dong et al., 2020). . . . .                                                                                                                                     | 7  |
| 2.2 | Demonstration of how Zernike moments capture information about an image Kulkarni et al. (2014). . . . .                                                                                                                   | 13 |
| 3.1 | Examples of different orchard types . . . . .                                                                                                                                                                             | 16 |
| 3.2 | Illustrating the organic spectroscopy over an orchard profile. . . . .                                                                                                                                                    | 16 |
| 3.3 | Estimated delineations of orchard tree crowns. . . . .                                                                                                                                                                    | 18 |
| 3.4 | Incorrect delineations of orchard tree crowns. Where reference polygons are illustrated as green-coloured polygon, while estimated delineations are represented by blue polygons and erroneous estimates are red. . . . . | 18 |
| 4.1 | Fundamental and state-of-the-art anomaly detection methods. . . . .                                                                                                                                                       | 25 |
| 4.2 | A representation of how isolation forests work by Hariri et al. (2021). Red represents an anomalous observation while blue represents a normal observation . . . . .                                                      | 29 |
| 6.1 | Architecture Diagram . . . . .                                                                                                                                                                                            | 38 |
| 6.2 | Comparison of baseline algorithms and the proposed methodology . . . . .                                                                                                                                                  | 43 |
| 7.1 | Subset of Figure A.1, depicting the Orchard and reference data of Orchard four of the training data. . . . .                                                                                                              | 47 |
| 7.2 | Subset of Figure A.2, depicting the Orchard DSM of Orchard four of the training data . . . . .                                                                                                                            | 48 |
| 7.3 | Subset of the profile plots presented in Section A.3, depicting the aerial reflectance and vegetative indices of Orchard four of the training data . . . . .                                                              | 49 |
| 7.4 | Boxplots illustrating the distribution of the shape descriptors and confidence of the first five orchards . . . . .                                                                                                       | 51 |

|      |                                                                                                                                                                                                                                       |    |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 7.5  | Boxplots illustrating the distribution of the spectral features of the first five datasets . . . . .                                                                                                                                  | 51 |
| 7.6  | Boxplots illustrating the distribution of texture descriptors of the first five datasets . . . . .                                                                                                                                    | 52 |
| 7.7  | Boxplots illustrating the distribution of the first 25 Zernike moments of the first five datasets . . . . .                                                                                                                           | 53 |
| 7.8  | Plot of pair-wise relationships between shape descriptors of the first five randomly selected training datasets. . . . .                                                                                                              | 54 |
| 7.9  | Plot of pair-wise relationships between spectral features and DSM of the first five randomly selected training datasets . . . . .                                                                                                     | 55 |
| 7.10 | Plot of pair-wise relationships between texture descriptors of the first five randomly selected training datasets. . . . .                                                                                                            | 56 |
| 7.11 | Subset of the clustered heatmaps presented in Appendix A, depicting the correlations within feature groupings of Orchard 4 of the training data. . . . .                                                                              | 57 |
| 7.12 | Summary of ablation study . . . . .                                                                                                                                                                                                   | 59 |
| 8.1  | Receiver operating characteristic curve of the detection algorithms over the first 20 orchards . . . . .                                                                                                                              | 62 |
| 8.2  | Precision-Recall curve of the detection algorithms over the first 20 orchards . . . . .                                                                                                                                               | 63 |
| 8.3  | Critical difference diagram of AUC demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithms . . . . .                                                                     | 64 |
| 8.4  | Critical difference diagram demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithm on a subset of the data dominated by over-segmentations . . . . .                     | 65 |
| 8.5  | Critical difference diagram demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithm on a subset of the data dominated by under-segmentations and false-postives . . . . . | 65 |
| 8.6  | Depicting change in time vs increase in number of observations . . . . .                                                                                                                                                              | 66 |
| 8.7  | Summary of time complexity . . . . .                                                                                                                                                                                                  | 66 |
| 8.8  | Boxplot summarising the performance over the test data sets, on algorithms trained via HPOD's meta-learning and on default hyperparameters . . . . .                                                                                  | 67 |
| 8.9  | Visual demonstration of what the classifiers capture in localized areas of orchards 74, 78, and 89 . . . . .                                                                                                                          | 68 |
| 8.10 | The outlier scores of IForest and the GBOD to illustrate how they are appointing outliers . . . . .                                                                                                                                   | 69 |
| 8.11 | Anomalies and their corresponding attributions relative to the 99 <sup>th</sup> percentile of the outlier score for each dimension . . . . .                                                                                          | 70 |

|      |                                                                                                            |     |
|------|------------------------------------------------------------------------------------------------------------|-----|
| A.1  | Depicting the orchard and reference data of the the first 5 randomly selected training datasets. . . . .   | 88  |
| A.2  | Depicting the 2-D and 3-D DSM data of the the first 5 randomly selected training datasets. . . . .         | 90  |
| A.3  | Profile plot of red aerial reflectance band of the first 5 randomly selected training datasets. . . . .    | 91  |
| A.4  | Profile plot of green aerial reflectance band of the first 5 randomly selected training datasets. . . . .  | 92  |
| A.5  | Profile plot of blue aerial reflectance band of the first 5 randomly selected training datasets. . . . .   | 93  |
| A.6  | Profile plot of REG aerial reflectance band of the first 5 randomly selected training datasets. . . . .    | 94  |
| A.7  | Profile plot of NIR aerial reflectance band of the first 5 randomly selected training datasets. . . . .    | 95  |
| A.8  | Profile plot of EVI vegetative index of the first 5 randomly selected training datasets. . . . .           | 96  |
| A.9  | Profile plot of GNDVI vegetative index of the first 5 randomly selected training datasets. . . . .         | 97  |
| A.10 | Profile plot of NDRE vegetative index of the first 5 randomly selected training datasets. . . . .          | 98  |
| A.11 | Profile plot of NDVI vegetative index of the first 5 randomly selected training datasets. . . . .          | 99  |
| A.12 | Profile plot of OSAVI vegetative index of the first 5 randomly selected training datasets. . . . .         | 100 |
| A.13 | Profile plot of SAVI vegetative index of the first 5 randomly selected training datasets. . . . .          | 101 |
| A.14 | Bending energy. . . . .                                                                                    | 102 |
| A.15 | Illustration of ASM outliers of the first 5 randomly selected training datasets. . . . .                   | 103 |
| A.16 | Illustration of contrast's left tail outliers of the first 4 randomly selected training datasets. . . . .  | 104 |
| A.17 | Illustration of contrast's right tail outliers of the first 5 randomly selected training datasets. . . . . | 105 |
| A.18 | Illustration of correlation outliers of the first 5 randomly selected training datasets. . . . .           | 106 |
| A.19 | Clustermap of texture features of the first 5 randomly selected training datasets. . . . .                 | 107 |
| A.20 | Clustermap of spectral features of the first 5 randomly selected training datasets. . . . .                | 108 |
| A.21 | Clustermap of shape descriptors of the first 5 randomly selected training datasets. . . . .                | 109 |
| B.1  | Summary of performance captured over 70 different orchards. . . . .                                        | 110 |
| B.2  | Reference data of orchards observed OOS. . . . .                                                           | 116 |

B.3 Orchard 62, depicting anomalous delineations captured by the Geary-based method . . . . . 117

# List of Tables

|     |                                                                                                                                                                                                                             |     |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Descriptions of some coarse shape descriptors. . . . .                                                                                                                                                                      | 12  |
| 3.1 | Feature set used to perform anomaly detection . . . . .                                                                                                                                                                     | 20  |
| 4.1 | Advantages and Disadvantages of Non-parametric Model-Based Methods . . . . .                                                                                                                                                | 26  |
| 4.2 | Advantages and Disadvantages of Proximity-Based Methods . . . . .                                                                                                                                                           | 28  |
| 4.3 | Advantages and Disadvantages of Subspace Learning Methods . . . . .                                                                                                                                                         | 30  |
| 6.1 | Data description of the 70 datasets included in the training split . . . . .                                                                                                                                                | 39  |
| 7.1 | Final feature set . . . . .                                                                                                                                                                                                 | 59  |
| 8.1 | Average and standard deviation values of different estimators' AP and AUC-ROC measures over 70 orchards. The <i>p-value</i> depicts the statistical significance of the difference between GBOD and the base-lines. . . . . | 64  |
| 8.2 | Depicting the estimators' average AUC-ROC and average precision values over the 31 OOS orchards, when hyperparameters have been selected with HPOD . . . . .                                                                | 67  |
| B.1 | Average Precision . . . . .                                                                                                                                                                                                 | 110 |
| B.2 | Area Under the Curve of the Receiver Operating Curve . . . . .                                                                                                                                                              | 112 |
| B.3 | Time . . . . .                                                                                                                                                                                                              | 114 |

# Chapter 1

## Introduction

Individual tree crown detection and delineation (ITCD) algorithms are a critical player in precision agriculture. This delineation defines the boundary and location where a single tree appears to exist. Using remote sensing and computer segmentation, they give access to advanced monitoring that aid farmers in improving crop yields, utilisation of resource inputs, and improvements in management strategies. ITCD algorithms enable operational high-throughput phenotyping of tree crowns in orchards to collect decision-making information. This highlights that extraction or delineation of the “individual” tree crowns is one of the most critical components of orchard management. ITCD algorithms have advanced from classical machine vision techniques to more reliable and generalisable machine learning approaches. While these methods effectively define boundaries, they continue to estimate several anomalous delineations. Three commonly occurring anomalies in ITCD are over-segmentation (individual trees estimated as many), under-segmentation (or hedging — many trees recognised as one) and false positives (ground pixels estimated as a tree crown) — [Figure 1.1](#). Poorly estimated delineations significantly limit the efficacy of precision agriculture. Over-segmentations and false positives are detrimental as this would trigger farmers to distribute more fertiliser or irrigation to delineations that appear small or dying in comparison to their counterparts. Under-segmentation results in overestimation of yield or may signal growers to implement pruning or thinning strategies for crowns that appear over-grown. Either case results in a waste of resources and a potential reduction in economic yield.

A systematic approach for the identification of poor delineations in ITCD algorithms under the absence of reliable reference data has yet to be introduced. Most ITCD literature recognise poor delineations, few authors have directly attempted to resolve it using heuristic and supervised anomaly detection methods. Given the cost and complexity of labelling data — along with the impact on model generalisation and label accuracy — it is essential to consider an approach that is not reliant on dependable reference data. It is also necessary to cater for all types of orchards.

This work develops and evaluates a reproducible process for the identification of poor delineations occurring amongst ITCD algorithms as a post-processing step using a salient feature set — based on texture features, shape descriptors and spectral indices — and unsupervised anomaly detection. A pipeline was developed to enable automatic, generalisable and efficient anomaly detection in new orchards where reference data is scarce and unreliable, and hyperparameter tuning is undesirable. In order to achieve this, a meta-learner was used to approximate hyperparameters for estimators detecting over new orchards without reference data. A novel approach to multidimensional spatial anomaly detection based on the local Geary  $c$  statistic in a multivariate context (GBOD) is also proposed. This method is able to exploit the physical neighbourhood inherent in the data; thus adding to the literature surrounding multidimensional spatial anomaly detection.

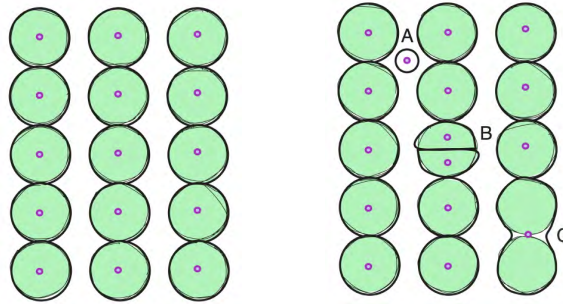


Figure 1.1: Illustrating anomaly classifications in the context of this study: A) false positive. B) Over-segmentation. C) Under-segmentation.

## 1.1 Aims and Objectives

The principle aim of this work was to develop an automatic and generalisable scheme to identify incorrectly estimated delineations often occurring in ITCD algorithms. To this end, a number of research questions were posed and evaluated:

1. How proficient is the proposed feature set at characterising over-segmentations, under-segmentations and false positive delineations in every type of orchard?
2. To what extent is a data-driven unsupervised approach feasible for detecting poor estimates in ITCD algorithms?
3. Does the proposed algorithm outperform the baseline?

## 1.2 Thesis Structure

Chapter 2, *Background*, introduces precision agriculture, highlighting the relevance and the pitfalls of ITCD. Chapter 3, *Data*, proposes a feature set to conduct this study. Chapter 4, *Anomaly Detection*, gives a robust review of all the plausible

methods that can be utilised. It also highlights a gap in literature surrounding multidimensional spatial anomaly detection. Chapter 5, *Proposed Algorithm*, presents a novel approach to add to the literature surrounding multidimensional spatial anomaly detection. Chapter 6, *Experimental Study*, details the set of experiments conducted to answer the research questions and produce a result that satisfies the research aims. Chapter 7, *Exploratory Results*, and 8, *Experimental Results*, are the findings of the tests conducted, finding that this approach is indeed plausible for generalisable and automatic anomaly detection for poor tree crown delineations. Chapter 9, *Discussion*, provides an interpretation of the findings in the results sections. Chapter 10, *Conclusion*, provides the concluding remarks.

# Chapter 2

## Background

This chapter provides a short introduction to precision agriculture and the importance of error-free individual tree crown delineations. It highlights the process of phenotyping that relies on remotely sensed data via drones and efficient computer vision techniques like segmentation. Having ascertained the importance of ITCD algorithms in precision agriculture, this chapter emphasises the appearance of poor delineations that arise, and attempts to solve this in previous literature.

### 2.1 Precision Agriculture and Remote Sensing

An orchard is a plantation of trees that have been purposefully created for food production<sup>1</sup>. Cultivators use orchard management techniques to achieve optimal results from their crop, as management decisions can vary plant growth, fruit quality and yield. Controlling these outcomes requires resource-intensive activities such as pruning, irrigation, fertilisation, thinning and harvesting (Roussos and Gasparatos, 2009; Zhang et al., 2019a). In response to growing concerns — regarding expensive, limited and environmentally harmful resources — precision agriculture has been introduced to help growers sustainably achieve high economic yield (Stafford, 2000). It is an advanced management technology that employs geospatial or computer vision techniques to analyse and interpret spatial data on plant, animal, or soil conditions at a local scale, equipping growers with information to apply site-specific treatments (Stafford, 2000; Alves Varella et al., 2015; Lynham, 2022). This practice incorporates modern technology to enhance the quantity and quality of production while reducing resource requirements associated with farm management. Precision agriculture uses sophisticated tools such as sensors, drones, computer vision and precision irrigation systems, enabling farmers to monitor and manage their crops with greater efficiency (Plant et al., 2000). These techniques allow yield prediction, harvesting

---

<sup>1</sup>“Orchard.” Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/orchard>. Accessed 19 Jun. 2024.

time judgement, disease attack warnings, pruning assessments and irrigation management (Zhang et al., 2021). Unmanned aerial vehicles (UAVs) act as an excellent monitoring tool as they provide high-throughput phenotyping that precisely measures crop health status, geometry, physiology and water stress in real time (Zhang et al., 2019a; Ampatzidis and Partel, 2019).

Plant phenotyping describes relevant information on plant traits: physiological and morphological characteristics (Mu et al., 2018; Stafford, 2000). More recently, remote sensing has been used as a dependable, fast, and non-destructive means of phenotyping (Ye et al., 2023; Huang et al., 2020). It is the process of collecting data from a distance, by exploiting propagated signals such as optical signals or reflected energy (Schneider and Xhafa, 2022). This allows for data collection via many different mediums, such as the human eye, cameras, radar, sonar and others. Typically remote sensing relies on reflected electromagnetic energy, such as image data. This energy, accompanied by analysis techniques, allows the capture of morphological and biophysical information necessary for orchard management activity, enabling growers to assess canopy health, crop growth status, nutrition status, and other factors to guide the application of inputs (Mu et al., 2018; Stafford, 2000).

**Unmanned Aerial Vehicles (UAVs).** UAVs have been chosen as a modern and popular choice for plant phenotyping due to flight scheduling flexibility, low cost and recent advances in sensor miniaturisation (Caruso et al., 2019). They are efficient and inexpensive sources of information for precision agriculture able to perform real-time monitoring at reasonable spatial and temporal resolutions. The utilisation of UAVs manages to mitigate the disadvantages (i.e. limited control, finite spatial resolutions, inconvenient revision times) of other aerial photogrammetry options. These devices can be fitted with various sensors to capture colour, spectral, temperature and geometric features (Plant et al., 2000; Usha and Singh, 2013; Zhang et al., 2021). The combination of this capable aerial device and appropriate sensor selection enables state-of-the-art plant phenotyping.

**Sensors.** RGB, multispectral, thermal, and hyperspectral sensors are frequently used in this domain. Each sensor has its purpose, and the usage of each is determined by the application necessary. RGB sensors are the most widely used for capturing morphological properties. This sensor is both simple and low cost, compared to its available counterparts (Mu et al., 2018; Zhang et al., 2021). Thermal sensors capture temperature changes, which share a strong correlation with transpiration rate and photosynthesis making it an ideal sensor to capture water stress for applications such as irrigation. Multispectral and hyperspectral sensors use the reflectance from emitted electromagnetic signals to capture information on physiological properties. Changes in these physiological properties affect spectral absorption. This results in the reflected signal capturing structural features, chemical composition and many other important biophysical characteristics (Usha and Singh, 2013). Various spectral bands and vegetative indices exhibit noteworthy correlations with these properties. Multispectral sensors are the more affordable of the two, but they only collect up

to seven spectral bands in the visible and near-infrared electromagnetic region. Hyperspectral sensors are more detailed as they capture hundreds of narrow bands, however this increases the complexity during analysis (Zhang et al., 2021).

## 2.2 Individual Tree Crown Detection and Delineation

ITCD algorithms are image segmentation procedures necessary for precise orchard management. Detection is the action of locating trees, and delineation is the process of outlining tree-crowns (Ke and Quackenbush, 2011), where a tree crown is the upper part of a tree consisting of leaves and branches. These ITCD techniques enable researchers and growers to differentiate individual trees as accurately as possible. They are not only valuable for capturing information on fruit tree canopies where often there is a pattern in the layout of the canopy, but they are also popular in capturing forest inventory where one may observe more dense and mixed forestry (Ke and Quackenbush, 2011; Huang et al., 2018). Distinguishing the boundaries of each crown is a method which essentially allows the realisation of high-throughput phenotyping of tree canopies (Mu et al., 2018; Ampatzidis and Partel, 2019). Accurately defining these trees further allows growers to action upon mechanised means of resource/input allocation (Roussos and Gasparatos, 2009; Pforte et al., 2012; Abioye et al., 2020). The appropriate application of ITCD enhances the efficiency of precision agriculture in a versatile and cost-effective manner.

**Classical Image Analysis.** Many authors have attempted to solve this complex problem in unique ways. In the early 1960s, visual interpretation of RGB photography was used as an alternative to field measurement for phenotyping. This increased the efficiency of data collection, but was still resource-intensive. With the progression of sensor technology and computational power, image analysis techniques for ITCD have become widely explored and available to enable efficient phenotyping (Ke and Quackenbush, 2011). Some of the earliest cases of capable ITCD algorithms date back to the early 1990s, when Pinz (1991) utilised Vision Expert Systems to solve the problem. The literature was advanced by other authors such as Gougeon (1995), Culvenor (2002) and Wang et al. (2004) who demonstrated the usefulness of classical computer vision segmentation algorithms to perform ITCD. These authors presented the valley-following algorithm, region-growing algorithm and most recently marker-controlled watershed segmentation, respectively. To this day many authors successfully implement these segmentation algorithms with great results. Huang et al. (2018) utilised bias field images and filtering as a pre-processing detection step followed by marker-controlled watershed segmentation. Lisiewicz et al. (2022) used watershed segmentation with local maximum filtering as a detection step. Roth et al. (2019) has more uniquely utilised algorithms such as the circular Hough transform to perform segmentation.

**Machine Learning Image Analysis.** The classical image analysis techniques are very useful in this space and have continued to be updated in the past many years. However, they are not without their faults. They have demonstrated varying performance over differing species due to their method of segmentation. Furthermore, they are not easily generalisable to different scenes. Machine learning techniques, particularly deep learning has been introduced to remedy this (Zhao et al., 2023). Since 2012, convolutional neural network based ITCD methods have become the standard. These models abstract the crown information in a comprehensible way and are generalisable over different scenes (Zhao et al., 2023; Li et al., 2023). Authors such as Ampatzidis and Partel (2019), Cheng et al. (2020) and Li et al. (2023), have utilised machine learning to perform ITCD, demonstrating its efficiency and advantages in the domain.

**Problems with ITCD Algorithms.** Anomalies are repeatedly identified in ITCD research; as this is a common observation, it is worth exploring the identification of common errors, why they occur and solutions to solving them. In essence, it may prove crucial to provide a tool to researchers to identify these inaccuracies as a post-processing step in ITCD.

### 2.2.1 Difficulties in ITCD

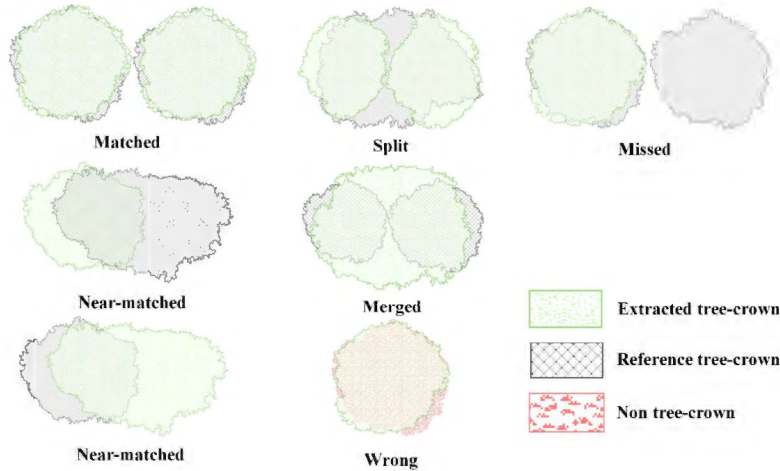


Figure 2.1: Depicting the common results observed in ITCD algorithms (Dong et al., 2020).

ITCD tasks often encounter a degree of error in their results. The common anomalies observed, as depicted in Figure 2.1, are over-segmentation (split-cases), under-segmentation (merged-cases), false positives (wrong-cases), false negatives (missed-cases) and offset (near-matched cases) delineations (Roth et al., 2019; Cheng et al., 2020). These erroneous estimates are observed both in orchard and forest canopy segmentation tasks (Zhen et al., 2016). Over-segmentation is when an individual

tree crown has been estimated as many. Under-segmentation or hedging is when many trees have been recognised as one. False positives in this context relate to areas of the orchard floor that have been identified as a tree crown. False negatives are canopies that have been left unidentified (Roth et al., 2019; Ampatzidis and Partel, 2019). Offset delineations represent an estimate that overlaps with a crown — this estimate is either too small, too large or misaligned and does not characterise the crown correctly. This work addresses over-segmentation, under-segmentation and false positive delineations. False negatives are not represented by a delineation and are therefore not considered further as they may require a different approach for identification. Anomaly detection of this kind may be useful as a separate study much like Leckie et al. (2016) and Donald G. Leckie and Gougeon (2016).

Little literature looks to discuss or resolve poor delineations, most have gone as far as to include these observations in their accuracy measure but have not progressed beyond this (Mu et al., 2018; Li et al., 2023). Often papers that do look into the anomalies do so by making use of their reference data, to engage in a supervised anomaly detection (Vargas-Muñoz et al., 2019; Lisiewicz et al., 2022). Ultimately, the goal is to be able to make predictions of canopy boundaries based on some machine learned model in spite of unreliable field measurements. Investing in schemes that can identify anomalies would greatly benefit this line of literature. The process of identifying anomalies can be done via anomaly detection algorithms, which are introduced in Chapter 4.

## 2.3 Related Work on Anomalous Segmentations

Anomaly detection on manual and estimated delineations is not uncommon. It is regularly used as a post-processing step in contouring tasks that either lack precise annotations or possess unreliable delineations. Kahneman (2011) has stated that “humans are incorrigibly inconsistent in making summary judgments of complex information”, further underscoring humans’ inherent difficulty in precise manual annotation (Akoglu et al., 2015). Manually delineating objects is laborious and time-consuming, while instance segmentation remains exceptionally challenging. Hence, anomaly detection has been successfully used as a post-processing tool for delineation tasks lacking precise annotations or unreliable delineations (Altman et al., 2015; Papaconstadopoulos et al., 2021; Wu et al., 2022). Vargas-Muñoz et al. (2019), endeavoured to rectify detections and delineations in rural building mappings, where annotations of buildings suffer in quality and quantity. Shumilo (2020) observed field border annotations and categorisations, to identify incorrectly detected fields and problematic delineations in field classification tasks. Papaconstadopoulos et al. (2021) utilised anomaly detection as a post-processing mechanism to aid in correcting auto-contouring on low-contrast structures, in medical images. These papers demonstrate that post-processing of delineations is essential in many domains. These ideas can be readily adapted to ITCD delineations.

### 2.3.1 Anomaly Detection of ITCD Algorithms

A few authors have attempted to resolve poor tree crown delineation. [Leckie et al. \(2016\)](#) recognises that good delineations are fundamental to attaining the most beneficial forest inventory attribution for efficient decision-making in forest management. Therefore, [Leckie et al. \(2016\)](#) and [Donald G. Leckie and Gougeon \(2016\)](#) have attempted to resolve this issue using template-matching in forest inventory studies. [Ampatzidis and Partel \(2019\)](#) remedied poor orchard delineations using heuristic techniques based on tree spacing and NDVI. [Cheng et al. \(2020\)](#) looked to gain information where tree canopies intermingled, where it utilised a heuristic based on the average length of canopies to identify under-segmentation. [Lisiewicz et al. \(2022\)](#) used a random forest classifier to learn and predict poor delineations from expertly crafted reference data. [Vargas-Muñoz et al. \(2022\)](#) also used a random forest classifier trained on an estimated probability map to identify false positive and negative delineations.

## 2.4 Feature Engineering

Qualities of trees in ITCD studies are often illustrated through morphological and phenological properties. This section highlights key variables often used in vegetative and image analysis studies that capture these properties.

### 2.4.1 Digital Surface Model

It is helpful to understand elevation modelling nomenclature to understand its usefulness. Digital elevation models (DEM) are gridded raster data used to describe the spatial representation of the Earth's surface height information. A digital surface model (DSM) is a DEM that captures surface information including ground objects such as vegetation and buildings. A digital terrain model (DTM) is a DEM without any ground objects, which represents a "bare-earth" surface ([Julzarika and Harintaka, 2019](#); [Alganci et al., 2018](#)). The definitions of DTM and DEM are often used interchangeably or sometimes synonymously, dependent on where the study is based ([Huang et al., 2020](#)). The importance of the availability of DSM information stems from its traditional use in tree delineation research. [Lisiewicz et al. \(2022\)](#), [Dimitrios Panagiotidis and Chiteculo \(2017\)](#) and [Yousef Erfanifard and Kamińska \(2018\)](#) used canopy height models to delineate tree crowns. This is important as canopy height models are DSM adjusted for slope, elevation and aspect, which is done by subtracting the DTM from the DSM. This is the conventional way that DSM data has been used in these studies. Other researchers, such as [Mu et al. \(2018\)](#), have made use of only DSM data to delineate trees without adjustments. Additionally [Huang et al. \(2020\)](#) demonstrated a multitude of other papers that rely on DSM data to characterise canopy architecture.

## 2.4.2 Spectral Features

Aerial reflectance bands — reflected electromagnetic energy — and vegetative indices give representation to the physiological properties of tree crowns. Physiological properties are captured via absorption rates of the constituents of biochemical molecules within plants. Chlorophyll in particular is a pigment that characterises vegetation greenness. This pigment along with other light-absorbing pigments are the greatest contributors to absorption rates in plants (Manolakis et al., 2016). Electromagnetic bands are useful at classifying vegetation characteristics. In particular, blue and red exhibit very low reflectance over healthy green plants, meaning that light in these wavelengths will mostly be absorbed by these plants. Amongst these same healthy green plants, the full reflectance of the green wavelength will be absorbed. Near-Infrared (NIR) illustrates high reflectance when exposed to vegetation, the value of this reflectance is significantly higher. Red-edge (REG) defines a range of wavelengths existent within the NIR bandwidth that exhibits a strong spike in reflectance as the wavelength increases into NIR (Manolakis et al., 2016). This is the general spectrum of reflectance that researchers encounter when observing healthy vegetation.

In multispectral imagery a finite number of spectral bands are capable of distinguishing vegetation. Vegetative indices have been introduced as a means to deliver more specific vegetative information in much broader categories, while also differentiating green vegetation from other objects in a scene. They are robust measurements that reduce the effect of exogeneous factors (sun angle, atmosphere, etc.) on information. Additionally, they are sensitive to different types of plant physiologies. These indices are formed via ratioing techniques to create composites out of various spectral bands (Manolakis et al., 2016). There are a large number of available vegetative indices to choose from<sup>2</sup>. Radočaj et al. (2023) researched vegetative indices, while considering the subtleties when dealing with precision agriculture. The most popular of these being normalized difference vegetation index (NDVI). It is an index adept at identifying vegetation, along with many other interesting factors to farmers. As a ratio between NIR and Red, it clearly defines chlorophyll content, hence illustrating the presence of green vegetation. Normalized difference red-edge index (NDRE), enhanced vegetation index (EVI), soil-adjusted vegetation index (SAVI), green-normalized difference vegetation index (GNDVI) and optimized soil-adjusted vegetation index (OSAVI) have been developed as optimisations of NDVI under specific settings. These alongside simple ratio (SR) and global environmental monitoring index (GEMI) are the most commonly used vegetative indices (Radočaj et al., 2023).

NDRE is more sensitive to low chlorophyll content, and is useful for assessing the presence of vegetation at early and late periods in the growth stages. This is due to the use of REG, an index especially sensitive to changes in vegetation (Boiarskii and

---

<sup>2</sup>All of which can be found in the [Index Database](#)

Hasegawa, 2019). Densely vegetated areas can demonstrate obscure NDVI values due to the added effect of the canopy background and the atmospheric influences. EVI, is sensitive to dense regions and therefore can be used as an alternative. Conversely, in sparsely vegetated areas, soil reflectance can obscure the results of NDVI; SAVI has been used to mitigate this by reducing the influence of soil brightness in sparse regions (Pettorelli et al., 2005). OSAVI is a specialised version of SAVI, where the soil correction factor is optimised to a value that reduces the impact of soil in both high and low density vegetation coverage (Rondeaux et al., 1996). An alternative measure that reduces the effect of soil background is GNDVI. By replacing the red band in NDVI with green, there is more reflectance observed by the green band when encountering soil and less reflectance when encountering vegetation; essentially allowing for a better visualisation of vegetation in sparse areas (Radočaj et al., 2023).

### 2.4.3 Shape Descriptors

Shape descriptors are used to quantify shape structures. They describe the characteristics of a geometry, therefore simplifying the representation of their appearance. They are used to distinguish unique shapes or retrieve similar ones. There are two classes of these descriptors: contour-based and region-based features. Contour-based descriptors are features that have been extracted from the boundary of the shape. Region-based descriptors are extracted from the whole shape. These are further split into structural and global approaches, where this respectively treats the entire shape as a single entity or in parts (Zhang and Lu, 2004).

A good shape descriptor should be able to distinguish defective shapes and group similar ones, irrespective of scale, translation, rotation and affine transformation. High-level features alone are capable of distinguishing dissimilar shapes, whereas low-level, complex, shape descriptors are more adept at matching similar shapes because they represent more acute details (Zhang and Lu, 2004). Coarse descriptors of shape such as area, perimeter, radius, major axis and minor axis are the most popular in summarising characteristics of the shape. These measures provide information on the size of the shape and how much space the shape takes up. Other coarse descriptors can be seen in Table 2.1. Moment invariants are an example of a powerful low-level shape descriptor. These are one of the most common region based techniques. Amongst moment invariants, Zernike moments perform the best, in that they provide the most robust information (Zhang and Tan, 2002). Since these methods utilise all the pixels within the shape, Zernike moments are also useful as texture descriptors. It should be noted that while considering moment invariants, lower order moments do have physical meaning, while higher order moments are more difficult to interpret (Zhang and Lu, 2004).

Table 2.1: Descriptions of some coarse shape descriptors.

| Data Property                            | Description                                                                                                                                                                                                                                                                                    |
|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Circularity                              | Describes how close to a circle a shape is (Zdilla et al., 2016). A value of 1 is a perfect circle; values closer to 0 are less circular.                                                                                                                                                      |
| Roundness                                | Is similar to circularity but is insensitive to irregular boundaries (Zdilla et al., 2016).                                                                                                                                                                                                    |
| Eccentricity (elongation or ellipticity) | The proportion of the minor axis length to the major axis length or the ratio of the width to the length of the object bounding box (Dražić et al., 2010). For a ratio close to 1, the shape is roughly circular; a shape is more elongated as it increases from 1.                            |
| Compactness                              | A measure of how dense a shape is. It demonstrates how the shape packs its area relative to its perimeter. A circle is considered the most compact shape possible (Bribiesca, 1997).                                                                                                           |
| Convexity                                | Convexity describes the relative amount that the polygon deviates from a convex shape (Pourghahramani and Forssberg, 2005).                                                                                                                                                                    |
| Solidity                                 | Much like convexity, gives a measure of how convex or concave a shape is (Zdilla et al., 2016). A value of 1 signifies a solid object, and a value less than 1 indicates an irregular shape. The solidity of a convex shape is 1; this value decreases the closer to concavity the shape lies. |
| Bending Energy                           | Describes how much energy would have to be used to transform a line segment into the shape of interest. Similar shapes will have similar energy, while dissimilar polygons will not (Young et al., 1974).                                                                                      |

Figure 2.2 shows the information captured at each moment. Zernike polynomials were created by von F. Zernike (1934) and extended to image analysis by Teague (1980). Zernike moments,  $Z_{nm}$ , are defined by von F. Zernike (1934) as follows,

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}^*(x, y) f(x, y), \quad x^2 + y^2 \leq 1,$$

where,

$$V_{nm}^*(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp^{jm\theta}$$

and

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \times \frac{(n-s)!}{s! \left(\frac{(n+|m|)}{2} - s\right)! \left(\frac{(n-|m|)}{2} - s\right)!} \rho^{n-2s}$$

$\rho$  and  $\theta$  are the radius and angle of the pixel  $(x, y)$  with respect to the center of gravity of the shape.  $V_{nm}^*(x, y)$  denotes any one of the orthogonal Zernike polynomials over the unit circle. The symbol  $n$  denotes the order of the Zernike moment and  $m$  the number of oscillations around the pixel (the angular order or repetitions) (von F. Zernike, 1934). Zernike moments are helpful in that similar shapes will have moments whose metric distances are very close to one another, while anomalies should be distinguishable (Zhang and Lu, 2004).

|                     |     |     |       |      |      |     |       |       |      |
|---------------------|-----|-----|-------|------|------|-----|-------|-------|------|
| Order               | One | Two | Three | Four | Five | Six | Seven | Eight | Nine |
| Character           |     |     |       |      |      |     |       |       |      |
| Moment in character |     |     |       |      |      |     |       |       |      |

Figure 2.2: Demonstration of how Zernike moments capture information about an image Kulkarni et al. (2014).

Figure 2.2 illustrates how Zernike moments capture information about an image. It is an example of how the lower order moments capture coarse descriptors such as centroid, area, symmetry and elongation. While higher order moments capture finer less interpretable details.

#### 2.4.4 Texture Features

The final set of features worth considering are those of texture features. These are used to classify regions of interest in an image via calculations of intensities and spatial distribution of colours. Texture can be viewed as a repeating pattern of local variations in image intensities and colour arrangements (Terashima, 2002). Grey-level Co-occurrence texture descriptors are the most popular to date. The co-occurrence matrix (GLCM) is a statistical method used to measure the intensities of grey-tone images (Haralick et al., 1973; Tuceryan and Jain, 1993). The matrices capture how pixel intensities interact with each other. Hence describing intensity relationships within a region using second-order statistical features (Ghalati et al., 2022). GLCM,  $P_d$ , depicts relative frequencies between two neighbouring cells separated by a distance,  $d$ , with intensities  $i$  and  $j$  (Haralick et al., 1973).  $P_d$ , for a displacement vector of  $\mathbf{d} = (dx, dy)$  is,

$$P_d(i, j) = |\{(r, s), (t, v) : I(r, s) = i, I(t, v) = j\}|$$

where  $I(r, s) = i$  and  $I(t, v) = j$  are image intensities at location  $(r, s)$ . In alternative phrasing,  $(i, j)$  represents the frequency of intensity values  $i$  and  $j$  separated by a distance  $d$ , given  $(r, s), (t, v) \in N \times N$  (or  $(r, s)$  and  $(t, v)$  are locations in an  $N \times N$  grid),  $(t, v) = (r + dx, s + dy)$  and  $|\cdot|$  is the cardinality of a set (Tuceryan and Jain, 1993).

Second-order statistical textural features, Haralick features (Haralick et al., 1973), are calculated from the co-occurrence values. These include angular second-moment (ASM) =  $\sum_i \sum_j P_d(i, j)$ , contrast (Con) =  $\sum_i \sum_j (i - j)^2 P_d(i, j)$  and correlation (Corr) =  $\frac{\sum_i \sum_j (i - \mu_x)(j - \mu_y) P_d(i, j)}{\sigma_x \sigma_y}$  (Haralick et al., 1973; Tuceryan and Jain, 1993). These values are calculated using the normalised versions of the co-occurrence values. ASM is a measure of homogeneity of intensities in an image, where a fine texture is closer to 1 and a coarse texture is closer to 0. Homogeneous images will have finer textures as there will be less intensity transitions between pixels. Contrast describes the amount of local variations of pixel intensities. A high contrast value indicates very sharp differences, or a coarse texture; while values closer to zero indicate a much smoother texture. Correlation is a measure of linear dependency between image intensities, areas of constant texture will have low contrast with variable noise, while grainy surfaces exhibit more linear structure (Haralick et al., 1973). These features have been used in texture classification tasks (Tuceryan and Jain, 1993).

## 2.5 Overview

Precision agriculture is heavily reliant on accurate and precise ITCD. This makes it ever more crucial to identify poor delineations estimated from these algorithms. Some studies have attempted this but existing literature relies heavily on heuristics and well-defined reference data to identify poor delineations. No systematic and generalisable approach that is independent of expensive and reliable reference data has been suggested, especially when the presence of reference data is doubtful. This paper proposes a pipeline to solve this problem in such a way, utilising feature groupings such as spectral, shape and texture features to inform reliable, proven and reproducible anomaly detection. This provides a generalisable approach to the detection of poor delineations in ITCD algorithms.

# Chapter 3

## Data and Feature Engineering

This chapter details the data sources provided and the engineered features that make up the basis of this work.

### 3.1 Study Area and Materials

The research has access to 316 unique orchards, each possessing anywhere from 100 to 10000+ trees<sup>1</sup>. RGB aerial imagery, digital surface models (DSM), aerial reflectance bands, and the estimated delineations have been provided for each orchard. The imagery hails from different sites with very expansive orchards which maintain several tree species, both between and within the orchards. One could expect to see any number of species, be it apples, oranges, olives, and others. This rich selection of scenarios provides this study with an exhaustive set of edge cases, while also providing the foundation for this study to demonstrate generalised solutions over different orchard types. [Figure 3.1](#) exemplifies the large number of different orchard types encountered, each of these examples exhibiting varying orchard conditions. Many similarities are shared between the orchards such as, the presence of trees, the existence of a walkway and a roughly linear planting pattern. However, it can be seen that tree species (size, shape, colour, fruit), growing conditions (soil, sunlight, temperature, water stress, plant type, arability), local planting pattern (tree spacing, row spacing, irregular planting patterns) and orchard maintenance (foliage, hedged crowns, unmaintained walkways, uniformly pruned trees, different sized trees) amongst very many traits of the orchard differ greatly between the examples. These differences between the orchards make this problem non-trivial, as certain rule-based ([Donald G. Leckie and Gougeon, 2016](#); [Leckie et al., 2016](#)) and scene specific ([Ampatzidis and Partel, 2019](#)) approaches will not result in a general approach that works over all orchards. These issues motivated a novel approach to this type of study utilising shallow feature mappings — hand-crafted, low-level

---

<sup>1</sup>The data sources in this study were provided courtesy of [Aerobotics](#)

features.

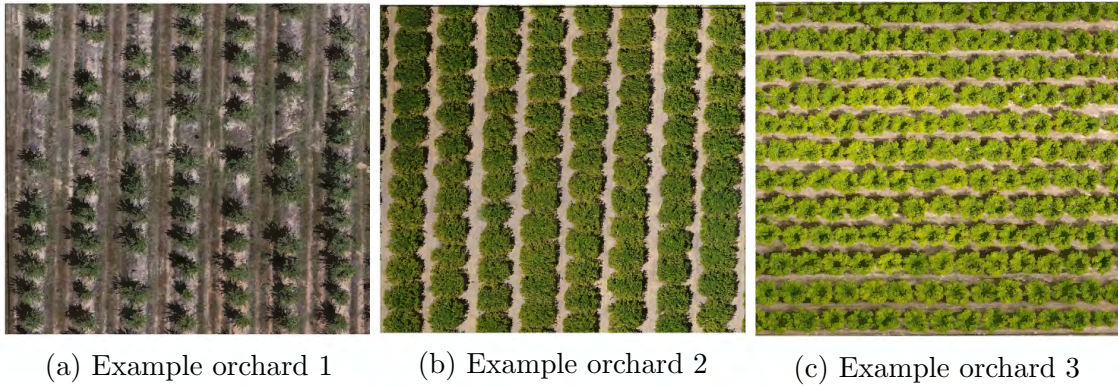
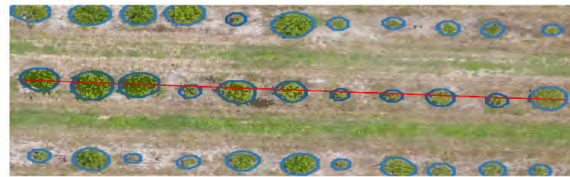
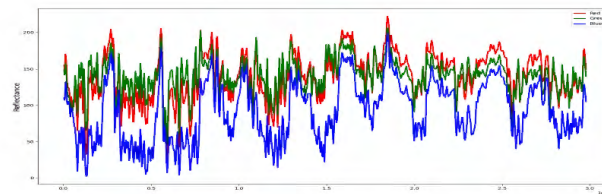


Figure 3.1: Examples of different orchard types

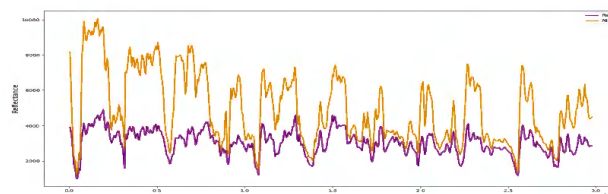
**Aerial Reflectance Bands.** Several spectral bands have been provided: red, green, blue, red-edge and near-infrared. It should be noted that RGB imagery — [Figure 3.1](#) — is simply a composite image comprising red, green and blue spectral bands.



(a) Profile over a number of healthy trees.



(b) Reflectance of Red, Green and Blue wavelengths captured over the profile.



(c) Reflectance of NIR and REG wavelengths captured over the profile

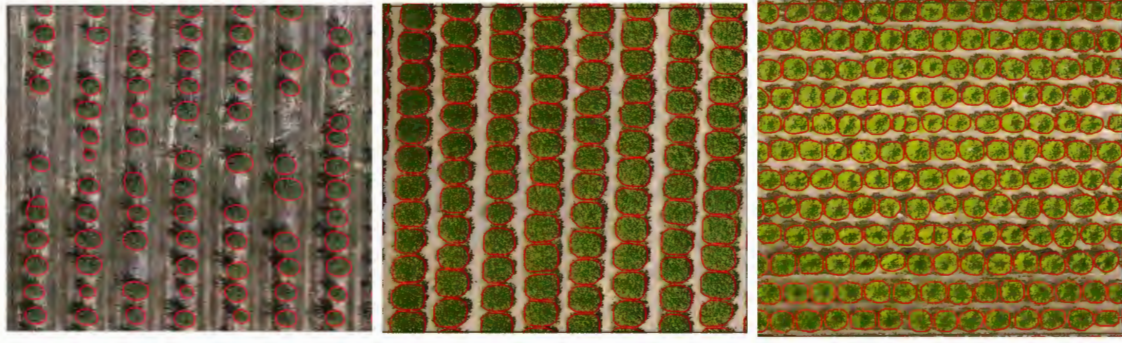
Figure 3.2: Illustrating the organic spectroscopy over an orchard profile.

Using [Figure 3.2a](#) as an example, [Figure 3.2b](#) illustrates that very green and healthy

vegetation results in low red and blue reflectance values, while soil pixels have much higher peaks. Green reflectance is difficult to characterise as the existing green in both soil and tree crowns are given similar importance, resulting in levels that are not easily distinguishable. [Figure 3.2c](#) depicts high reflectance values for both REG and NIR. These bands are able to capture vegetation information more clearly owing to their greater reflectance values and the notable differences between the soil pixel troughs and the healthy vegetation peaks. Empirically, NIR and REG portray similar information, however, NIR looks to be a stronger indicator of healthy vegetation between the two similar indices because of the much greater variability between soil and tree crowns captured. Although, NIR is capable of distinguishing vegetation within the scope of this study, these features alone are too weak. The next section identifies popular and strong identifiers that are used to confidently identify vegetation.

**Digital Surface Model.** DSM data for each orchard has also been supplied. The purpose of such a data source is to aid in identifying false-positive delineations; intuitively soil will have a lower elevation than tree crowns.

**Delineations.** [Figure 3.3](#) demonstrates the final data source provided. These are the estimated delineations of the tree crowns. These delineations have been identified using a mask region-based convolutional neural network (Mask R-CNN) ([He et al., 2020](#)) under an undisclosed model specification. The Mask R-CNN estimated these delineations with a level of confidence. Within these delineations hide a few poor delineations. Human verified reference data was provided alongside these estimated delineations. This is a combination of machine delineations and manual delineations, that have been labelled as such. This data was used to assess performance accuracy of the anomaly detection schemes used. This data also came in the form of boundary polygons. Since this reference data is human-verified, it is more reliable than the estimated delineations, but is also not certainly reliable due to human error exposure. There are various ways that human error may corrupt data. These include the potential for duplicated anomalies, annotation errors, recording errors, missed anomalies and contamination ([Han et al., 2022](#)).



(a) Example orchard 1 with delineations (b) Example orchard 2 with delineations (c) Example orchard 1 with delineations

Figure 3.3: Estimated delineations of orchard tree crowns.

To label the reference data as inlying (normal instances) or outlying (anomalous instances), a comparison was made between the centers of the estimated delineations and the boundary of the reference delineations. If the center of the estimated delineation did not fall within a reference, this estimate was a false-positive. If the centers of many estimated delineations were within a single reference, then these estimates were over-segmentations. As a final check, if many centers of the reference delineations were within a single estimated delineation, then this estimate was under-segmented. Figure 3.4 demonstrates examples of the poor delineations that have been estimated. Figure 3.4a demonstrates false-positive delineations. This is observed by how there is no reference polygon within a number of estimated delineations. Figure 3.4b, demonstrates an over-segmented delineation, this is captured by how one reference polygon is occupied by more than one estimated delineation. Finally, Figure 3.4c illustrates an example of under-segmentation, this is shown by how more than one reference polygon is captured within a single estimated delineation.

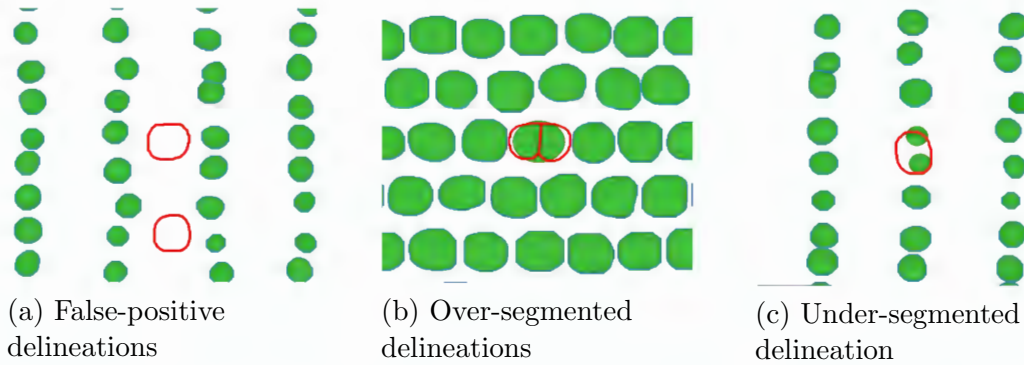


Figure 3.4: Incorrect delineations of orchard tree crowns. Where reference polygons are illustrated as green-coloured polygon, while estimated delineations are represented by blue polygons and erroneous estimates are red.

## 3.2 Feature Engineering

This section describes the set of features — [Table 3.1](#) — curated from the data sources, which have been discussed in groupings of vegetative indices, shape descriptors and texture features. Hand-crafted, low-level and basic shallow feature mappings have been used, as opposed to complex, task-specific, high-level, and abstract deep feature maps. Shallow features capture simpler, more localised information; these features are also more general and transferable across tasks, which is especially useful to our use case ([Santhosh et al., 2020](#); [Smiti, 2020](#); [Pang et al., 2021](#)). In doing this, there is more control to clearly distinguish between outlying and inlying data instances. This also provides simplicity in attributing the cause of outlieriness in poor delineations.

**Vegetative Indices.** A subset of the most useful and popular vegetative indices that cater to the nuances of orchards have been considered. These composite bands are more discriminative of vegetation than aerial reflectance bands. The vegetative indices used are based on [Radočaj et al. \(2023\)](#), who has listed several widely used indices, including: NDVI, EVI, GNDVI, SAVI, NDRE and OSAVI. The inclusion of spectral indices is meant to identify false positive delineations. The hypothesis being that tree crowns will have a significantly different vegetative valuation in comparison to the soil/canopy floor. This concept has been demonstrated by [Figure 3.2](#), where crown and ground pixels illustrate contrasting values. Inherently, one may expect all the trees in an orchard to be at a similar growing stage; therefore, it may be useful to assume that all trees will exhibit similar values within each vegetative index, whereas soil pixels will demonstrate vastly different values.

**Texture Features.** Following [Zhang et al. \(2019b\)](#) and [Papaconstadopoulos et al. \(2021\)](#), GLCM has been used to assess the accuracy of delineations and reject erroneous contours. Due to the nature of this research it is only necessary that these descriptors distinguish tree crowns from soil pixels, therefore, Haralick features should suffice. Higher-order statistical features that capture more complex characteristics would be problematic for this application as these may capture unnecessary additional information such as differences in species and health. Furthermore, most statistics-based approaches describe texture in the form of a histogram or a group of histograms ([Ghalati et al., 2022](#)). Again demonstrating that Haralick features are preferred in this work due to design and dimensionality constraints. GLCM does not cater for rotation, scale, illumination or orientation invariance ([Ghalati et al., 2022](#)). This is trivial as tree crowns take on the shape of circular “blobs”, and therefore appear to share very similar structural properties between crowns. Consequently, they are all comparable whether the tree crowns are rotated or represented at different orientations. Within an orchard the scale at which the textures are being compared are unchanging. However, one should be wary of illumination invariance.

Table 3.1: Feature set used to perform anomaly detection

| Description               | Variable                                     | Description                                                                                                        | References                 |
|---------------------------|----------------------------------------------|--------------------------------------------------------------------------------------------------------------------|----------------------------|
| Shape Descriptors         | Compactness                                  | $Com = \frac{CP^2}{4\pi CPA}$                                                                                      | Bribiesca (1997)           |
|                           | Eccentricity                                 | $Ecc = \frac{MinAL}{MajAL}$                                                                                        | Zhang and Lu (2004)        |
|                           | Convexity                                    | $Conv = \frac{CP}{MajAL}$                                                                                          | Hentschel and Page (2003)  |
|                           | k-Curvature                                  | $\kappa = \frac{convexPerimeter}{ x''y' - x'y'' ^{\frac{3}{2}}}$                                                   | Young et al. (1974)        |
|                           | Bending Energy                               | $E_c = \frac{1}{N} \sum_{p=1}^N \kappa(p)^2$                                                                       | Young et al. (1974)        |
|                           | Circularity                                  | $Circ = \frac{CP^2}{4\pi CPA}$                                                                                     | Zdilla et al. (2016)       |
|                           | Roundness                                    | $Rou = \frac{CP^2}{\pi MajAL^2}$                                                                                   | Hentschel and Page (2003)  |
|                           | Solidity                                     | $Sol = \frac{convexArea}{CPA}$                                                                                     | Zdilla et al. (2016)       |
|                           | Zernike Moments                              | $Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y V_{nm}^*(x, y) f(x, y)$                                                    | Zhang and Lu (2004)        |
|                           | Angular Second Moment                        | $ASM = \sum_{i,j=0}^{levels-1} P_{i,j}^2$                                                                          | Haralick et al. (1973)     |
| Haralick Texture Features | Contrast                                     | $Con = \sum_{i,j=0}^{levels-1} P_{i,j}(i-j)^2$                                                                     |                            |
|                           | Correlation                                  | $Corr = \sum_{i,j=0}^{levels-1} P_{i,j} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$ |                            |
| Aerial Reflectance Bands  | Band 2, Blue (0.45-0.51 $\mu$ m)             | Blue                                                                                                               | -                          |
|                           | Band 3, Green (0.53-0.59 $\mu$ m)            | Green                                                                                                              | -                          |
|                           | Band 4, Red (0.64-0.67 $\mu$ m)              | Red                                                                                                                | -                          |
|                           | Band 5, Red Edge (0.68-0.73 $\mu$ m)         | Reg                                                                                                                | -                          |
|                           | Band 6, Near-Infrared (0.85-0.88 $\mu$ m)    | NIR                                                                                                                | -                          |
| Vegetation Indices        | Normalized Difference Vegetation Index       | $NDVI = \frac{NIR-RED}{NIR+RED}$                                                                                   | Rouse et al. (1974)        |
|                           | Normalized difference red-edge Index         | $NDRE = \frac{NIR-REG}{NIR+REG}$                                                                                   | Barnes et al. (2000)       |
|                           | Green Normalized Difference Vegetation Index | $GNDVI = \frac{NIR-GREEN}{NIR+GREEN}$                                                                              | Buschmann and Nagel (1993) |
|                           | Soil Adjusted Vegetation Index               | $SAVI = \left( \frac{NIR-RED}{NIR+RED+L} \right) \cdot (1+L)$                                                      | Huete (1988)               |
|                           | Optimised SAVI                               | $OSAVI = 1.16 \cdot \frac{NIR-RED}{NIR+RED+0.16}$                                                                  | Rondeaux et al. (1996)     |
|                           | Enhanced Vegetation Index                    | $EVI = 2.5 \cdot \frac{NIR+6\cdot RED-7.5\cdot BLUE+1}{NIR-RED}$                                                   | Huete et al. (2002)        |
| Other Indices             | Confidence                                   | Conf                                                                                                               | -                          |
|                           | Digital Surface Model                        | DSM                                                                                                                | -                          |

In [Figure 3.3c](#), differing levels of sunlight exposure can be seen throughout the orchard due to cloud coverage, time of day, or angle of photo. This is one example where illumination invariance will change the outcome of GLCM. Global anomaly detection methods will struggle under these circumstances, but it is expected that local anomaly detection will be able to cater for this varying illumination. This hypothesis arises from the capability of local anomaly detection schemes to embed context, as neighbouring trees (in the case of spatial anomaly detection) and similar trees (in the case of relative densities) are likely to exhibit comparable illumination within their relative densities. The subset of Haralick features used in this study are ASM, Con and Corr.

**Shape Descriptors.** [Donald G. Leckie and Gougeon \(2016\)](#), [Leckie et al. \(2016\)](#) and [Papaconstadopoulos et al. \(2021\)](#) have used shape descriptors as a key aspect in their analysis. This study mimics [Papaconstadopoulos et al. \(2021\)](#) who has opted for more flexible and generalised descriptions of delineations. These descriptors describe the morphology of the tree crown. Area (CPA), perimeter (CP), radius, major axis (MajAL) and minor axis (MinAL), although popular, are not robust coarse descriptors and have therefore not been considered further. Additionally, some orchards retain a number of trees often in different growing stages to the rest — these crowns are not anomalous, they may either be poorly maintained or in a different growing stage — ultimately resulting in crowns having similar shapes but differing sizes. For this reason it is useful to consider more robust shape descriptors that are scale invariant. Convexity (Conv), eccentricity (Ecc), bending energy ( $E_c$ ), circularity (Circ), compactness (Com), roundness (Rou) and solidity (Sol) represent some of the more robust shape descriptors used in this study. Zernike moments ( $Z_{nm}$ ) have been chosen as a fine region-based global descriptor ([Zhang and Lu, 2004](#)). Circ, Rou and Ecc are measures that are sensitive to elongation, they help classify under- and over-segmented estimates from inliers. Rou captures the same information as Circ but is less sensitive to boundary irregularities. By ignoring irregularities while focusing on the differences between inlying and outlying shapes, it holds the potential to be more discriminative than Circ. Conv highlights that over-segmentations have a more concave shape due to how the ITCD algorithm forces many estimates into a small area, therefore forcing parts of the polygon to bend away from an expected convex shape. Sol provides similar information as Con but in a region-based manner ([Hentschel and Page, 2003](#)). Finally,  $E_c$  is a measure that should be able to give inliers very similar levels of energy, while outliers with varying shapes may stray from this expectedly common bending energy.  $Z_{nm}$  are also considered in this study as fine shape descriptors and translation, orientation, scale and geometric invariant texture descriptors ([Zhang and Tan, 2002](#)).

### 3.3 Summary

Using the data sources, texture features, shape descriptors and spectral indices were engineered to produce a feature set — [Table 3.1](#) — to identify poor delineations. It has been hypothesised that shape descriptors are specifically included to deal with over- and under-segmentations, vegetative indices capture false-positives, while texture descriptors can be used to capture all kinds. The proposed feature set enables unsupervised anomaly detection that is focused on capturing patterns in the data.

# Chapter 4

## Anomaly Detection

Anomaly detection (AD) — otherwise known as outlier detection — is the observation of events that deviate from the expected pattern of well-defined normal events in a dataset ([Hawkins, 1974](#)). This chapter details relevant information to understand anomaly detection. First, it looks at how to define different types of anomalies, the learning paradigms that can be implemented to resolve this problem and how anomalousness/outlierness is scored. Following this background information, the remainder of the chapter highlights relevant fundamental and state-of-the-art algorithms.

### 4.1 Background

#### 4.1.1 Types of Anomalies

Traditionally, one may encounter three types of anomalies: point, contextual, and collective ([Chandola et al., 2009](#); [Santhosh et al., 2020](#)). A point anomaly (or global outlier) is a single observation that deviates from the pattern of the rest of the data. A false positive and under-segmented delineations are theorised to be global anomalies in this study, as they represent observations that vary greatly at a global scale. A collective anomaly is a grouping of data points that differ with respect to the entire dataset. Individual observations within the collection may not be anomalies, but their occurrence together makes them anomalous. A contextual anomaly (or local outlier) is an irregular data point only in a specific context but not otherwise. These are reliant on the extraction of contextual information from the data. When the behavioural attributes in a specific context deviate significantly, that's when we observe contextual anomalies. Context can be embedded using locational information, or a description of the neighbourhood relative to an instance. By incorporating contextual information both a point anomaly and a collective anomaly can be viewed as a contextual anomaly. Both under-segmentations and false positives can be turned into contextual outliers. However, when only considering relative densities, over-segmentations are hypothesised to be the local outliers present in this

study, as they may be common out of context.

### 4.1.2 Learning Paradigms

Many anomaly detection techniques exist, which can be grouped into several learning paradigms — supervised, semi-supervised, unsupervised and hybrid — depending on the presence of data labels (Chandola et al., 2009; Santhosh et al., 2020). Anomalies are rare entities, making them challenging to label, and this rarity of anomalous observations often causes class imbalances (Chalapathy and Chawla, 2019). In general, labelling datasets in anomaly detection applications is expensive. While a fully labelled dataset allows the usage of supervised learning techniques, these often require a balanced representation of anomalous and normal instances to build a predictive model. The sparseness of anomalies and the expense of data labelling are the main shortcomings of supervised approaches. Alternatively, it is possible to only have one labelled class — normality or abnormality — where the aim is to model this one class to detect the other. This is the case for semi-supervised methods. Unsupervised learning approaches are the most widely used as they do not rely on training data. Unsupervised models perform anomaly detection based on the intrinsic properties of the data (Hodge and Austin, 2004; Chandola et al., 2009; Chalapathy and Chawla, 2019). The last paradigm that one may expect to encounter is hybrid learning algorithms. These combine the abovementioned techniques and others to form hybrids. Active learning, where learning happens through interactive labelling of data (Chen et al., 2018) and self-supervised learning are examples of hybrid techniques (Santhosh et al., 2020).

### 4.1.3 Anomaly Scoring

Anomaly detection algorithms use one of two methods to reflect outlierness — the degree to which an observation is outlying — of data instances. The first of these happens by giving each observation a score, as a real number, that depicts their degree of outlierness. While other methods may directly assign a label (anomaly or normal) to each instance (Chandola et al., 2009). Using the scoring approach to anomalous flagging provides the user with more information than binary labels. The caveat to this is a lack of standardisation, making scores incomparable between methods (Wang et al., 2019). Therefore it is useful to classify observations in a binary manner. To transform anomaly classifications to binary, studies have applied thresholding to the real number scores, as demonstrated below,

$$\text{label} = \begin{cases} \text{outlier,} & \text{if } s(x) \geq \tau \\ \text{inlier,} & \text{if } s(x) < \tau, \end{cases}$$

where  $s$  denotes the anomaly score, and  $\tau$  the decision threshold (Ruff et al., 2021). The choice of threshold plays a key role in minimising the appearance of type I (false alarms) and type II (missed anomalies) errors. An increase in  $\tau$  would result in more

type type II errors and a decrease would result in more type I errors. Visualisations, like the receiver operating curve or precision-recall curve, are crucial for deciding the appropriate threshold (Ruff et al., 2021).

## 4.2 Fundamental and State-of-the-Art Algorithms

The literature presented in Section 2.3.1 identified poor delineations using heuristics and supervised models, both heavily reliant on reference data and visual interpretation. However, unreliable reference data implies that neither approach is suitable for detection. Due to the expectation of unreliable reference data, this study considers unsupervised techniques. These techniques enable automatic and generalisable anomaly detection over every type of orchard, independent of reliable reference data.

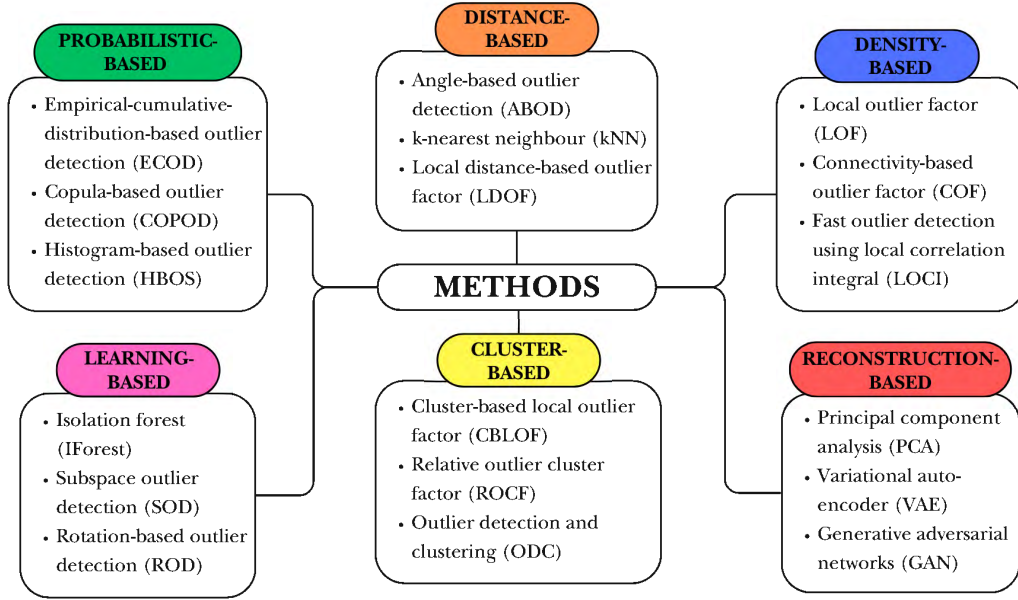


Figure 4.1: Fundamental and state-of-the-art anomaly detection methods.

This section presents anomaly detection methods which aim to identify inaccuracies in a generalisable and automatic way. The choice of approach is not limiting, however, selecting the right method remains crucial for optimising detection within a dataset. The chosen approach will influence scalability, speed, modelling accuracy, model distribution and model capabilities (Hodge and Austin, 2004). Widely used and state-of-the-art anomaly detection algorithms suitable for this task can be grouped under model-based, distance-based, density-based, reconstruction-based, subspace learning and spatial methods, as depicted in Figure 4.1 (Hodge and Austin, 2004; Chandola et al., 2009; Ruff et al., 2021).

### 4.2.1 Diagram-Based Methods

One of the simplest techniques is the usage of boxplots to identify extreme values in univariate or multivariate datasets. In the univariate case, anomalies are points that lie beyond the whiskers of these plots (Hodge and Austin, 2004). In the multivariate case, inter-attribute dependencies must be accounted for; Laurikkala et al. (2000) suggests using a reduced sub-ordering based on the Mahalanobis distance measure to embed the dependencies before identifying outliers. There are several simplistic and diagrammatic approaches that can be used to visualise anomalies, much like boxplots. These methods often specifically require human intervention unless guided by some heuristic (Hodge and Austin, 2004). Even with a well-defined heuristic, these techniques lack accuracy (Shi et al., 2017).

### 4.2.2 Model/Probabilistic-Based Methods

Table 4.1: Advantages and Disadvantages of Non-parametric Model-Based Methods

| Method       | Advantages                                                                                                                      | Limitations                                                                                             |
|--------------|---------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| Kernal-based | + Detect global outliers<br>+ Detect local outliers                                                                             | - Computationally Expensive<br>- Hyperparameter Intensive<br>- Curse of Dimensionality                  |
| HBOS         | + Detect global outliers<br>+ Computationally efficient<br>+ Supports high dimensionality                                       | - Cannot detect local outliers<br>- Sensitive to bin sizes<br>- Unaware of inter-attribute dependencies |
| COPOD        | + Hyperparameter-free<br>+ Computationally efficient<br>+ Aware of inter-attribute dependencies<br>+ Capable in high dimensions | - Cannot detect local outliers                                                                          |
| ECOD         | + Computationally efficient<br>+ Hyperparameter free<br>+ Supports high dimensionality                                          | - Cannot detect local outliers<br>- Unaware of inter-attribute dependencies                             |

Model-based methods learn the parameters of candidate algorithms to fit a hypothesis to the data (Chandola et al., 2009; Santhosh et al., 2020). These are further split into parametric and non-parametric sub-groups (Wang et al., 2019). Parametric methods are often supervised and thus not considered further. Under non-parametric techniques, the hypothesis set is not defined a priori; the given data determines the model structure. These have been further separated into smaller groupings: histogram-based, kernel function-based, Bayesian approaches and many

others (Latecki et al., 2007; Chandola et al., 2009; Santhosh et al., 2020; Smiti, 2020). Histogram-based outlier score (HBOS) (Goldstein and Dengel, 2012), uses estimated densities to determine anomaly scores. Copula-based outlier detection (COPOD) (Li et al., 2020) observes outliers as tail events of empirical Copulas. One of the latest and most dependable non-parametric detection methods (Zhao et al., 2019) is the empirical cumulative distribution functions for outlier detection (ECOD) (Li et al., 2022). This algorithm takes a d-dimensional dataset,  $\mathbf{X}$ , and fits empirical cumulative distribution functions,  $\tilde{F}_{\mathbf{X}}$ , and utilises the tail probabilities of these distributions to determine outlierness. Table 4.1 illustrates the limitations and advantages for these types of methods.

### 4.2.3 Proximity-Based Methods

In these methods normal instances occur in dense neighbourhoods (a region where many data instances cluster), while anomalies fall outside of these or appear far from their closest neighbour (Chandola et al., 2009). Proximity methods can be split into two categories: distance-based and density-based methods. Knorr and Ng (1998) first introduced the meaning of a distribution-free distance-based outlier. In distance-based methods, similarity based on distance related metrics are calculated between observations and their nearest neighbours to determine outlierness. Anomalies are observations that do not have enough neighbours or those that appear far from their closest neighbour (Chandola et al., 2009; Smiti, 2020). Distance based methods are mostly used for detecting global outliers (Chawla and Sun, 2006; Wang et al., 2019). k-nearest neighbours based outlier detection (kNN) (Knorr and Ng, 1998; Ramaswamy et al., 2000), angle-based outlier detection (ABOD) (Kriegel et al., 2008) and local distance-based outlier factor (LDOF) (Zhang et al., 2009) are some of the most common distance-based methods. Angle-based outlier detection (ABOD) (Kriegel et al., 2008) is a unique distance based method that leverages the variance of all possible weighted angles between differing permutations of paired-distance vectors to produce outlier scores. However, its consideration of every possible permutation of the entire dataset is slow. fastABOD, an adaptation presented in Kriegel et al. (2008), uses pairs between the k-nearest neighbours instead of the entire dataset to speed up the run-time.

Density-based methods emphasise the detection of local outliers. If the relative density of an observation differs significantly from its neighbourhood or an instance lies in a low density region, it may be considered an anomaly (Chandola et al., 2009; Smiti, 2020). Local outlier factor (LOF) (Breunig et al., 2000) is the foundation of density-based methods. LOF endeavours to distinguish outlying objects relative to the density of their local neighbourhood, enabling outlier detection in different densities. It is the average of the ratio between the relative density of an instance and the relative densities of the k-nearest neighbours of an instance. Many adaptations of LOF have been created and it remains one of the most widely used and benchmarked methods to date. Some popular adaptations include connectivity-based outlier factor

(COF) (Tang et al., 2002), Fast outlier detection using the local correlation integral (LOCI) (Kitagawa et al., 2003). While other methods not based on LOF have been recently introduced: Tang and He (2017) and Ning et al. (2018). Table 4.2 shows the advantages and limitations of using these methods.

Table 4.2: Advantages and Disadvantages of Proximity-Based Methods

| Approach | Method | Advantages                                                         | Limitations                                          |
|----------|--------|--------------------------------------------------------------------|------------------------------------------------------|
| Dist.    | kNN    | + Detect global outliers                                           | - Curse of Dimensionality                            |
|          |        | + Distribution-free                                                | - Cannot detect local outliers                       |
|          | ABOD   | + Hyperparameter free                                              | - Sensitive hyperparameters                          |
|          |        | + Capable in high dimensions                                       | - Computationally expensive                          |
| Density  | LDOF   | + Detect global outliers                                           | - Curse of dimensionality when using kNN alternative |
|          |        | + Detect local outliers                                            | - Cannot detect local outliers                       |
|          | LOF    | + Detect local outliers                                            | - Curse of dimensionality                            |
|          |        | + Detect global outliers                                           | - Cannot handle complex data distributions           |
|          | COF    | + Computationally efficient                                        | - Sensitive to densities                             |
|          |        | + Detect global outliers                                           | - Sensitive hyperparameter                           |
|          | LOCI   | + Detect local outliers                                            | - Curse of dimensionality                            |
|          |        | + Robust in low density regions                                    | - More computationally expensive than LOF            |
|          | LOCI   | + More robust detector than LOF in non-outlier low density regions | - Sensitive hyperparameter                           |
|          |        | + Handles multi-granularity                                        | - Computationally expensive                          |

#### 4.2.4 Cluster-Based Methods

These methods group similar observations together into clusters: both small clusters and objects that don not fit into any cluster can be considered anomalies. Clustering is not anomaly detection, but can be used for this. The efficacy of cluster-based techniques for anomaly detection depends on the algorithm’s success in capturing the composition of normal instances within clusters (He et al., 2003; Chandola et al., 2009; Shi et al., 2017; Wang et al., 2019). Due to this dependency and the hyperparameter intensive nature of these methods, they will not be considered further as they are sensitive to changes in data, which is impractical for generalisability and automation.

### 4.2.5 Learning-Based Methods

These are a subset of model-based methods, but deserve their own section due to the vast literature available on them.

**Subspace Learning.** This offers a means to perform anomaly detection with different permutations of relevant subspaces (Aggarwal, 2016). The curse of dimensionality masks outliers, as irrelevant features act as noise in an increasingly sparse dataset. Therefore, detecting anomalies in a relevant subset of dimensions helps reduce noise and highlight outliers as they may only exist in locally relevant subspaces. Isolation forest (IForest) (Aggarwal, 2016) is one of the most widely used subspace learning methods. IForest builds an ensemble of isolation trees and anomalies are identified as observations that have short average path lengths on the isolation trees, as seen in Figure 4.2. This method sequesters instances using multiple binary trees, by recursively branching each leaf node on the observations of a randomly selected dimension (Liu et al., 2008).

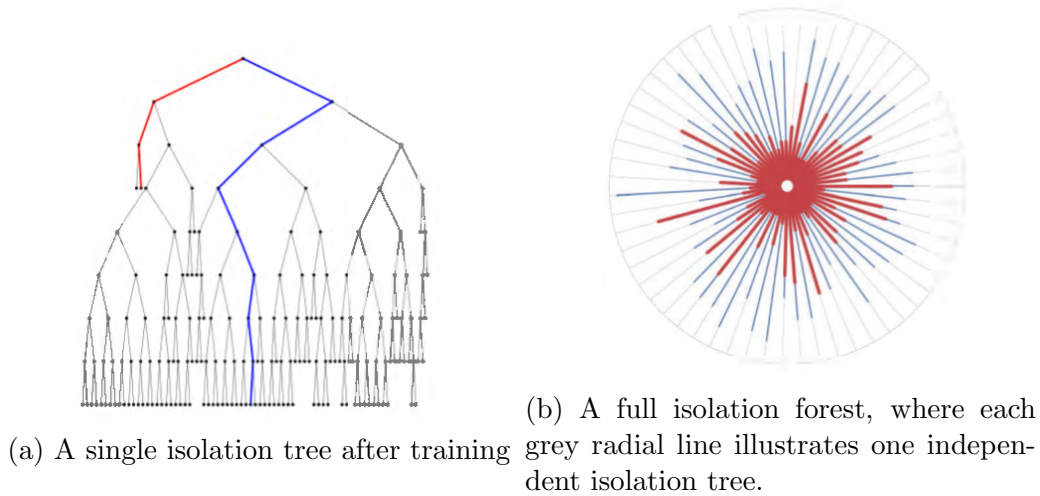


Figure 4.2: A representation of how isolation forests work by Hariri et al. (2021). Red represents an anomalous observation while blue represents a normal observation

The grey lines in Figure 4.2b illustrates the full depth of each tree, the blue lines illustrate the depth that an individual normal instance reached and the red radial lines illustrate the depth that an individual anomalous instance reached. These same normal and anomalous observations are shown to uniquely traverse a single tree in Figure 4.2a. In Figure 4.2 the anomaly is quickly isolated, while the normal points isolate much closer to the maximum depth. For this reason the average depth of isolation for each point over all trees in a forest will indicate outlierness (Liu et al., 2008).

Table 4.3: Advantages and Disadvantages of Subspace Learning Methods

| Method | Advantages                                                                     | Limitations                                                                                     |
|--------|--------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| ROD    | + Hyperparameter free<br>+ Distribution Free<br>+ Supports high dimensionality | - Computationally expensive                                                                     |
| SOD    | + Detect local outliers<br>+ Supports high dimensionality                      | - Computationally expensive<br>- Sensitive hyperparameter<br>- Idealistic modelling assumptions |
| IF     | + Supports high dimensionality<br>+ Model-free<br>+ Computationally efficient  | - Axis-parallel split bias                                                                      |

Rotation-based outlier detection (ROD) ([Almardeny et al., 2022](#)) is another subspace learning method that establishes outlier scores via rotations of three-dimensional subspaces. Subspace outlier detection (SOD), obtains the axis-parallel subspace defined by an observations neighbours, and determines the observation degree of deviation from this to distinguish outlierness ([Kriegel et al., 2009](#)). The advantages and disadvantages of subspace learning methods have been depicted in [Table 4.3](#).

**Deep Learning-Based Methods.** These are another class of learning-based techniques. These typically also fall under parametric model-based methods ([Santhosh et al., 2020](#)). There is evidence indicating their better performance in detecting outliers, their ability to learn deep feature mappings, and their more flexible and superior ability to better discriminate the boundary between inliers and outliers. Although they demonstrate significant performance, these methods are not ideal for the problem description. These methods rely on the availability of large labelled datasets to learn the characteristics of anomalies, which is presumed inaccessible in this study ([Wang et al., 2019](#); [Kwon et al., 2019](#)). They are hyperparameter intensive, computationally expensive, and difficult to interpret ([Li et al., 2022](#)).

#### 4.2.6 Spectral/Reconstruction-Based Methods

These methods are designed to project data into a lower-dimensional space, where normal and anomalous instances frequently show marked differences ([Chandola et al., 2009](#); [Wang et al., 2019](#)). A fundamental example of this technique is principal component analysis based outlier detection (PCA) ([Shyu et al., 2003](#)), which utilises robust principal components analysis, to determine outlierness by the distance between an observation from its component. There are many variations that rely on PCA ([Hawkins, 1974](#); [Parra et al., 1997](#)). The primary limitation of PCA is that it is effective only for linear relationships among features ([Ruff et al., 2021](#)). In contrast, many manifold learning techniques can be alternatives for obtaining a lower-dimensional subspace from nonlinear data. However, these methods often re-

quire extensive tuning of numerous hyperparameters. They are difficult to interpret, complex to structure and computationally complex (Wang et al., 2019).

#### 4.2.7 Graph-Based Methods

Graphs are a structured representation of data based on nodes interconnected by edges. Graph-based methods are those which exploit inter-dependencies, as well as local and global correlations existing in this representation. They are efficient at capturing relational anomalies. These methods are often complex to represent and computationally complex (Akoglu et al., 2015). Nevertheless, the relevance of these graph based methods in this study will be presented in the following subsection.

#### 4.2.8 Spatial Anomaly Detection

The existence of spatial data is unique as it comprises spatial attributes, that carry locational information (coordinates) of entities, and non-spatial attributes, which possess behavioural information (Shukla and Lalitha, 2023). Many studies have shown the benefits of using specialised techniques that consider spatial attributes, addressing spatial dependence, heteroscedacity and heterogeneity (Ernst and Haesbroeck, 2017; Shukla and Lalitha, 2023). Since the delineations presented in this study occur in a spatial context, it is possible to exploit relational dependencies via physical neighbourhoods enabling the capture of spatial anomalies (Shi et al., 2017). These are defined by Shekhar and Chawla (2003) as, “a spatially referenced entity whose non-spatial attribute values are significantly different from those in its spatial neighbourhood”. While considering graph-based methods, physical/spatial neighbourhoods can be distinguished via graph connectivity, inherently capturing spatial relationships and dependence, allowing for the identification of these spatial anomalies (Shekhar et al., 2002; Chawla and Sun, 2006). Multiple univariate spatial dependency and detection methods have been constructed to exploit this (Anselin, 1995; Shi et al., 2017, 2018; Deng et al., 2018). However, there is limited literature that uses spatial dependence in a multivariate context (Filzmoser et al., 2014), and even less that conduct spatial outlier detection in a multidimensional setting. Most spatial outlier detection focuses on a single indicator. Due to the nature of this study and the crafted multidimensional feature set to identify outliers, this paper proposes a novel approach for multidimensional spatial anomaly detection.

### 4.3 Summary

This chapter has highlighted the most crucial information about anomaly detection. In the background, three types of anomalies were characterised. It has been hypothesised that false positive and under-segmentations fall under global outliers, while over-segmentations are an example of local outliers. Different learning approaches were explored, however, only unsupervised approaches have been considered further

owing to the emphasis on good detection despite lacking reference data. The chapter has demonstrated a number of potential solutions and identifies a lack of literature surrounding multivariate spatial anomaly detection.

# Chapter 5

## Proposed Algorithm

Chapter 4 introduced a large number of anomaly detection algorithms, many of which are capable of solving this problem under the design of the feature set. However, none of the methods are able to take advantage of the physical neighbourhood and the multi-dimensionality of the orchard datasets. Furthermore, there exists little literature in the domain of spatial anomaly detection that fully takes advantage of multi-attribute feature sets. To this end, this chapter introduces a simple yet effective use of local Geary's  $c$  statistic in a multivariate context for multi-dimensional spatial anomaly detection. This is also known as multivariate local indicators of spatial association (Multivariate LISA).

This section is divided into two parts. The first will give the theoretical background behind Multivariate LISA, the fundamental portion of the algorithm, while the second part will describe how to utilise this for automatic, generalisable and parameter-free anomaly detection.

### 5.1 Formulation of Multivariate LISA

Univariate local spatial autocorrelation measures the relationships between an observation and its immediate neighbourhood, thereby describing spatial dependence and heterogeneity in a local setting. In doing so, it becomes straightforward to illustrate anomalous locations — more often than not spatial neighbourhoods (Anselin, 1995) — whose behavioural attributes deviate in their neighbourhood (Cressie, 1993; Anselin, 1995; Filzmoser et al., 2014). This finding is consistent with Shi et al. (2017) who stated that “The construction of spatial proximity relationships and the measurement of non-spatial attribute differences are two critical factors in discovering spatial anomalies accurately and sufficiently”. This perspective is further supported by other studies: Chawla and Sun (2006) Shekhar and Chawla (2003). This idea can be cast into a multidimensional case using local Geary's  $c$  statistic in a multivariate context, introduced by Anselin (2019). This is simply an extension of the

local Geary's  $c$  statistic or local indicators of spatial association (LISA) (Anselin, 1995). Geary (1954) introduced a univariate global spatial autocorrelation measure, Geary's  $c$ ,

$$c = \frac{\frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{2 \sum_i \sum_j w_{ij}}}{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} \quad i \neq j \quad (5.1)$$

where  $x_i$  is the observation of the attribute at location  $i$ ,  $\bar{x}$  is the attributes mean,  $n$  is the number of observations, and  $w_{ij}$  is the spatial weights matrix which embeds the structure of the physical neighbourhood. The numerator of Equation 5.1 is the weighted square difference between an observation  $i$  and all its neighbours  $j$ , while the denominator is the unbiased variance of the attribute. The Geary's  $c$  statistic is calculated between an observation and all its neighbours and the sum of the statistics enables the global measure of spatial autocorrelation. Anselin (1995), later developed a local variation of this measure, called local indicators of spatial association (LISA or local Geary's  $c$ ),

$$c_i = \sum_{j \in \mathcal{N}(i)} w_{ij} (x_i - x_j)^2. \quad (5.2)$$

Equation 5.2 is simply the weighted square distance between an observation,  $i$ , and its immediate neighbours,  $\mathcal{N}(i)$ . This formulation can be extended to the multivariate context. If we consider  $X$  and  $Y$  to be different dimensions, the squared distance,  $d_{ij}^2$ , in two-dimensional attribute space between observations at location  $i$  and  $j$  is given by,

$$d_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2. \quad (5.3)$$

This can be seen as the squared Euclidean distance between two observations. It is crucial that the physical neighbourhood is incorporated. Therefore, it is useful to consider the weighted average of Equation 5.3 between location  $i$  and all its neighbours,

$$\sum_{j \in \mathcal{N}(i)} w_{ij} d_{ij}^2 = \sum_{j \in \mathcal{N}(i)} w_{ij} [(x_i - x_j)^2 + (y_i - y_j)^2] \quad (5.4)$$

$$= \sum_{j \in \mathcal{N}(i)} w_{ij} (x_i - x_j)^2 + \sum_{j \in \mathcal{N}(i)} w_{ij} (y_i - y_j)^2 \quad (5.5)$$

$$= c_{1,i} + c_{2,i} \quad (5.6)$$

where  $c_{p,i}$  denotes, the local Geary's  $c$  statistic at location,  $i$ , for dimension,  $p$ . This demonstrates that the local Geary's  $c$  statistic is additive over the attribute space. It can therefore be further expanded for any number,  $K$ , of attributes,

$$c_i = \sum_{p=1}^K c_{p,i} = \sum_{p=1}^K \sum_{j \in \mathcal{N}(i)} w_{ij} (x_{ip} - x_{jp})^2. \quad (5.7)$$

Equation 5.7 depicts the equation of local Geary's  $c$  statistic in a multivariate context for a single observation (or location).

## 5.2 Algorithm of Geary-Based Outlier Detection (GBOD).

Equation 5.7 can be used to perform contextual anomaly detection given a spatial dataset  $\mathbf{X}_{n \times p}$  with  $n$  observations and  $p = 1, \dots, K$  dimensions. This is given in three parts. First, we generate the spatial weights matrix,  $\mathbf{W}_{n \times n}$ , based on Delauney triangulation to account for the physical neighbourhood (Shi et al., 2017). Secondly, the adjacency matrix is used to obtain the multivariate LISA,  $\mathbf{c}_i$ , for every observation. Finally, we define a function  $f : \mathbb{R}^+ \rightarrow [0, 1]$  that scales a positive real number to a probability. This final step ensures consistency, making the method directly comparable to the state-of-the-art. As no function directly maps  $f : \mathbb{R}^+ \rightarrow [0, 1]$ , the statistic was first zero centred by subtracting the mean of the scores from the scores, transforming  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ . In this state, the inverse logarithm or the expit function,  $\text{expit}(x) = \frac{e^x}{1+e^x}$ , performed the appropriate scaling,  $f : \mathbb{R} \rightarrow [0, 1]$ . This is demonstrated in Algorithm 1.

---

### Algorithm 1 Proposed Algorithm using Multivariate LISA

---

**Input:**  $\mathbf{X} \in \mathbb{R}^{n \times p}$  - input dataset

**Output:**  $\mathbf{O} \in [0, 1]^{n \times 1}$  - Anomaly score

- 1: Use Delauney triangulation to obtain spatial weights matrix,  $\mathbf{W}_{n \times n}$ .
  - 2: **for** each dimension  $j$  in  $1, \dots, d$  **do**
  - 3:     **for** each observation  $i$  in  $1, \dots, n$  **do**
  - 4:         Calculate the univariate Local Geary's c statistic,  $c_i$ , of observation  $x_i$ :  

$$c_{p,i} = \sum_{j \in \mathcal{N}(i)} w_{ij} (x_{ip} - x_{jp})^2 \quad \forall j \neq i$$
  - 5:         Resulting in a matrix of scores,  $\mathbf{C}_{n \times p}$ , for each observation and dimension
  - 6:     **end for**
  - 7: **end for**
  - 8: Calculate the row sum of  $\mathbf{C}$ , to get the Multivariate LISA score:  

$$\mathbf{c}_i = \sum_{p=1}^K c_{p,i}$$
  - 9: Calculate the mean of  $\mathbf{c}$ :  $\mu_c = \frac{\sum_{i=1}^n c_i}{n}$
  - 10: **for** each score  $i$  in  $1, \dots, n$  **do**
  - 11:     Zero centre the score of observation  $x_i$ :  $\mathbf{c}_i = \mathbf{c}_i - \mu_c$
  - 12:     Set the outlier score of observation  $x_i$  to scale  $[0, 1]$ :  $O_i = \frac{e^{c_i}}{1+e^{c_i}}$
  - 13: **end for**
  - 14: **return** Outlier Scores  $\mathbf{O} = (O_1, O_2, \dots, O_n)$
- 

First, we look to define an appropriate spatial weights matrix,  $\mathbf{W}_{n \times n}$ , to represent the physical neighbourhood. Delauney triangulation is used to define the neighbourhood criterion. For a comprehensive explanation on this triangulation, see Berg et al. (2008). It is the most direct and optimised graph representation for point patterns (Bivand et al., 2008; Shi et al., 2017). The neighbour relationships are defined in a neighbours matrix by the triangulation, while the links between observations are

appropriately assigned weights. For this application, the weights matrix is an adjacency matrix where edges between vertices are given a weight of 1 and vertices that are not neighbours of one another are given a weight of 0. Having collected the spatial weights matrix, Equation 5.7 is now used. Subsequently calculating the local Geary's  $c$  statistic for each observation of each dimension,  $x_{ip} \in \mathbf{X}_{n \times p}$ , where  $i$  denotes the observation/location. Finally, to make the statistic directly comparable to other anomaly detection techniques, the local Geary's  $c$  statistic is transformed to the appropriate scale. As there is no function that directly maps  $f : \mathbb{R}^+ \rightarrow [0, 1]$ , a first step would be to make the statistic zero centred by subtracting the mean of the scores from the scores. This would transform  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ . In this state the expit function,  $\text{expit}(x) = \frac{e^x}{1+e^x}$ , is able to perform the appropriate scaling from the real number line to a probability,  $f : \mathbb{R} \rightarrow [0, 1]$ .

The usage of this methodology for multivariate spatial anomaly detection at the time of this study has not been observed, making this a novel utilisation of Multivariate LISA. Coupled with this, the proposal to use Delauney triangulation for the neighbourhood criterion and the transformations performed aid in making this approach both hyperparameter-free and highly interpretable.

### 5.3 Interpretability

Interpretability provides a means to assess model explainability in a reliable way. This is vital as it provides researchers with insights on which dimensions are most responsible for the identification of the outlier. This is possible since Multivariate LISA finds the local Geary's  $c$  statistic for each dimension and for each observation. Therefore, allowing attribution of outlieriness for anomalous instances. If an observation has a large outlier score, the additive nature of the Multivariate LISA should identify which of the contributing dimensions have the highest scores. We can follow a similar logic as Li et al. (2020) and Li et al. (2022) to identify the causal dimensions of anomalous instances. For an anomalous detection we can plot the outlier score against the dimension, and for comparison plot the 99<sup>th</sup> percentile of the dimension's scores. Wherever, the outlier score of the instance crosses or meets the percentile illustrates which dimensions contribute the most to outlieriness.

### 5.4 Constraints

This method is not without its limitations. It suffers from the curse of dimensionality (Anselin, 2019). The use of Euclidean distance to formalise the definition of Multivariate LISA, intrinsically, makes this a distance-based method performed within a context. This means that the method is a robust and powerful algorithm often associated with global outlier detection in a local framework, making it useful in detecting local outliers. Nevertheless, this reliance on distance makes this method sensitive to dimensionality (Aggarwal, 2016).

## 5.5 Summary

The chapter has elaborated on the the proposed algorithm, GBOD. This method is an adaptation of [Anselin \(2019\)](#) for the purposes of this research. Via the optimal selection of neighbours the method aims to contextually define anomalies. It looks to demonstrate interpretability advantages and data sparsity limitations. Its ease of use, hyperparameter-free and explainable nature make it increasingly attractive to utilise. Having described this in detail, it is necessary to assess how well it performs against the state-of-the-art. The following chapters look into this.

# Chapter 6

## Experimental Study

This chapter describes the set of experiments used to comprehensively assess the saliency of the data, the effectiveness of the detection schemes and the viability of this approach for similar studies.

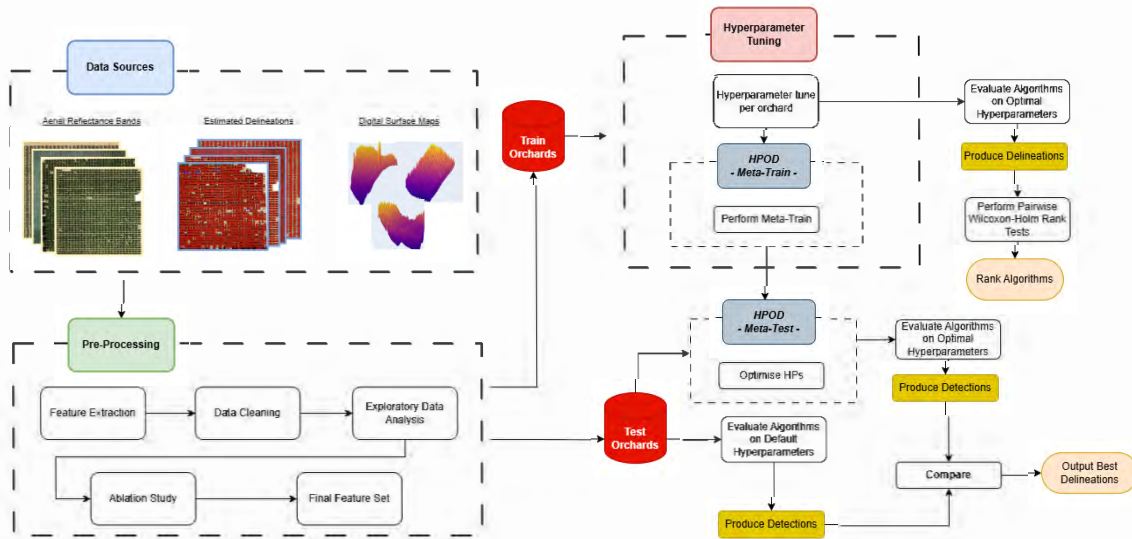


Figure 6.1: Architecture Diagram

Figure 6.1 describes the study design used to answer the research questions. Using the data sources, the first step in the pipeline was to extract the relevant features. Various pre-processing steps were conducted to clean the data and to find the most salient subset. Following data extraction for each orchard, the orchards were split into training sets, containing hand-made reference data and test sets, which are assumed to not have reference data — although the test sets do have reference data, it is best to assume that they do not to better understand how the methods perform out-of-sample (OOS). Thereafter, the models built on the train set were

hyperparameter tuned, and the results of this passed on to HPOD to train the meta-learner. Concurrently, the tuned models built on each orchard of the training set were evaluated to assess the performance amongst optimal models, enabling appropriate ranking. To observe whether these algorithms can be used OOS, the trained HPOD is used to optimise OOS hyperparameters. These optimal models thereafter produced detections and were compared to the results at default settings. This proposed pipeline is laid out in detail in the following subsections.

## 6.1 Dataset, Train/Test Split

Of the 316 orchards, 70 have been randomly selected as a training set. The training set has been summarised in Table 6.1, which illustrates the vast array of differing orchards. Multiple datasets are often used to benchmark methods in studies and the training set has been used for this. A subset of 70 orchards provides sufficient evidence to substantiate modelling effects (Li et al., 2020; Han et al., 2022; Xu et al., 2023). The training set is also used to obtain several optimal hyperparameter settings as a necessary step for HPOD to find hyperparameters for the test set. Without replacement, 31 orchards have also been randomly selected as a test set used to assess the generalisability and viability of automation in this study.

A heuristic defined in chapter 3 was used to distinguish inliers from outliers. However, discriminating between the different types has been done empirically for each orchard and demonstrated in Table 6.1. Table 6.1 shows the number of observations,  $N$ , the number of observed anomalies indicated by the reference data, Anomaly (ratio), and the types of anomalies. It illustrates the different categories as: false positive,  $F_p$ , over-segmentation,  $O - S$ , and under-segmentation,  $U - S$ . These distinctions are useful in attributing the performance in particular orchards by the methods and type of anomaly. Following the assumptions made in chapter 3, under-segmented and false-positive delineations are hypothesised to be easily captured by global methods, while local methods are expected to better capture over-segmentation.

Table 6.1: Data description of the 70 datasets included in the training split

| Data      | $N$  | # Anomaly (Ratio) | # $F_p$ | # $O - S$ | # $U - S$ |
|-----------|------|-------------------|---------|-----------|-----------|
| Orchard 1 | 1523 | 28 (1.84%)        | 6       | 22        | -         |
| Orchard 2 | 1148 | 2 (0.17%)         | -       | -         | 2         |
| Orchard 3 | 2049 | 49 (2.39%)        | 49      | -         | -         |
| Orchard 4 | 1281 | 30 (2.34%)        | 6       | 24        | -         |
| Orchard 5 | 1024 | 3 (0.29%)         | 3       | -         | -         |
| Orchard 6 | 862  | 6 (0.70%)         | 1       | 5         | -         |
| Orchard 7 | 312  | 2 (0.64%)         | -       | 2         | -         |
| Orchard 8 | 1487 | 4 (0.27%)         | 3       | -         | 1         |

|            |      |              |     |    |    |
|------------|------|--------------|-----|----|----|
| Orchard 9  | 1113 | 4 (0.36%)    | 4   | -  | -  |
| Orchard 10 | 1122 | 27 (2.42%)   | 23  | 5  | -  |
| Orchard 11 | 406  | 10 (2.46%)   | 9   | 1  | -  |
| Orchard 12 | 2234 | 26 (1.16%)   | 26  | -  | -  |
| Orchard 13 | 791  | 14 (1.77%)   | 14  | -  | -  |
| Orchard 14 | 5169 | 10 (0.19%)   | 10  | -  | -  |
| Orchard 15 | 1211 | 80 (6.61%)   | 78  | 2  | -  |
| Orchard 16 | 3785 | 40 (1.06%)   | 3   | 37 | -  |
| Orchard 17 | 915  | 16 (1.75%)   | 15  | 1  | -  |
| Orchard 18 | 1004 | 19 (1.89%)   | -   | 19 | -  |
| Orchard 19 | 1011 | 9 (0.89%)    | 7   | 2  | -  |
| Orchard 20 | 9931 | 102 (1.03%)  | 98  | 4  | -  |
| Orchard 21 | 1377 | 0 (0.0%)     | -   | -  | -  |
| Orchard 22 | 294  | 6 (2.04%)    | 2   | 3  | 1  |
| Orchard 23 | 2702 | 128 (4.74%)  | 122 | 6  | -  |
| Orchard 24 | 3366 | 122 (3.62%)  | 109 | 10 | 3  |
| Orchard 25 | 294  | 2 (0.68%)    | -   | -  | 2  |
| Orchard 26 | 1130 | 18 (1.59%)   | 17  | 1  | -  |
| Orchard 27 | 3641 | 381 (10.46%) | 380 | 1  | -  |
| Orchard 28 | 546  | 2 (0.37%)    | 2   | -  | -  |
| Orchard 29 | 380  | 12 (3.16%)   | 11  | 1  | -  |
| Orchard 30 | 647  | 5 (0.77%)    | -   | 5  | -  |
| Orchard 31 | 1078 | 30 (2.78%)   | 28  | 2  | -  |
| Orchard 32 | 652  | 17 (2.61%)   | -   | 17 | -  |
| Orchard 33 | 1231 | 23 (1.87%)   | -   | 21 | 2  |
| Orchard 34 | 1772 | 44 (2.48%)   | 41  | 3  | -  |
| Orchard 35 | 1293 | 33 (2.55%)   | 1   | 32 | -  |
| Orchard 36 | 2853 | 21 (0.74%)   | 1   | 1  | 19 |
| Orchard 37 | 2538 | 49 (1.93%)   | 11  | 38 | -  |
| Orchard 38 | 1566 | 11 (0.70%)   | 9   | 2  | -  |
| Orchard 39 | 3216 | 37 (1.15%)   | 34  | 1  | 2  |
| Orchard 40 | 981  | 1 (0.10%)    | -   | 1  | -  |
| Orchard 41 | 399  | 3 (0.75%)    | 1   | 1  | 1  |
| Orchard 42 | 1061 | 42 (3.96%)   | 42  | -  | -  |
| Orchard 43 | 475  | 5 (1.05%)    | 3   | 2  | -  |
| Orchard 44 | 733  | 50 (6.82%)   | 41  | 9  | -  |
| Orchard 45 | 1299 | 45 (3.46%)   | 40  | 5  | -  |
| Orchard 46 | 2666 | 33 (1.24%)   | -   | 33 | -  |
| Orchard 47 | 1723 | 31 (1.80%)   | 31  | -  | -  |
| Orchard 48 | 814  | 35 (4.30%)   | 23  | 12 | -  |
| Orchard 49 | 229  | 0 (0.0%)     | -   | -  | -  |

|            |      |            |    |    |    |
|------------|------|------------|----|----|----|
| Orchard 50 | 1476 | 2 (0.14%)  | 2  | -  | -  |
| Orchard 51 | 1610 | 52 (3.23%) | 52 | -  | -  |
| Orchard 52 | 1167 | 1 (0.08%)  | -  | 1  | -  |
| Orchard 53 | 238  | 9 (3.78%)  | 7  | 2  | 1  |
| Orchard 54 | 4691 | 70 (1.49%) | 56 | 3  | 11 |
| Orchard 55 | 2090 | 25 (1.20%) | 25 | -  | -  |
| Orchard 56 | 1094 | 30 (2.74%) | 30 | -  | -  |
| Orchard 57 | 1885 | 12 (0.64%) | 10 | 2  | -  |
| Orchard 58 | 6759 | 9 (0.13%)  | 9  | -  | -  |
| Orchard 59 | 1304 | 1 (0.08%)  | 1  | -  | -  |
| Orchard 60 | 1862 | 15 (0.81%) | 14 | 1  | -  |
| Orchard 61 | 825  | 13 (1.56%) | 12 | 1  | -  |
| Orchard 62 | 287  | 5 (1.74%)  | 2  | 2  | 1  |
| Orchard 63 | 1023 | 2 (0.20%)  | 1  | -  | 1  |
| Orchard 64 | 1047 | 31 (2.96%) | 30 | 1  | -  |
| Orchard 65 | 383  | 4 (1.04%)  | 2  | -  | 2  |
| Orchard 66 | 2221 | 47 (2.12%) | -  | 47 |    |
| Orchard 67 | 1433 | 21 (1.47%) | -  | 19 | 2  |
| Orchard 68 | 4435 | 77 (1.74%) | -  | 76 | 1  |
| Orchard 69 | 2316 | 71 (3.07%) | 59 | -  | 12 |
| Orchard 70 | 1254 | 0 (0.0%)   | -  | -  | -  |

## 6.2 Exploratory data analysis

This section is used to better understand the features, their relationships with one another, and their relevance in detecting poor delineations. The exploratory data analysis (EDA) was performed on the first five orchards of the training set. This has been done over multiple datasets as each orchard may have a different relationship with the feature set, where the first five vary enough to capture the principle ideas of the data and the typical orchard. The EDA is used to understand the datasets and guide hypothesis testing, modelling and feature selection. This small hand-picked sample size of varying orchards has been used to gain an understanding of what the data may expose. A small size has been selected due to time constraints.

The first step was understanding the bare orchards and estimated delineations. This is done to illustrate the scope of the problem. Thereafter, a univariate analysis explored each feature independently. Profile plots individually examined DSM, aerial reflectance bands and vegetative indices to assess the discrimination properties of these features. Following this, the distributional properties of each feature were captured. The univariate analysis has also been used to capture the presence of outliers for each feature, reinforcing several hypotheses about what the data captures. Thereafter, a bivariate analysis was conducted to assess the relationship between

features. This explores multicollinearity, varying densities and bivariate outlier relationships. Collinearity demonstrates if two or more variables are strongly linear dependent, illustrated by the correlation between variables. However, it is often difficult to infer causality in the presence of multicollinearity, as collinear variables represent noise. Therefore, it is not helpful to maintain all features with strong relationships, especially when the curse of dimensionality is in effect. Another aspect of bivariate analysis is the ability in assessing the presence of varying densities, this may highlight which orchards and variables can benefit from local outlier detection. The bivariate outlier relationships may also highlight noisy features.

With a clearer understanding of the set of candidate features, the potential exists to select a more salient feature set or identify missing features. Hence, dimensionality is reduced while potentially improving detection and computational complexity.

### 6.2.1 Ablation Study

An ablation study has been used as a final step in the EDA. The ablation study has been used to assess the saliency of the proposed feature set. The feature groupings — shape, texture, spectral, DSM and confidence — were individually removed from the dataset, and each subset underwent anomaly detection on the methods at default settings to examine the impacts of inputs. All things being equal, this assessment was used to capture noisy features. As a control experiment, anomaly detection was performed on the full feature set that was captured after the exploratory data analysis. Thereafter, several experiments on different combinations of the data were conducted. First, we remove the confidence variable, while keeping all other variables and perform detection. Next, we removed DSM, while keeping all other variables, and performed detection. This simply allows the exploration of the performance differential offered by these two features independently. The third experiment removed shape descriptors, the fourth removed spectral indices, the fifth removed texture and the final experiment removed Zernike moments. AUC and AP were collected for the experiments and summarised over the 70 orchards. With a better understanding of the data, it became possible to assess the performance of detection methods more effectively.

## 6.3 Baseline Algorithms

The proposed algorithm is compared to five baseline methods described in the literature: ABOD, ECOD, IForest, LOF and PCA. As depicted in [Figure 6.2](#), a technique from each relevant AD approach has been selected for a robust comparison. These methods can be found in [Zhao et al. \(2019\)](#). LOF and IF have been chosen as they are fundamental methods and have been used as one of the most benchmarked schemes to date ([Aggarwal, 2016](#)). ECOD is a state-of-the-art method used in this study, which tops a number of benchmarks in terms of speed and performance ([Li](#)

et al., 2022; Han et al., 2022). ABOD is a common distance-based method, as it attempts to mitigate the problems often seen by most schemes in its category (Kriegel et al., 2008; Li et al., 2022; Almardeny et al., 2022). Finally, PCA acts as a feasible, hyperparameter-free spectral-based method.

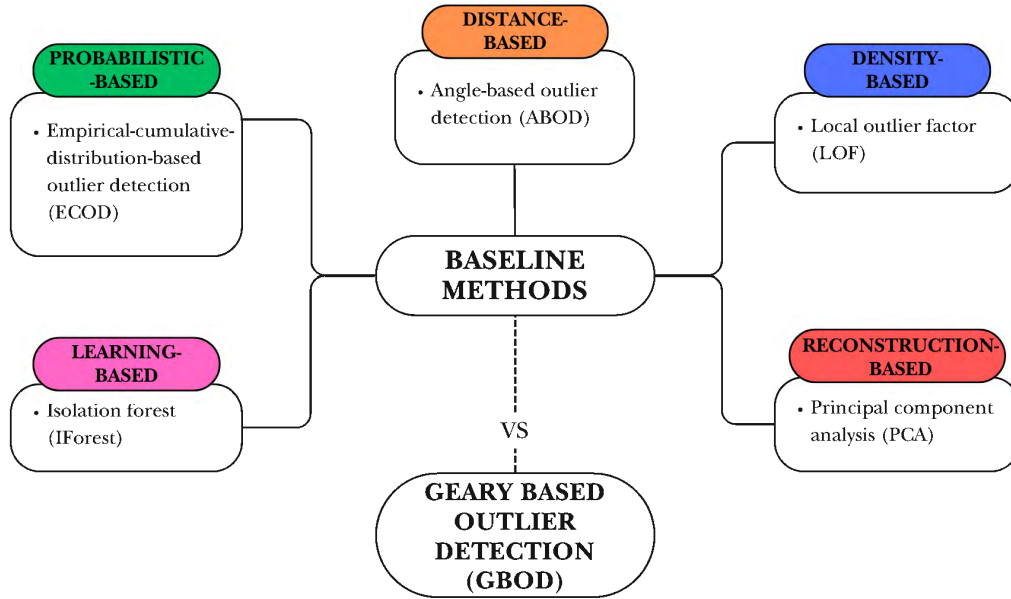


Figure 6.2: Comparison of baseline algorithms and the proposed methodology

Representative deep methods have been excluded from the analysis to allow a fair comparison between the proposed methodology and other shallow representative unsupervised algorithms. This choice has also been made to highlight the discriminative properties of the curated feature set and consider the possibility of unreliable references. Cluster-based methods have also been excluded from further study due to their hyperparameter-intensive nature. These methods are not generalisable or reusable due to the dependence on appropriate clusters for good detection. Similarly, manifold learning techniques are also hyperparameter intensive, making them complex to achieve generalisability and automation in this study.

## 6.4 Evaluation Metrics

The model performance was evaluated using the most popular performance measures in anomaly detection such as receiver operating characteristic (ROC), area under the curve (AUC-ROC), precision-recall (PR) curve, and average precision (AP). These measures are often utilised in tandem (Ruff et al., 2021; Davis and Goadrich, 2006). ROC curves are used in binary decision problems, demonstrating the relationship between correctly classified and incorrectly classified observations as a result of a

changing threshold. While considering, true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN), the ROC curve illustrates the change in the true positive rate,  $TPR = \frac{TP}{TP+FN}$  against the false positive rate,  $FPR = \frac{FP}{FP+TN}$ , dependent on the change in the threshold,  $\tau$ . The area under this visual is the accuracy measure AUC-ROC. PR curves are an alternative, specifically used for skewed class distributions, the common case in anomaly detection. They are used to highlight the performance of models on the anomalous classes, using precision and recall, where precision is specifically calculated in terms of the minority class. This curve depicts the change in Precision =  $\frac{TP}{TP+FP}$  against Recall =  $\frac{TP}{TP+FN}$ , across  $\tau$ . AP summarises the precision-recall curve as the weighted mean of precisions over all thresholds. This is denoted as  $AP = \sum_n (R_n - R_{n-1}) \times Pr_n$  where  $P_n$  and  $R_n$ , denotes precision and recall at the  $n^{th}$  threshold, respectively (Almardeny et al., 2022). This can otherwise be seen as the area under the PR curve (Davis and Goadrich, 2006; Ruff et al., 2021). These measures will be used to demonstrate a robust evaluation of detection performance accuracy.

## 6.5 Hyperparameter Tuning

Certain models — ABOD, IForest, LOF and PCA — require hyperparameter tuning to obtain their best performance. Tuning was performed mimicking the design used by Li et al. (2020), Han et al. (2022) and Almardeny et al. (2022) for validation analysis. These studies used a more inductive setting as opposed to the transductive nature of unsupervised learning. Where the inductive setting describes using training data to adapt a model and using this final hypothesis to classify data; in essence an inductive approach is analogous to a classification/supervised scheme (He and Yu, 2010; Rossi et al., 2018; Han et al., 2022). The transductive setting describes an approach that does not require the building of a final hypothesis that can predict test data. Instead, it describes the use of the learning algorithm to capture appropriate patterns in a dataset without classification (Vapnik, 2013; He and Yu, 2010). This is the ubiquitous manner in which unsupervised models are designed.

In each experiment, or this case orchard, 70% of the orchard data was used for training and 30% for validation. The training-validation split was done using stratified sampling to ensure that relative class frequencies were preserved (Han et al., 2022). The detection method was thereafter trained and validated over 5 independent trials for each configuration suggested in the search space. Over these trials the averaged AP was saved to capture the performance of the setting. The “Optuna” framework (Akiba et al., 2019) was utilised to explore the hyperparameter settings described below and select the best settings for each orchard.

For LOF the search space for the *number of neighbours* has been set between the range 10 and 150 at an interval of five, using distance metrics Euclidean, Manhattan, Chebyshev and Minkowski (Breunig et al., 2000). This range accounts for the presence of unwanted statistical fluctuation observed in *number of neighbours* < 10

while the upper bound considers the maximum number of outliers that can share the same density. For the more computationally complex ABOD, the search space of the *number of neighbours* was selected within the range 5 and 100 at an interval of five. PCA does not typically require tuning, however, incremental dimensions between two and the total *number of features* were explored for robustness. IForest used a search space for the *number of trees* in the range of 100 and 2000 at an interval of 100, and several dimensions between zero and one less than the total number of features were assessed for the maximum number of features, (*max features*). The *number of trees*, decides the number of isolation trees on which to train a forest, while *max features* indicates the number of features to use to train a tree.

As this work aims to achieve generalisability and automation, it is essential to appropriately select hyperparameters out-of-sample, or when the reference data is unknown. This is a significant point, as reference data is often unreliable or unavailable in anomaly detection studies. To this end, the research emulated hyperparameter optimisation for unsupervised outlier detection (HPOD) (Zhao and Akoglu, 2022) to satisfy this requirement. By utilising all the training orchards, a meta-learner is developed based on hyperparameter settings, internal performance measures, and meta-features captured within the search spaces of the methods. Thereafter, sequential model-based Bayesian optimisation is used to evaluate OOS hyperparameters from the meta-learner and the test sets' meta-features, proposed hyperparameters and subsequent internal performance measures. The full details of implementation can be found at Zhao and Akoglu (2022). Using the meta-learning approach suggested by Zhao and Akoglu (2022), hyperparameters were evaluated for datasets that lacked reference data. In the case of this study, this relates to the 31 randomly selected test datasets.

## 6.6 Model Comparison

The primary goal of this section was to assess the viability of the proposed algorithm against the optimised baseline methods in a transductive manner. This section also assessed the performance of the optimised models, in a fully unsupervised manner. Essentially, we are benchmarking the proposed algorithm against the baseline and state-of-the-art methods. Now that the best hyperparameters have been selected over the 70 orchards, it is possible to determine which model(s) are best suited for the task of anomaly detection for poor tree crown delineations. While using the optimal hyperparameter settings of each method and orchard, detection was carried out over all 70 orchards and identifications were produced. The reference data is thereafter compared to assess the performance of these models. This enables the realisation of ROC and PR curves and subsequently the calculation of AUC and AP measures for performance. Following this, the paired Wilcoxon-Holm signed rank test was used to compare the statistical significance between the performance of the methods across the multiple datasets (Demšar, 2006). Using this a critical difference diagram was

used to illustrate this statistic and rank the performance of the models from best to worst (Ismail Fawaz et al., 2019; Xu et al., 2023; Han et al., 2022). The CD diagram uses a thick horizontal line to group classifiers that are not significantly different together. Some orchards have a particular anomaly type in abundance. To this end, critical difference plots were created based on the subset of orchards dominated by over-segmentations and a subset dominated by both false positive delineations and under-segmentations. This was to examine whether over-segmentations are indeed local outliers, hence demonstrating better performance when coupled with local detection. And the hypothesis that false-positives and under-segmentations are global outliers, demonstrating better performance with global detection.

## 6.7 Generalisation Tests, Out-of-Sample

As a final test, HPOD’s meta-learning, which was trained on the training set, is used to optimise hyperparameters for the algorithms on the test set. The results of this were compared to evaluated algorithms at default settings, to observe the effect of HPOD on generating optimal models. Thereafter, the performance of the classifiers OOS was illustrated in localised areas. This visually depicts how well each model captured each type of anomaly, and where the classifiers struggle.

## 6.8 Environment Setup

The experiments of this work were conducted on a workstation with an Intel i9-13900H CPU, NVIDIA GeForce RTX 4070 Laptop GPU GDDR6 @ 8GB, and 32 GB RAM. The implementation of the code can be found at: <https://github.com/edwardbaleni/ITCD-AnomalyDetection.git>

## 6.9 Summary

This chapter has highlighted the experiments conducted to answer the research questions and produce a plausible solution for the aim of this work. Figure 6.1 demonstrates the process of the experiments. Given a train set, the EDA and ablation study are used to derive the most useful attributes from the proposed feature set; a necessary step in distinguishing the most salient set of features for repeatable studies. Following this, hyperparameter tuning is performed to compare GBOD against the optimised estimators. This enabled the training of HPOD and the evaluation of the algorithms at optimal hyperparameters. Evaluation of the algorithms produce results that are able to resolve whether GBOD is competitive and which algorithm is best suited for the task. Finally, using the trained HPOD and the test set, the meta-learner is used to enable OOS outlier detection. The following chapter is the first of these steps, the EDA. Thereafter, the remainder of the experimental study.

# Chapter 7

## Exploratory Results and Feature Reduction

### 7.1 Exploratory Data Analysis

This analysis was conducted over the first five orchards described in [Table 6.1](#). An illustration of each can be seen in [Figure A.1](#). This chapter will present Orchard four, [Figure 7.1](#), as an example. As seen in [Table 6.1](#) not all orchards are heavily populated with anomalies, similarly, not all anomaly types exist in each orchard. [Figure 7.1b](#) demonstrates an ideal case depicting several poor delineations including over-segmented and false-positive delineations.

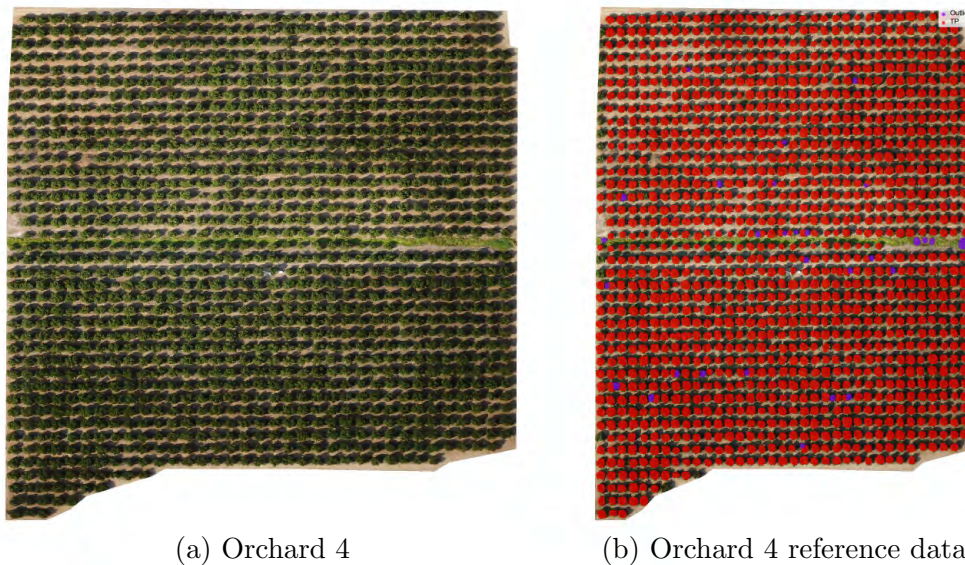
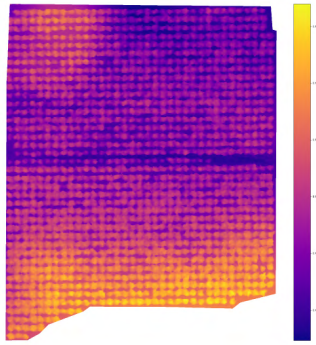


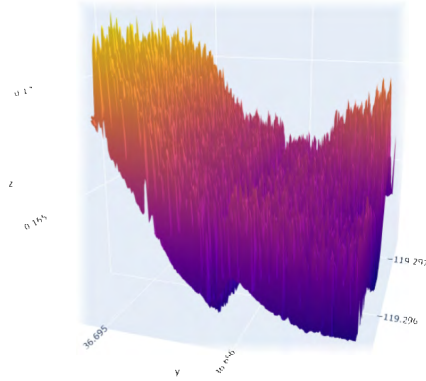
Figure 7.1: Subset of [Figure A.1](#), depicting the Orchard and reference data of Orchard four of the training data.

The variation in shadows cast is the first thing to notice over all the orchards in [Figure A.1](#). This difference between orchards demonstrates that the time and weather at which the photographs were taken are inconsistent. Orchard four — [Figure 7.1](#) — depicts varying illumination across the orchard, which may indicate that it was captured on a cloudy day. Orchards two, three and five — [Figure A.1](#) — also have shadows cast, which may indicate the photo being taken closer to dawn or dusk as opposed to noon or may indicate a cloudy day. Illumination in the orchard directly impacts the valuation of texture and aerial reflectance. Different times of day may indirectly affect the spectral absorption of aerial reflectance bands due to atmospheric changes resulting in attenuation. One can also see that no delineation embodies a perfect circle, although they do tend towards it. These imperfect circles all exhibit some level of varying boundary irregularities. This is considered when selecting the final feature set.

### 7.1.1 Profiles



(a) Orchard 4 2-D Representation



(b) Orchard 4 3-D Representation

Figure 7.2: Subset of [Figure A.2](#), depicting the Orchard DSM of Orchard four of the training data

[Figure 7.2](#) and [A.2](#) empirically demonstrate that DSM is discriminative, as it clearly distinguishes the tree crown from the orchard floor. However, since this is DSM and not canopy height model, the slope of the terrain on which the orchard exists is also apparent. This can be seen in the orchards depicted in [Figure A.2](#), and should be assumed to be the case for all orchards. The trend in the elevation data of each orchard can be linear, quadratic, or any higher-order polynomial. This inherently makes this feature difficult to handle despite its clear usefulness. Trees with peaks may be given a similar height to the orchard floor at other locations due to the slope. This would make this feature indiscriminate or even harmful for techniques that focus on global detection. Nevertheless, since local and spatial techniques incorporate local densities, this feature can still stand as a powerful discriminator between the orchard floor and tree crowns.

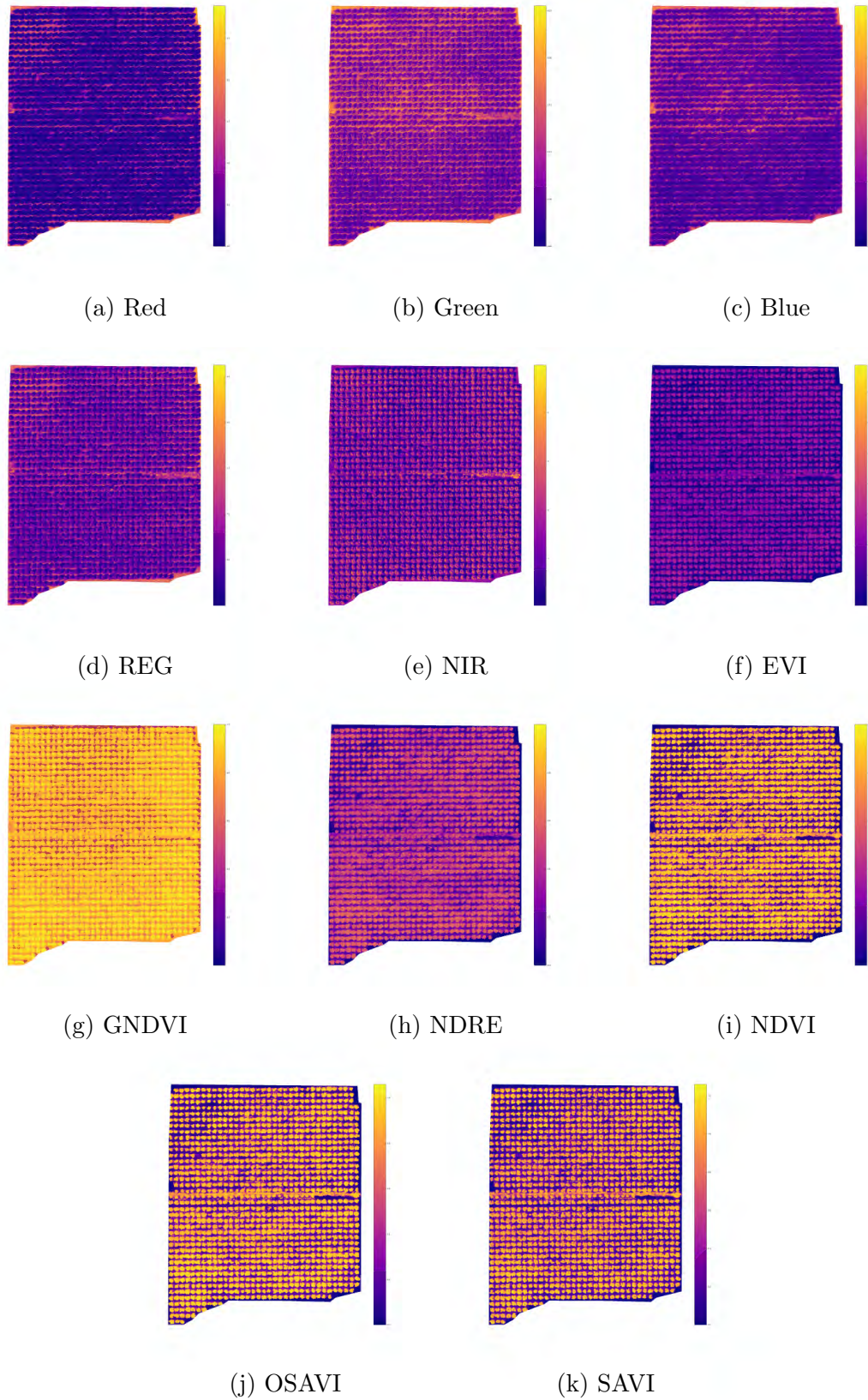


Figure 7.3: Subset of the profile plots presented in [Section A.3](#), depicting the aerial reflectance and vegetative indices of Orchard four of the training data

For example, [Figure 7.2](#) demonstrates much greater elevation at the bottom of the orchard and at the top left. However, there also exist peaks in between these areas. This is illustrated more clearly as a parabolic shape in [Figure 7.2b](#). This is different from the other orchards like [Figure A.2d](#) which demonstrate a linear trend. Therefore, it is not possible to describe some heuristic for detrending that could work for all cases. Because of the presence of trend in [Figure 7.2](#), it is expected that global detectors would identify trees in the trough of the orchard to have similar heights as the orchard floor in the greater elevation areas, making it difficult to distinguish soil and tree crowns appropriately. However, while using spatial methods, other trees in the same physical neighbourhood would be directly compared. While considering local densities, the small number of anomalies have significantly different relative densities, due to their other features, if they are considered neighbours to the large number of true positives in these regions. For an orchard with a trend, detection is problematic as we cannot compare a peak at the top of a slope with a peak at the bottom of the slope. However, using local and spatial anomaly detection methods enables this feature, as it will compare peaks on the slope with other peaks at similar levels; whereas while using global detection methods, DSM data may deteriorate results.

Alongside the DSM information, it is essential to understand the spectral profiles. The first of these to consider are the aerial reflectance bands over all five orchards, depicted in [Section A.3](#). For our example, these have been illustrated in [Figure 7.3](#). Visually, one can discern the different trees, but the difference in aerial reflectance value is not very contrasting or consistent. On the other hand, the vegetative indices (excluding GNDVI) give a much stronger contrast between soil and tree crowns. GNDVI as depicted in [Figure 7.3g](#), along with the other aerial reflectance bands are not useful indices to characterise tree crowns in the context of this study. Vegetative indices such as EVI, NDRE, NDVI, OSAVI and SAVI are much more expressive and discriminating of crowns and soil and are therefore considered further.

### 7.1.2 Univariate Analysis

The univariate analysis gives a low-level description of the data's discriminative properties and observes the presence of outliers. The skewness of the features in [Figure 7.4](#), [7.5](#), [7.6](#) and [7.7](#) — where the distribution is either heavily skewed left or right and most, if not all, the outliers exist to the right or left respectively — illustrates the binary nature of the selected attributes. It can also be seen that within each feature the shape of the distribution is consistent between orchards, indicating the features' usefulness in capturing the expected information across different types of orchards, albeit at different scales. Conf, as depicted in [Figure 7.4a](#), describes a relative confidence that the ITCD algorithm correctly classified the delineation. The left skew of its distribution indicates that there are some observations that the algorithm did not confidently classify. Circ and Com have very compressed distributions, with a large number of outliers that are spread far from the body of the

distribution. This observation encapsulates the sensitivity to boundary irregularities faced by these features. Figure 7.4a shows that the shape descriptors behave as theorised.

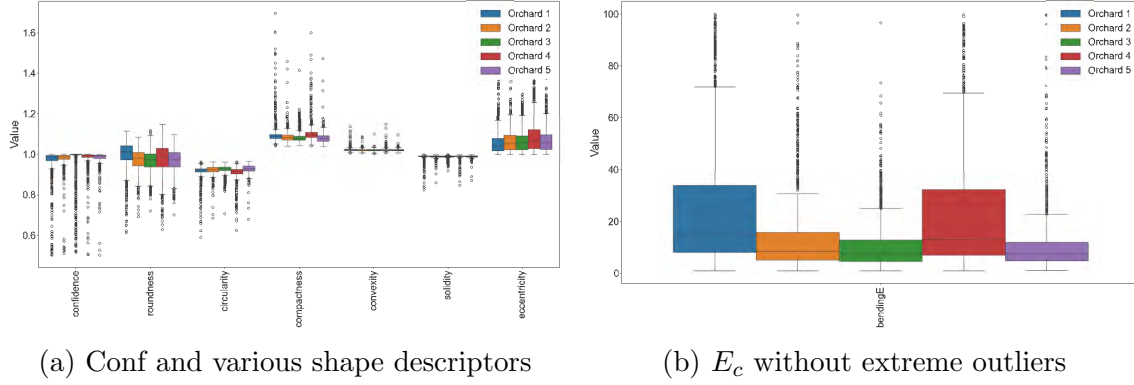


Figure 7.4: Boxplots illustrating the distribution of the shape descriptors and confidence of the first five orchards

Figure A.14 portrays that  $E_c$  has exceedingly large values, in the scale of  $\times 10^{37}$ , due to impressive outliers. While disregarding the most extreme outliers, Figure 7.4b shows that  $E_c$  does follow the expected distribution. The distribution is skewed to the right. This is the case as a value closer to 0 indicates that low energy is used to deform a shape, which is typically a circle or a rod. However, the median existing between 10 and 20 demonstrates that most of the shapes do deviate from a perfect circle, this median essentially portrays that all delineations have some level of boundary irregularities. The right tail of the distribution may present delineations that are excessively deformed.

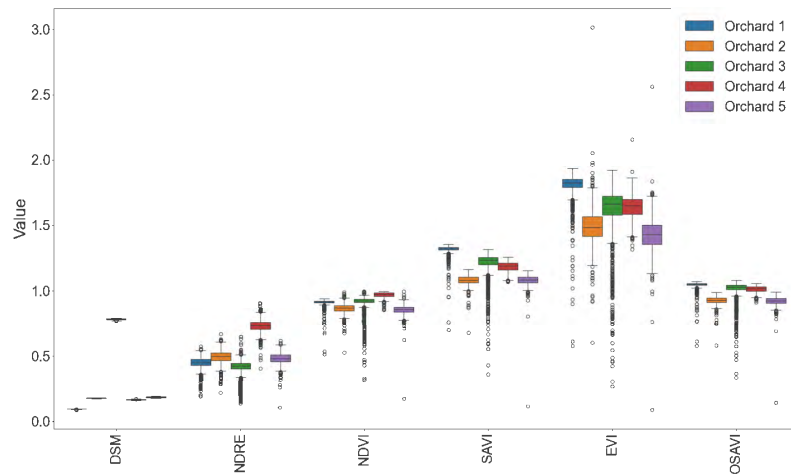


Figure 7.5: Boxplots illustrating the distribution of the spectral features of the first five datasets

The spectral features as depicted in [Figure 7.5](#) exist in a comparatively larger scale than the shape descriptors, except for DSM. They are all left skewed with relatively broad distributions demonstrated by their wider boxplots. This distribution can be directly compared to the profile plots in [Section A.3](#). The boxplots capture the maximum spectral value within the delineation for each vegetative index. The left whiskers directly relate to the soil pixels captured by the profile plots. This is the case for each of the vegetative indices, hence demonstrating that each is a decent discriminator. While soil pixels are expected to be observed on the left tail, right tail outliers could represent trees at different growing stages, or a different species of tree that has more chlorophyll content. The latter of which should be considered an outlier.

In the case of spectral features large variation is expected to indicate a great difference between soil and canopy pixels, this should be the same for texture features. However, the same assumption cannot be made for shape descriptors and DSM. In general, the shape of the crowns are almost circular, so the trees that pertain to this assumption should fit well within the distribution. It is important to note that some indices are sensitive to the boundary irregularities that these circular estimates have. Therefore, it is possible that large variations may be highlighting negligible differences.

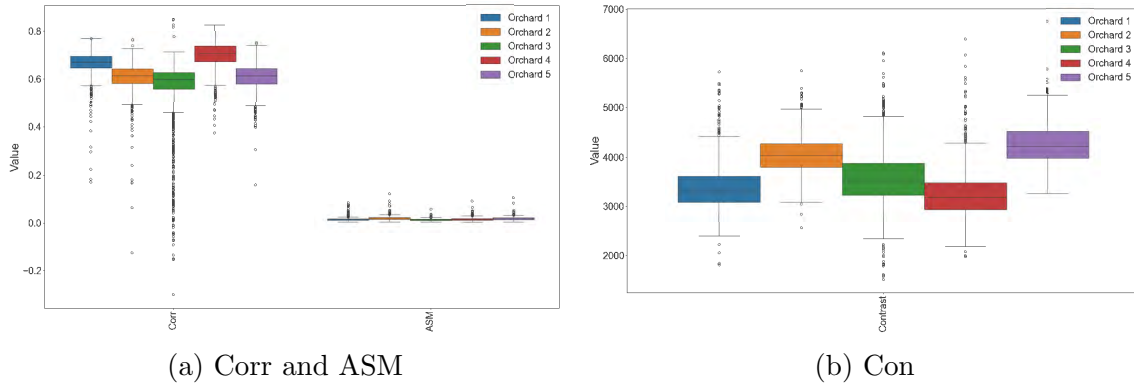


Figure 7.6: Boxplots illustrating the distribution of texture descriptors of the first five datasets

[Figure 7.6a](#) demonstrates that both Corr and ASM behave as expected, where the distribution of Corr and ASM are left and right skewed respectively. In the case of Corr the left tail indicates outliers, while the main body indicates true tree crowns. The right tail for ASM indicates outliers, while the main body indicates tree crowns. Con, [Figure 7.6b](#), is a complex attribute. Based on the notion that soil pixels have smoother textures than tree crowns, it is likely that the left tail represents soil pixels, while the remainder of the boxplot represents the desired crowns, a mixture of soil and grass textures, or even a mixture of tree crowns and soil. It is difficult to specifically attribute outlieriness here, so it is useful to take a deeper look at the what the

outliers of the texture feature boxplots are depicting. Con has left tail and right tail outliers. The left tail outliers of the Con distribution are illustrated in Figure A.16. This demonstrated that these outliers are uninformative to this study. Which goes against the assumption that low Con should equate to a soil surface. Figure A.16c is the only orchard where some of the false positive delineations actually relate to lower Con. Figure A.17 relates to the right tail outliers. These identify a large portion of over-segmented delineations, as these delineations have a combination of soil and tree crown pixels within the delineation, which results in a much greater Con valuation than if the delineation only encompassed either tree or soil. However, near-matched delineations have weakly defined boundaries, therefore, it should be allowed for true positive tree delineations to have a variation of tree crown and soil pixels. This is also the case for Figure A.17c and A.17e. At this stage, Con has been shown to be a weak discriminator for poor delineations in the context of this work. Similar to the right tail outliers of Con, Figure A.15 demonstrates that ASM is not distinguishing between false positives and true positives, it instead highlights over-segmented delineations. However, unlike Con, this feature rarely identifies true positives as outliers. Figure A.18 shows that Corr acts as expected, outlying correlation values identify between true positive delineations and false positives. However, this feature does also identify weaker trees that are in an earlier stage of the growth period, this is due to weak delineations capturing a lot of soil background and tree canopy.

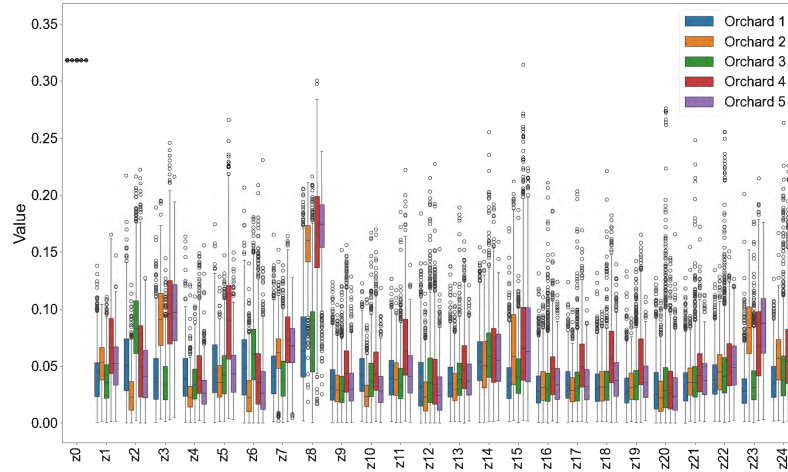


Figure 7.7: Boxplots illustrating the distribution of the first 25 Zernike moments of the first five datasets

Zernike moments are the final group of features as depicted in Figure 7.7. In particular,  $Z_{00}$  is not a useful feature for this study. It is a measure of centrality, which in the case of Zernike moments of our blobs would be near invariable since location is not considered in the measure. It has zero variation and no distribution to show. It therefore is uninformative and should be removed from the remainder of the study.

### 7.1.3 Bivariate Analysis

The off diagonals of Figure 7.8, 7.9 and 7.10 shows the kernel density plots, which give an indication of where the highest density of the data exists between two features, while the scatter plot gives an illustration of the bivariate relationship. Among the shape descriptors in Figure 7.8, Circ and Com have an almost perfectly linear relationship. Circ and Conv, Rou and Ecc, Com and Sol, and Conv and Sol have a strong negative relationships between the two. Conv and Com, and Circ and Sol both have strong positive correlations. These are all instances of potential multicollinearities that could prove problematic to the modelling of the data. These features seem to cluster heavily in one area, with a number of outliers present away from this clustering, much like the univariate case. While observing Conv, Sol, and Ecc, we see the presence of varying densities. This is better illustrated by observing their density plots, which is especially evident in orchard three. These varying densities may represent the existence of local densities and the need for local detection in the presence of such features.

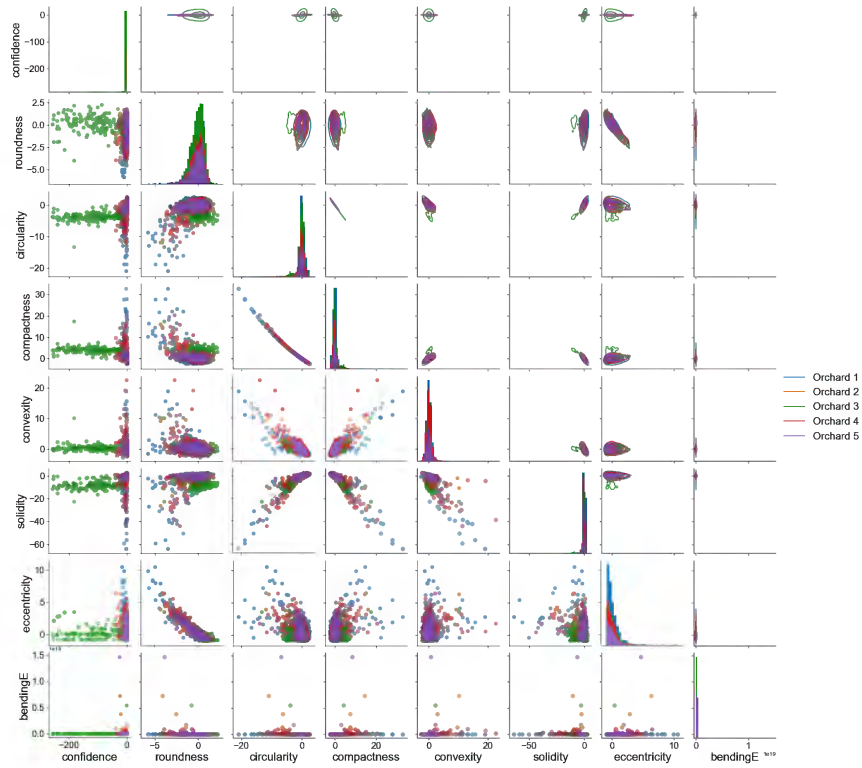


Figure 7.8: Plot of pair-wise relationships between shape descriptors of the first five randomly selected training datasets.

While looking at DSM — Figure 7.9 — there are no obvious outliers present in its

dimension, but there are obvious outliers in the dimension of the spectral indices being compared to DSM. There are no strong relationships present between DSM and the other spectral features, which is desirable as this will not cause multicollinearity issues. The same cannot be said for the vegetative indices. There are strong relationships present amongst all of the indices. After DSM, NDRE looks to have the weakest relationship to all other features present. There exists a correlation between NDRE and the other spectral indices, albeit weak. OSAVI, SAVI and NDVI exhibit a very strong linear relationship with one another over orchards one, three and five; while orchard two and four, demonstrate a moderate to strong correlation. On the other hand, EVI has a slightly more parabolic relationship with OSAVI, SAVI and NDVI. However, it still maintains a strong linear correlation with these features in orchards one, three and five; with moderate to strong correlations in orchards two and four. There exist multiple densities while observing DSM and NDRE with all other variables in the density plots for orchard three. This again brings light to the necessity for local outlier detection schemes.

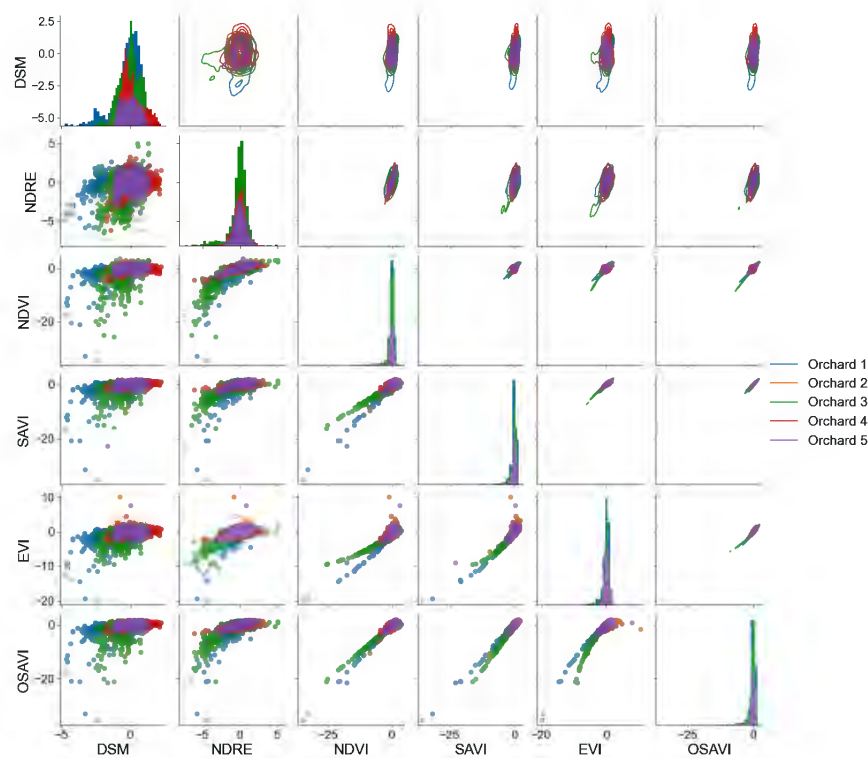


Figure 7.9: Plot of pair-wise relationships between spectral features and DSM of the first five randomly selected training datasets

The texture features depicted in [Figure 7.10](#) have no obvious relationships. Con, is a very peculiar feature. It clusters centrally, and does not express any obvious outliers when looking at the scatter plot. However, when observing the density plots

it seems to demonstrate varying densities for orchard 4. This is demonstrated by a much smaller, potentially sparse, cluster that forms on the right tail of Con. Again it is difficult to give this observation meaning. ASM does demonstrate some obvious outliers. This feature mostly clusters in one area. Corr seems to be the most defining and informative property for this study. It is very expressive in defining outliers in its dimension. There is also the presence of smaller varying densities when observing Corr. These smaller densities could be a grouping of outliers, or may demand local outlier detection in such a region.

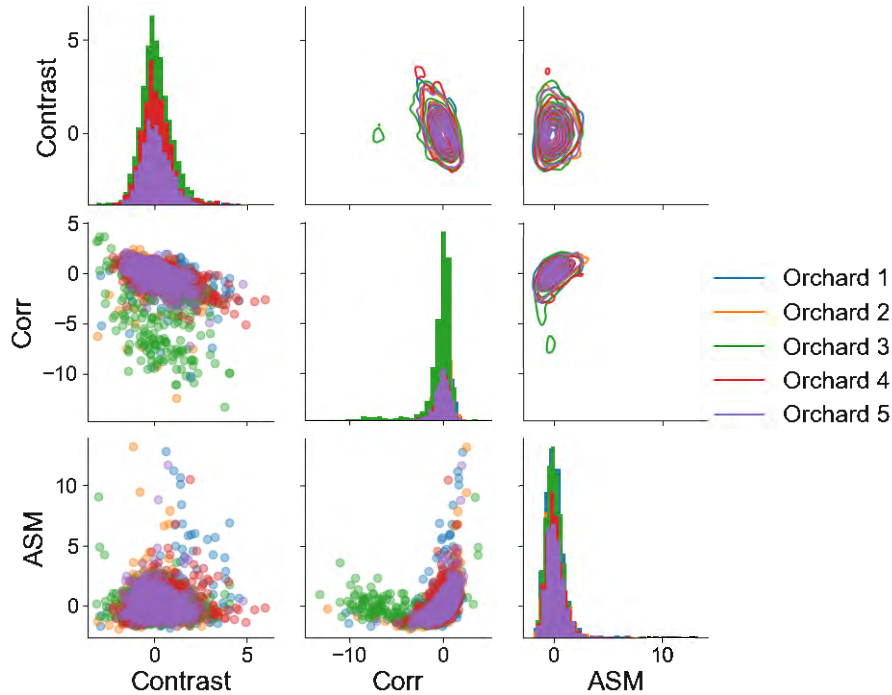


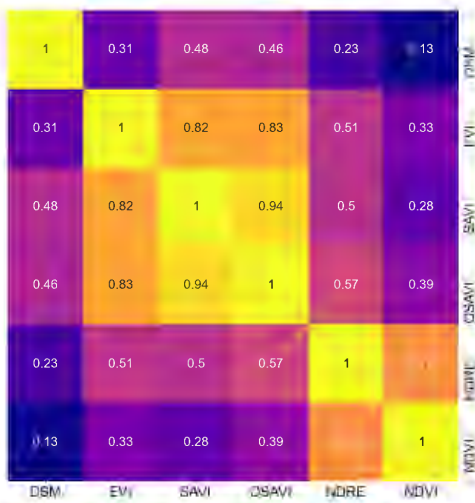
Figure 7.10: Plot of pair-wise relationships between texture descriptors of the first five randomly selected training datasets.

Figure A.20 illustrates the presence of strong correlations between all of the vegetative indices. Orchard four — Figure 7.11a — depicts only strong correlations between SAVI and OSAVI, which represent collinearity. It also demonstrates moderate to strong correlations between OSAVI, EVI and SAVI with one another and between NDVI with NDRE. The remaining vegetative indices maintain a moderate correlation with one another.

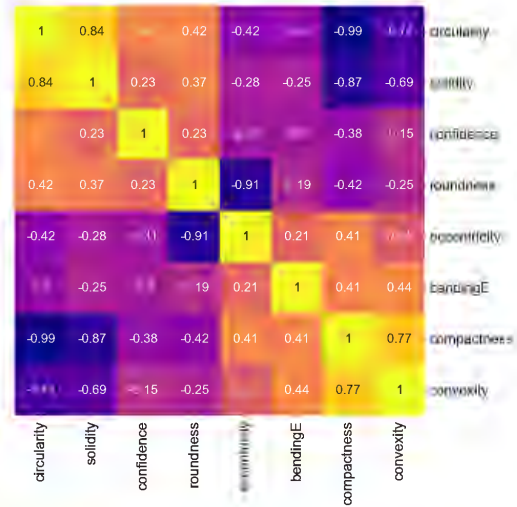
Similarly, Figure 7.11b agrees with 7.8. Comp has a strong negative relationship with Circ and Sol, and a strong positive relationship with Conv. Conv has a strong negative relationship with Circ and Sol. Ecc has a strong negative correlation with Rou and Sol has a strong positive correlation with Circ. Although these observations are common in Figure A.21, A.21c shows that some of these relationships are not as

profound. For example, Conv does not have a strong correlation with Sol and has a weaker correlation with Circ, while Com also has a weaker correlation with Conv. It is crucial to keep this in mind when considering the relevance of these features.  $E_c$ , unlike the other shape descriptors, does not have any relationships, which is ideal for the problem at hand.

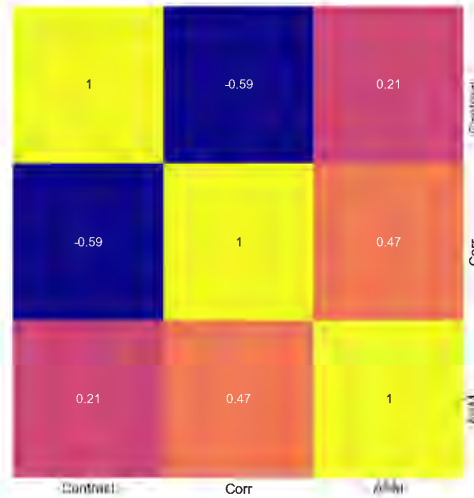
Finally, in the case of [Figure 7.11c](#), the texture descriptors have weak correlations. The strongest of which exists as moderate correlation between Con and Corr.



(a) Spectral Features



(b) Shape Descriptors



(c) Texture Features

Figure 7.11: Subset of the clustered heatmaps presented in [Appendix A](#), depicting the correlations within feature groupings of Orchard 4 of the training data.

## 7.2 Initial Feature Reduction

As a number of methods suffer from the curse of dimensionality, it is essential that the feature set captures the most salient set of variables. Many useless and collinear features have been identified in the EDA and need to be removed from the dataset.

As justified by the profile plots, the aerial reflectance bands and GNDVI are not worth further consideration because they are less discriminative than vegetative indices. OSAVI is an optimised version of SAVI that works in a variety of soil conditions, making it a versatile feature over many orchards. OSAVI and SAVI were the only spectral indices that demonstrated collinearity over all orchards, and as a result SAVI will be removed from further consideration. The other features had variable relationships dependent on the orchard, therefore disabling the conclusion of multicollinearities.

Con has been complex to attribute, and therefore cannot be included in this study. This complexity is damaging to the generalisability of this work. Although Con is adept at identifying false positives, it is difficult to understand what the left-tail outliers do to the modelling. ASM and Corr are kept for the remainder of the research as they provide great insight into textural qualities distinguishing over-segmented delineations and false positives from true positive delineations.

The shape descriptors demonstrated multicollinearities. The strongest collinearity exists between Com and Circ, both of which are sensitive to boundary irregularities. As both convey similar information, it is worth removing one of them to reduce the effects of collinearity and dimensionality. Circ is the more redundant option of the two because it conveys the same information as Rou. Com on the other hand seems more generalisable by definition. It is not reliant specifically on the circular description of the shape but also works well in cases of other shapes. Both Rou and Ecc are measures sensitive to elongation. Rou is a multifaceted index as it not only describes a measure of circularity but is also sensitive to the elongation of the shape (Zdilla et al., 2016). It also has a very strong negative correlation with Ecc. Rou can perform the duties of both Ecc and Circ and is therefore used over Ecc. This descriptor should be adept at identifying over- and under-segmented delineations.

Finally, the first Zernike moment,  $Z_{00}$ , has zero variation, making it uninformative in this study. It should therefore be removed for this reason.

## 7.3 Ablation Study

A further ablation study was done in the exploration phase to assess each feature grouping's quality. The feature groupings — shape, texture, spectral, DSM and confidence — were individually removed from the dataset, and each subset underwent AD on the methods at default settings to examine the impacts of inputs. All things being equal, this assessment was used to capture noisy features. AUC and AP were

collected for the experiments and summarised over the 70 orchards.

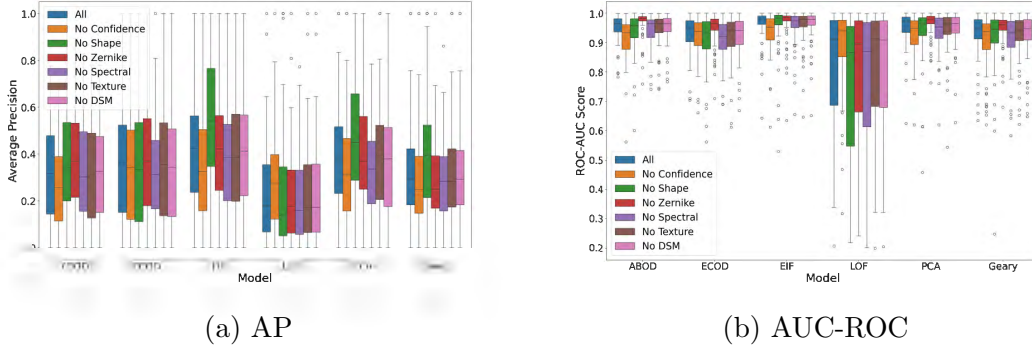


Figure 7.12: Summary of ablation study

Figure 7.12 depicts the results of the ablation study. The results showed that except for LOF, removing shape descriptors led to improved detection. Given the expectation that shape descriptors capture local outliers, coupled with the imbalance of orchards dominated by hypothesised global outliers, global detection methods are expected to corroborate this finding. Nevertheless, shape descriptors were retained to account for local methods. Removing Zernike moments had a net positive effect over the baselines and a minor negative effect for GBOD. Zernike moments drastically increase dimensionality and complicate interpretability and were therefore removed from the final feature set. The final feature set is captured in Table 7.1

Table 7.1: Final feature set

| Description        | Variable                               | Abbreviation |
|--------------------|----------------------------------------|--------------|
| Shape Descriptors  | Compactness                            | Com          |
|                    | Solidity                               | Sol          |
|                    | Convexity                              | Con          |
|                    | Bending Energy                         | $E_c$        |
|                    | Roundness                              | Rou          |
| Texture Features   | Angular Second Moment                  | ASM          |
|                    | Correlation                            | Corr         |
| Vegetation Indices | Normalized Difference Vegetation Index | NDVI         |
|                    | Normalized Difference Red-Edge Index   | NDRE         |
|                    | Optimised SAVI                         | OSAVI        |
|                    | Enhanced Vegetation Index              | EVI          |
| Other Indices      | Confidence                             | Conf         |
|                    | Digital Surface Model                  | DSM          |

## 7.4 Summary

The EDA explored a number of visualisations to help better understand the features. The profile plots demonstrated that aerial reflectance bands and GNDVI were lacking in discriminative power in comparison to the other spectral features. The univariate boxplots highlighted that  $Z_{00}$  was uninformative because it had zero variance. It also showed that Con was relatively unproductive towards the purpose of this research. The bivariate analysis depicted many existent collinearities removing the necessity of Circ and Ecc from the study. The ablation study was helpful in showing the utility of each feature grouping, finding that Zernike moments were more harmful than assumed, in most cases. The combination of the experiments eliminated 34 noisy features, greatly reducing the dimensionality of the data. With this completed, the experimental results were captured, as depicted in the next chapter.

# Chapter 8

## Experimental Results

This chapter assesses the results of the model evaluation performed with the optimal algorithms — methods with optimally tuned hyperparameters. This enabled comparisons between the proposed GBOD with baseline algorithms and the evaluation of model performance under different types of anomalies. Thereafter, this chapter examines generalisability and the viability of HPOD to automate anomaly detection for ITCD algorithms OOS.

### 8.1 Model Evaluation

Figure 8.1 and 8.2 illustrate a subset of representative orchards that demonstrate classifier performance. As seen in Table 6.1, this representativity is demonstrated by the orchards varying number of sizes, false positives, over-segmentations and under-segmentations. An EDA would be necessary to fully ascertain the representativeness of these orchards, however, this lengthy task would have delayed the execution of this study.

Figure 8.1 and 8.2 demonstrate that the classifiers, in general, perform better than the baseline — they are better than random guessing. However, LOF is shown to be a weak classifier for orchards three and 20. Figure 8.2, supports this as the classifier does not perform better than the baseline. Among the sample orchards, Figure 8.1 shows that the optimal threshold is high, which agrees with the small contamination levels present in Table 6.1. However, in some instances, such as orchards two, five, six and eight, the optimal threshold appears much lower. This demonstrates a preference for lenient thresholding when detecting poor delineations. In Figure 8.2, the classifiers of orchards two, five, six and eight dramatically decrease to the baseline, the AP valuation of the classifiers are also close to the contamination value; indicating underperforming classifiers for these scenarios.

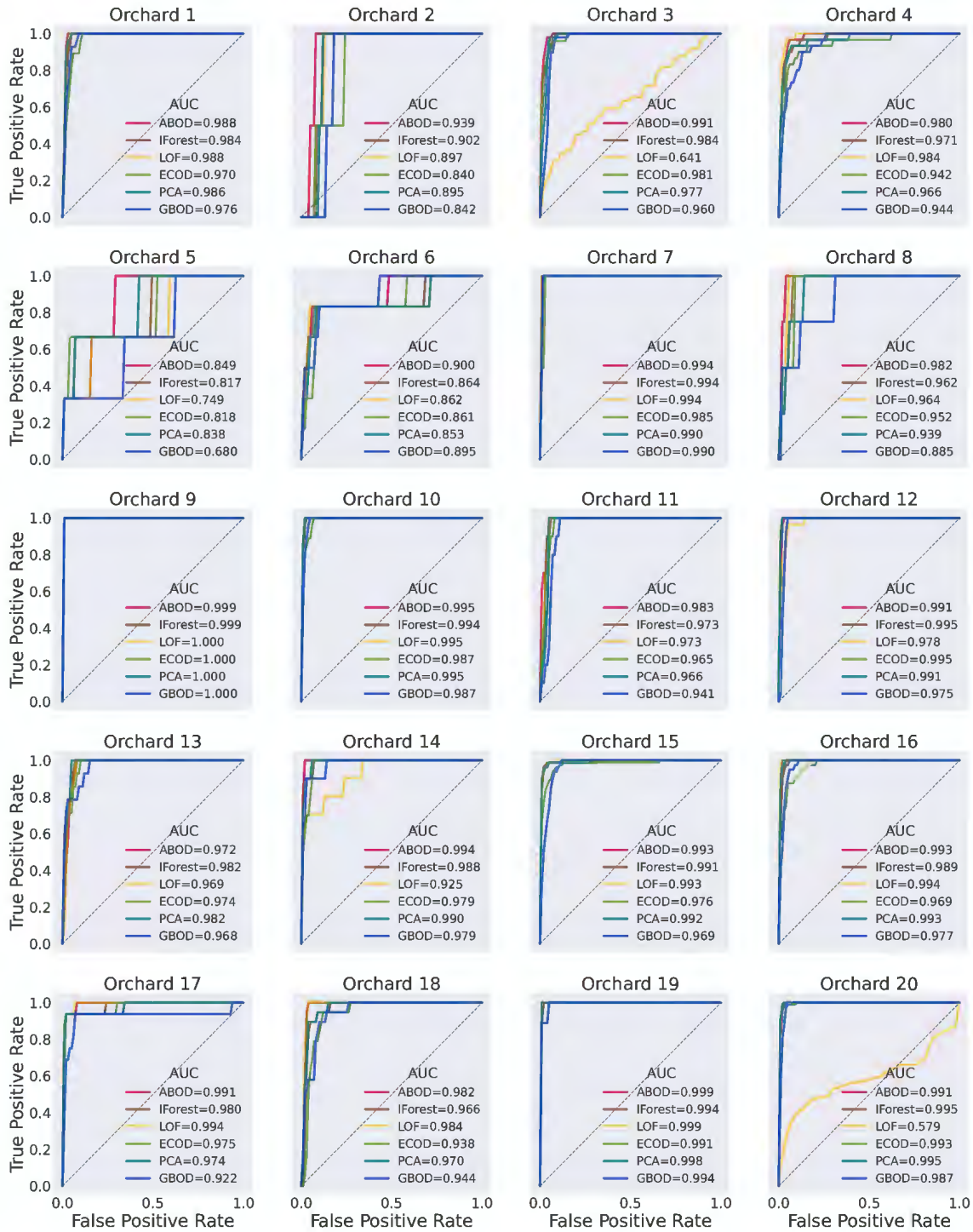


Figure 8.1: Receiver operating characteristic curve of the detection algorithms over the first 20 orchards

The AP measures observed in Figure 8.2 are relevant despite their low valuation. This has been demonstrated by the classifiers performing above the baseline.

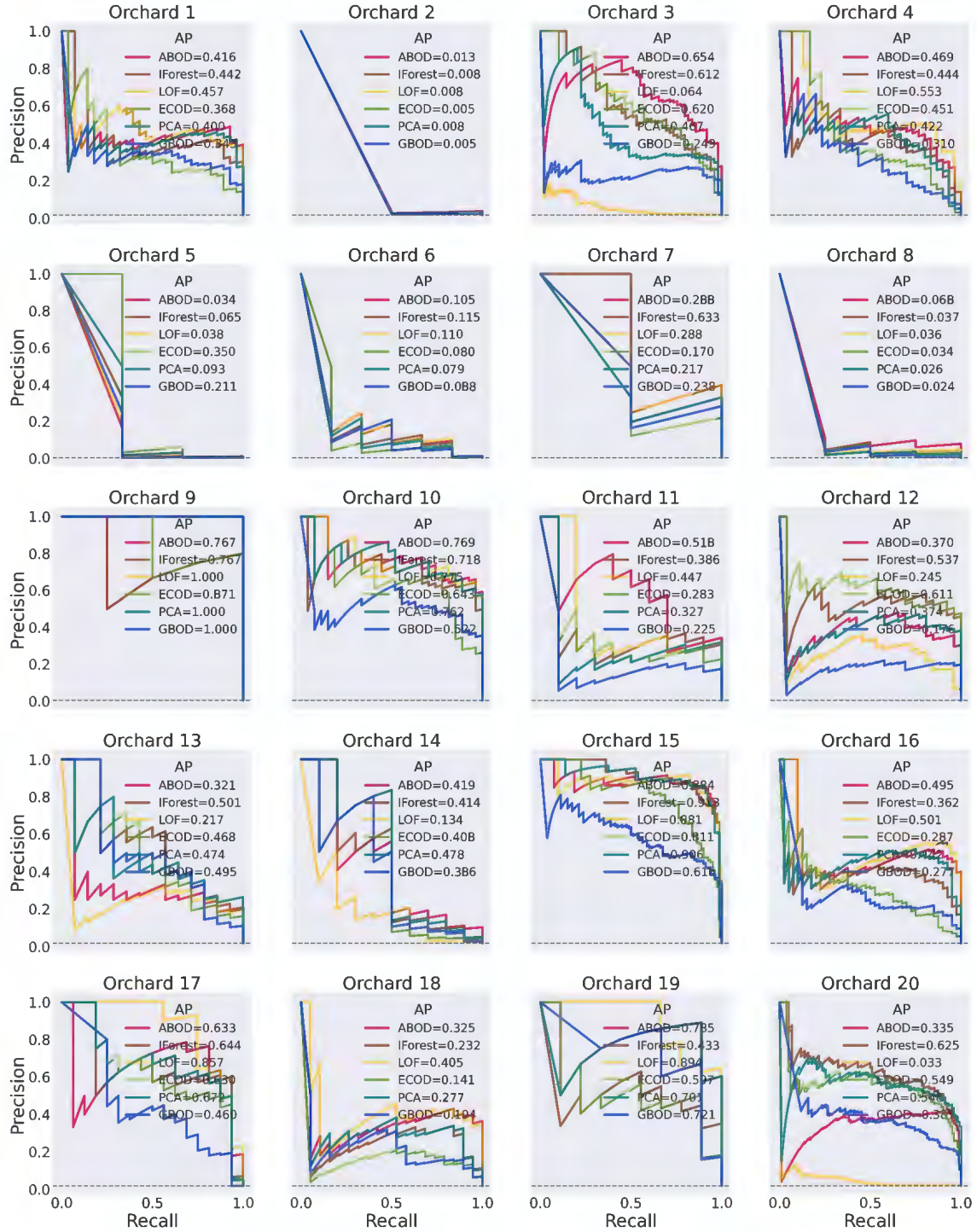


Figure 8.2: Precision-Recall curve of the detection algorithms over the first 20 orchards

Low AP values are also common in anomaly detection benchmarking papers with multiple datasets (Han et al., 2022; Xu et al., 2023; Almardeny et al., 2022). Performance measures like AUC-ROC and AP are helpful for model comparison.

### 8.1.1 Model Comparison

AP and AUC-ROC summarise PR and ROC curves, which enables the comparison of different classifiers. The summary statistics of the benchmarked performance are shown in [Table 8.1](#). Thereafter, paired Wilcoxon-Holm signed rank tests were used to examine statistical significance in algorithm performance ranking, which are summarised in critical difference diagrams (CD-diagram) that visually illustrate ranking ([Demšar, 2006](#)).

Table 8.1: Average and standard deviation values of different estimators’ AP and AUC-ROC measures over 70 orchards. The *p-value* depicts the statistical significance of the difference between GBOD and the baselines.

| Metric  |                  | ABOD                  | IForest               | LOF       | ECOD                  | PCA                    | GBOD      |
|---------|------------------|-----------------------|-----------------------|-----------|-----------------------|------------------------|-----------|
| AP      | <i>Avg.</i>      | <b>0.422 (3)</b>      | <b>0.437 (1)</b>      | 0.365 (5) | 0.397 (4)             | <b>0.423 (2)</b>       | 0.302 (6) |
|         | <i>Std. Dev.</i> | 0.230                 | 0.250                 | 0.249     | 0.241                 | 0.248                  | 0.190     |
|         | <i>p-val.</i>    | $7.39 \times 10^{-8}$ | $7.01 \times 10^{-9}$ | 0.005     | $3.97 \times 10^{-6}$ | $3.06 \times 10^{-11}$ | -         |
| AUC-ROC | <i>Avg.</i>      | <b>0.970 (1)</b>      | <b>0.964 (2)</b>      | 0.919 (6) | 0.946 (4)             | <b>0.962 (3)</b>       | 0.943 (5) |
|         | <i>Std. Dev.</i> | 0.042                 | 0.043                 | 0.130     | 0.057                 | 0.046                  | 0.058     |
|         | <i>p-val.</i>    | $2.12 \times 10^{-9}$ | $1.15 \times 10^{-9}$ | 0.036     | 0.395                 | $5.46 \times 10^{-8}$  | -         |

[Table 8.1](#) shows that the top three models are shared by global detection methods: ABOD, IForest and PCA. It also shows that GBOD, ECOD and LOF underperform, of which GBOD and LOF are local detection methods. At a confidence level of 99% and while considering AP, GBOD was the weakest performing model with statistical significance. However, in the case of AUC-ROC, there is not enough evidence to suggest that GBOD is statistically different in performance to both ECOD and LOF. The proposed methodology consistently underperforms in detecting poor delineations. This can be further seen in [Table B.1](#) and [B.2](#), where on average, over 70 orchards, both LOF and GBOD depict the worst performance. [Figure 8.3a](#) corroborates these findings, and demonstrates that IForest, ABOD, PCA and LOF are the best models, respectively. On the other hand, [Figure 8.3b](#) demonstrates that ABOD is the top performer with statistical significance. The next few models — IForest, PCA, and LOF — rank second but are statistically indistinguishable from one another.

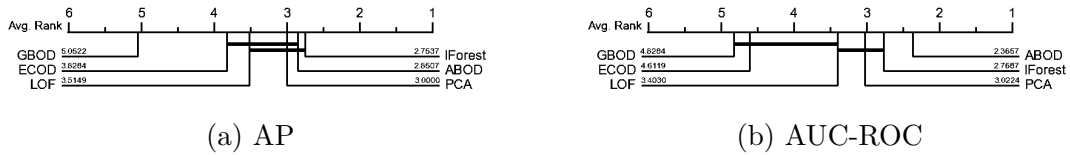


Figure 8.3: Critical difference diagram of AUC demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithms

It appears that the models with trained hyperparameters consistently performed better than hyperparameter-free models. It is also apparent that global outlier detection methods are preferable to local outlier detection methods for the detection of poor delineations.

### 8.1.2 Model Selection

Table 6.1 shows that certain orchards have particular types of anomalies in abundance, while other orchards contain a mixture of anomalies. This section identifies which models are best under particular settings. To this end, critical difference diagrams were created based on the subset of orchards dominated by over-segmentations, and of orchards dominated by both under-segmentations and false positives.

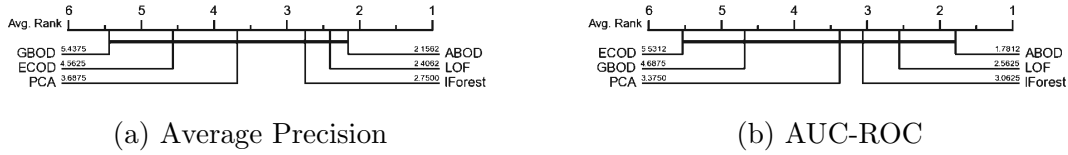


Figure 8.4: Critical difference diagram demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithm on a subset of the data dominated by over-segmentations

While only considering the subset of orchards that are dominated by over-segmented anomalies, a critical difference plot based off the performance measures AP and AUC-ROC were created — Figure 8.4. Due to there only being 16 orchards that are dominated by over-segmented anomalies, there is not enough evidence to provide statistical significance for the difference in rank performance in Figure 8.4. However, the implications of the ranks are still of interest. The average rank of LOF is significantly higher than in Figure 8.3. Although, ABOD is still ranked higher in this scenario, this observation demonstrates that LOF is potentially better at capturing over-segmentations than the other types. GBOD also looks to move up in rank in Figure 8.4b also demonstrating its preference for identifying over-segmentations.



Figure 8.5: Critical difference diagram demonstrating pairwise statistical difference comparisons amongst the proposed algorithm and the baseline algorithm on a subset of the data dominated by under-segmentations and false-postives

On the other hand, Figure 8.5a demonstrates that global outlier detector methods are superior at detection of false-positives and under-segmentations. This is demonstrated by the occupancy of global methods, PCA, IForest, ABOD and ECOD, of the

top four ranks, while the two local detectors, GBOD and LOF occupy the bottom rank. Figure 8.5b mostly agrees with Figure 8.3a and 8.3b in terms of the ranking of classifiers.

In terms of data, these findings may confirm that over-segmentation is a type of local outlier. Similarly, that under-segmentation and false-positives are a type of global outlier.

### 8.1.3 Runtime Efficiency

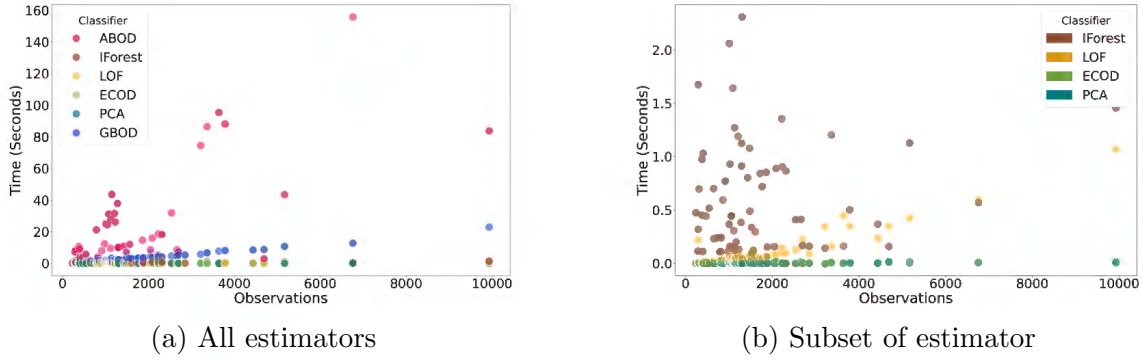


Figure 8.6: Depicting change in time vs increase in number of observations

The runtime of each model was captured, as depicted in Figure 8.6. ABOD is the most demanding in terms of time complexity. The sporadic time over the observations for ABOD can be explained by the change in number of neighbours for different orchards; as the *number of neighbours* greatly affect ABODs speed. ABOD is computationally inefficient for this problem at high-dimensions. Second to this is GBOD which demonstrates a sufficient time complexity for comparison in this study. It is computationally less complex than one of the best performing models in this study. Its increase is also linear, dependent on the number of observations.

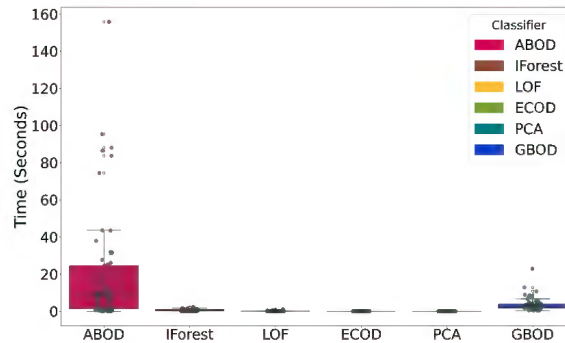


Figure 8.7: Summary of time complexity

Figure 8.7 gives a more general understanding of the time complexity expected. Figure 8.7 shows that PCA, ECOD, LOF and IForest are the fastest methods. Followed by GBOD and the slowest of all being ABOD. Much like with ABOD, IForest's

time is seemingly dependent on its hyperparameters as depicted in Figure 8.6b; the greater the number of trees used to train the forest, the slower the model evaluates.

## 8.2 Generalisation

Figure 8.8 shows that HPOD is preferable to default hyperparameters, especially in the case of LOF. The difference in distribution between default LOF and HPOD’s LOF, shows significant performance improvement. A slight performance increase is seen for ABOD and IForest. The distribution of AP and AUC scores is centered more leftwards for the default models than it is for the models that have been optimised using HPOD. Although not a significant difference between the two, the improvement in detection is preferable. There exists no added benefit to tuning PCA. GBOD is more competitive when compared to the default baseline methods, outperforming models like LOF.

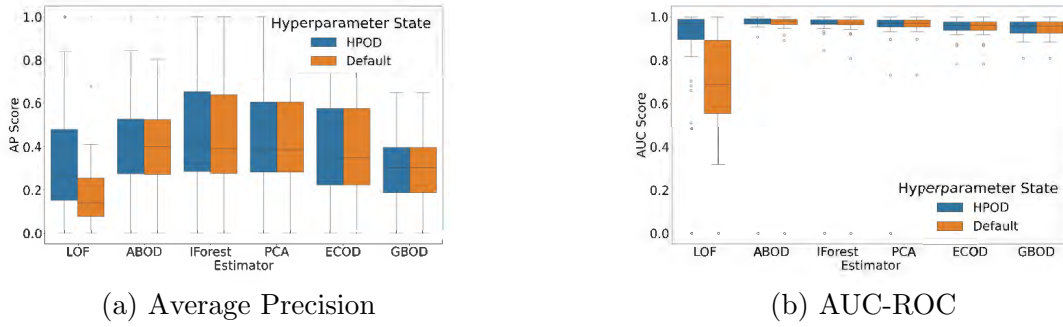


Figure 8.8: Boxplot summarising the performance over the test data sets, on algorithms trained via HPOD’s meta-learning and on default hyperparameters

### 8.2.1 Out-Of-Sample Results

Table 8.2: Depicting the estimators’ average AUC-ROC and average precision values over the 31 OOS orchards, when hyperparameters have been selected with HPOD

| Metric  |                  | ABOD             | IForest          | LOF       | ECOD      | PCA              | GBOD      |
|---------|------------------|------------------|------------------|-----------|-----------|------------------|-----------|
| AP      | <i>Avg.</i>      | <b>0.441 (1)</b> | <b>0.433 (2)</b> | 0.326 (5) | 0.379 (4) | <b>0.421 (3)</b> | 0.324 (6) |
|         | <i>Std. Dev.</i> | 0.232            | 0.239            | 0.247     | 0.232     | 0.230            | 0.205     |
| AUC-ROC | <i>Avg.</i>      | <b>0.946 (1)</b> | <b>0.937 (2)</b> | 0.875 (5) | 0.919 (4) | <b>0.928 (3)</b> | 0.918 (4) |
|         | <i>Std. Dev.</i> | 0.177            | 0.176            | 0.216     | 0.176     | 0.179            | 0.175     |

Table 8.2 depicts that global detection is dominant OOS. Average performance is upheld OOS, capturing decent performance relative to the hyperparameter tuned classifiers presented earlier.

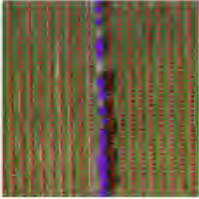
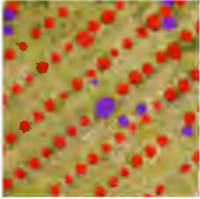
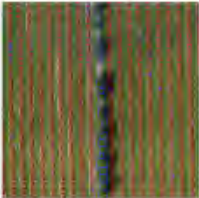
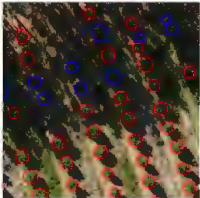
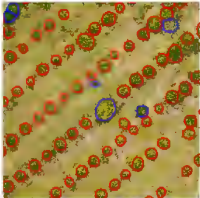
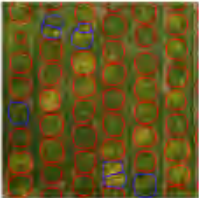
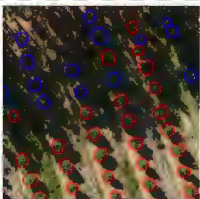
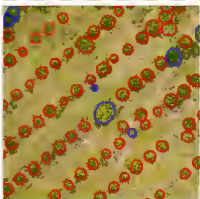
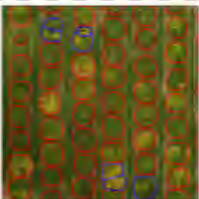
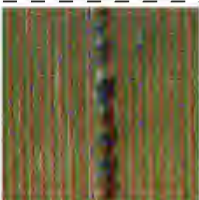
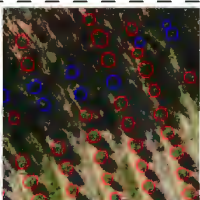
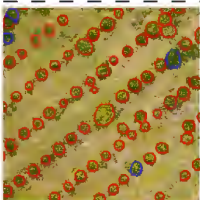
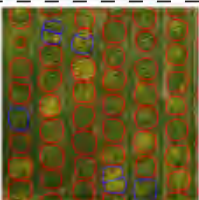
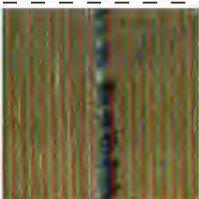
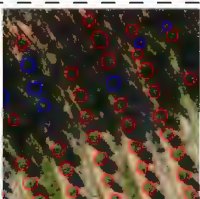
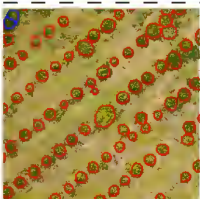
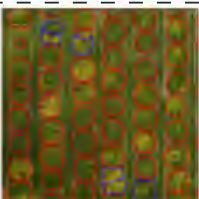
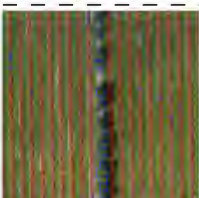
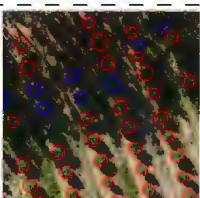
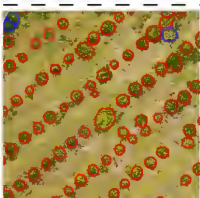
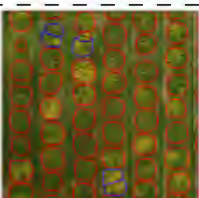
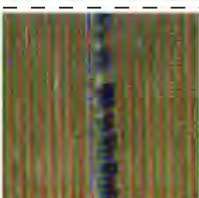
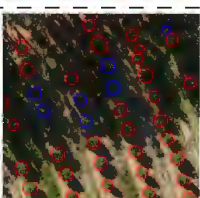
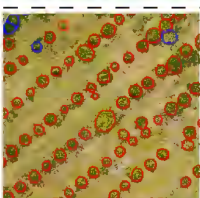
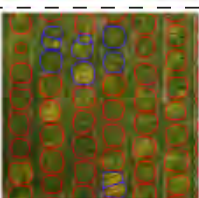
| Orchard 73                                                                          | Orchard 74                                                                          | Orchard 78                                                                          | Orchard 89                                                                           | Method    |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|-----------|
|    |    |    |    | Reference |
|    |    |    |    | ABOD      |
|   |   |   |   | IForest   |
|  |  |  |  | LOF       |
|  |  |  |  | ECOD      |
|  |  |  |  | PCA       |
|  |  |  |  | GBOD      |

Figure 8.9: Visual demonstration of what the classifiers capture in localized areas of orchards 74, 78, and 89

Figure 8.9 illustrates the performance of the classifiers OOS in localised areas of Figure B.2. Orchard 73, demonstrates how well the methods detect false positives. Global methods and GBOD, were efficient in capturing these anomalies, while LOF underperformed in this area. The classifiers were relatively successful in identifying the true anomalies in orchard 74; however, this was done with low precision as many delineations are incorrectly identified as poor. Orchard 78 was chosen to demonstrate the classifiers’ ability to detect under-segmentation. ABOD and IForest were capable of this, while in this scenario all other methods failed to realise this anomaly. Orchard 89, was dedicated to over-segmentation. All methods were efficient in capturing this type of anomaly. ABOD, IForest, and GBOD captured all the anomalies in this region. LOF, ECOD and PCA missed one over-segmentation. As the tree next to the true anomaly appeared more deformed, the concave shape was instead identified as anomalous due to the properties of the feature set.

### 8.3 A Closer Look at GBOD

Figure 8.10b illustrates that IForest highlights problematic points. On the other hand GBOD is highlighting problematic regions, as depicted in Figure 8.10a. This can also be seen in orchards 73, 74 and 89 in Figure 8.9 where the identified anomalies are clustered in areas.

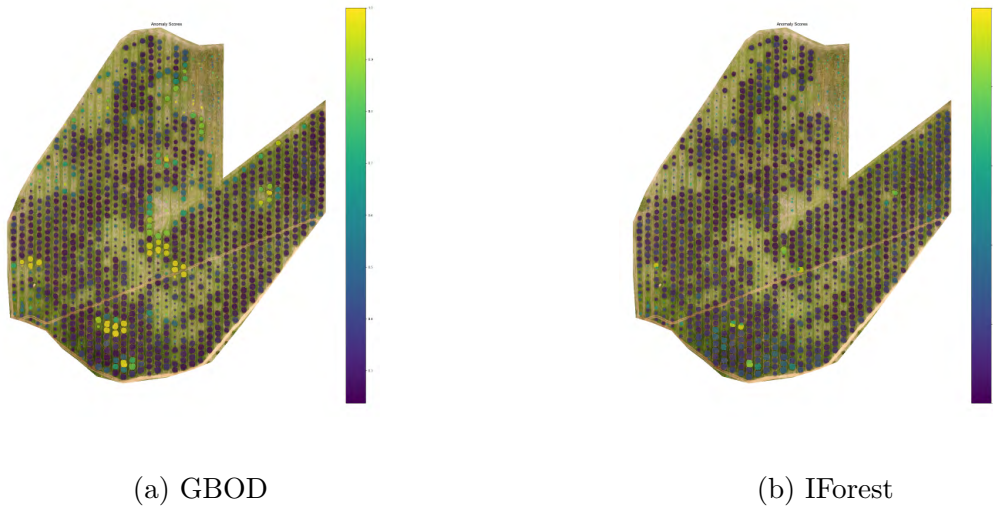


Figure 8.10: The outlier scores of IForest and the GBOD to illustrate how they are appointing outliers

#### 8.3.1 Attribution

Apart from ECOD, GBOD is the only method present that provides a way to attribute outlierness in this study. While looking at orchard 62 — Figure B.3 — as

an example, we can demonstrate the causes of outlieriness.

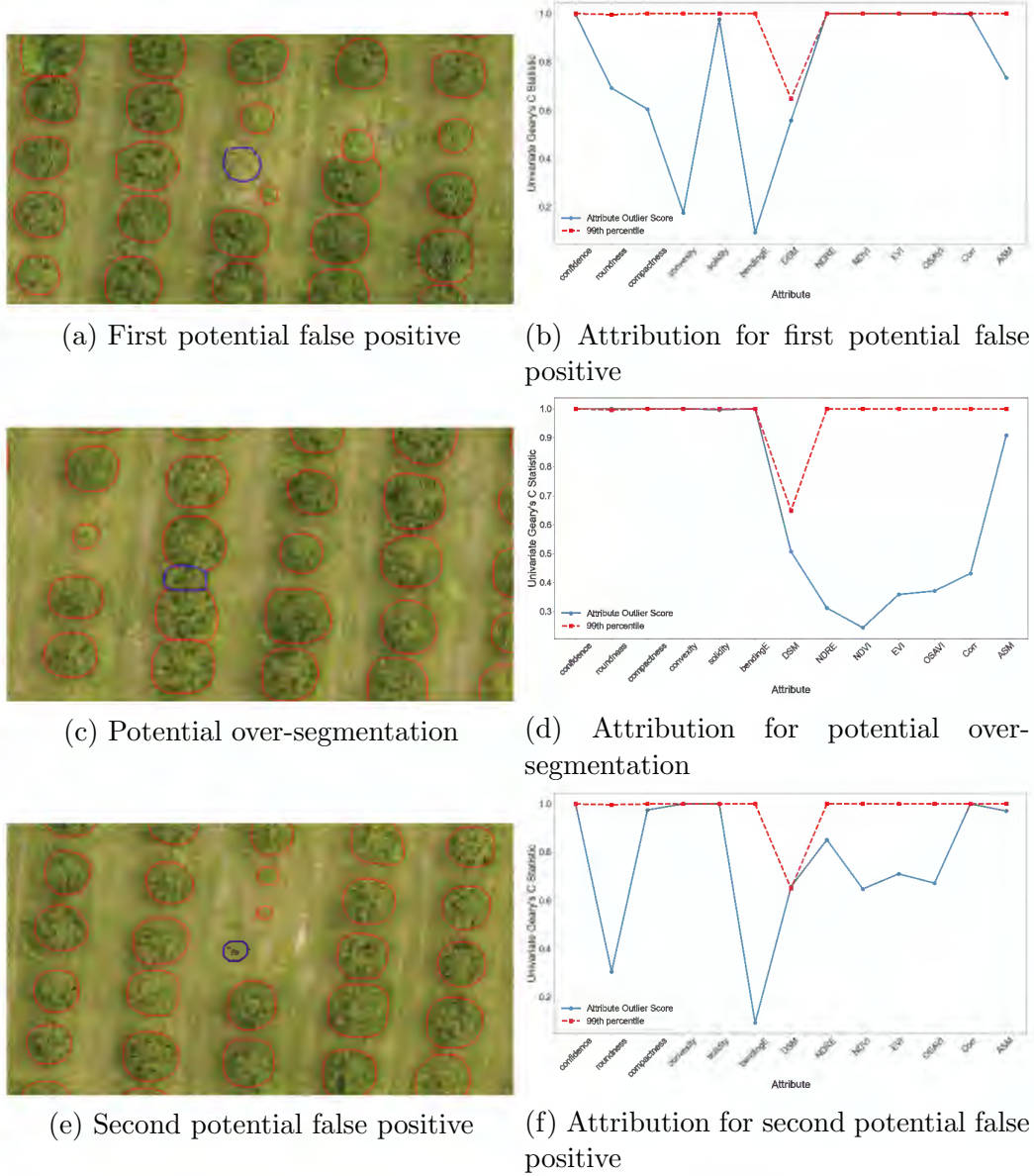


Figure 8.11: Anomalies and their corresponding attributions relative to the 99<sup>th</sup> percentile of the outlier score for each dimension

Figure 8.11a identifies the delineation as a false positive. Figure 8.11b attributes outlieriness to Conf, Sol, Corr and the spectral indices. Figure 8.11c identifies an over-segmented delineation, while Figure 8.11d attributes outlieriness to Conf, and the shape descriptors. Figure 8.11e also identifies a false-positive, attributing outlieriness to Conf, Conv, Sol, DSM, and Corr. GBOD's attribution was successful in confirming many of the hypotheses made about the data. First, it highlighted that Conf is universally important in estimating all anomaly types. It demonstrated that

the shape descriptors are the most crucial identifier for over-segmentations. Next it demonstrated that the spectral indices are important for identifying false positives in some cases.

## 8.4 Discussion

This work found [Table 7.1](#) to be an efficient shallow feature set capable of capturing poor delineations when coupled with shallow unsupervised methods in most cases. After evaluation the ROC- and PR-curves of the estimators were shown to exist above the baseline demonstrating that this data-driven unsupervised approach is feasible for detecting poor delineations. However, there exist orchards — two, five, six and eight — where the PR curve exists near its baseline. This revealed that the feature set has not captured some characteristics or that there are noisy features that mask the outlierness in these scenarios.

The pairwise comparison of the optimised methods illustrated, with statistical significance, that global outlier detection methods are better suited for identifying poor delineations. It also demonstrated that LOF and GBOD become more competitive when over-segmentations dominated orchards. However, the subset of orchards did not provide a large enough sample size to fully reveal the potential of local outlier detection methods or infer statistical significance. This insight aided in confirming the hypothesis that over-segmentations are local outliers, while under-segmentations and false-positives are global outliers. Future research should examine additional orchards dominated by over-segmentations to infer statistical significance.

Thereafter, the research found that allocating hyperparameters via HPOD benefited OOS performance. Therefore, while using the trained meta-learner, performing anomaly detection on any new orchard is possible without hyperparameter tuning or reference data. This demonstrated that the suggested pipeline is automatic and generalisable to new orchards. The OOS evaluation also revealed that the feature set could not define an anomaly in the presence of shadows/shade. While vegetative indices are illumination invariant, Haralick texture features are not. Indicating that despite being broadly applicable, the feature set lacks robustness in certain scenarios. This also disproved the hypothesis mentioned in [Chapter 3](#), that local outlier detection could appropriately handle illumination invariance in orchards. Subsequently, revealing that the detection algorithms and proposed feature set were not as capable in varying illumination conditions. In the case of a lacking feature set, future work should consider other features. Soil attributes are a potential add-on as well as the meticulous selection of low-level shape descriptors. Alternatively, it is possible that attempting to capture all manner of anomalies is detrimental to capturing any one type of anomaly. Solving each problem individually, may mitigate this.

Although the proposed method underperformed against the baselines, it succeeded in accomplishing its task. It was simplistic, interpretable, hyperparameter-free and

computationally efficient. Instead of identifying points as the baseline methods did, GBOD highlighted regions. Local spatial autocorrelation measures assess if neighbourhoods show spatial similarity or heterogeneity. Inherently, this has affected the detection quality for this particular use case. The method demonstrated potential for other domains. Such a method may instead perform better when tasked with identifying diseased or water-deficient parts of the orchard, where clustering is expected.

In terms of time complexity, GBOD demonstrated better performance than ABOD, highlighting another aspect that makes it competitive. The other methods, are amongst some of the faster anomaly detectors available in literature ([Han et al., 2022](#)), wining-out over GBOD despite its competitive speed.

# Chapter 9

## Conclusion

Despite advancements in precision agriculture, the unavoidable capture of poor delineations in ITCD algorithms continues to impede progress in the field. This research aimed to develop an automatic and generalisable approach to identify these anomalies. With this objective in mind, data-driven unsupervised anomaly detection was considered. Our data-centric approach enabled patterns to be objectively identified, avoiding the subjectivity inherent in scene-specific heuristics. This enabled the proposed process to be adaptable to new orchards and sensitive to more edge cases otherwise missed by top-down methods. Unsupervised anomaly detection leveraged the data while addressing the cost and inaccessibility of labelled data, labelling accuracy concerns, and their impact on model generalisation. Exploring this domain further highlighted a research gap in the literature surrounding multidimensional spatial anomaly detection. Therefore, this work proposed a novel multidimensional spatial anomaly detection technique based on multivariate LISA to identify anomalous tree crown delineations. The proposed method was compared with several baseline methods. Thereafter, this study assessed whether anomaly detection of ITCD algorithms can be automatic and generalisable via meta-learning.

In doing so we hoped to discover whether our data-driven unsupervised approach is feasible, if GBOD is competitive with the state-of-the-art and whether the proposed feature set is characteristic of poor delineations.

### 9.1 Key Findings

The proposed feature set consisted of spectral indices, shape descriptors, texture features, confidence and DSM. To ascertain the most salient feature set, an EDA and ablation study were conducted, determining that Circ, Ecc, SAVI, Con and  $Z_{nm}$  were noisy and therefore redundant for further analysis. The feature set was shown to be practical and characterised anomalies in the majority of orchards. However, it failed to account for illumination variability within orchards.

The results of the model evaluation proved that a carefully curated shallow feature set and unsupervised anomaly detection are effective at discerning anomalies. Ranking the benchmarked algorithms by average performance found that global anomaly detection methods — ABOD, IForest and PCA — are best suited for detecting poor delineations in ITCD literature. While only considering over-segmentations, local detectors — LOF and GBOD — exhibited improved performance.

The research also demonstrated the proposed algorithm’s ability to perform adequately against the state-of-the-art, adding to the literature surrounding multidimensional spatial anomaly detection. Still, it showed greater potential in other domains, due to its ability to detect anomalous regions.

The performance OOS — of estimators whose hyperparameters were evaluated by the trained HPOD — was not inferior to that of the tuned in-sample algorithms. This demonstrated that unsupervised anomaly detection of poor delineations can be automatic and generalisable to any new orchard without reference data or hyperparameter tuning.

## 9.2 Future Work

The research exposed various areas of improvement. The feature Con was able to identify offset delineations. The other texture descriptors were also capable of this to some extent. Future work that is insistent on capturing all anomalies, or specifically offset cases, should emphasise the importance of texture descriptors. GBOD is efficient at capturing anomalous regions, this may be an indicator of its capability to identify false negative delineations. Due to the spatial structure embedded in GBOD it is capable of highlighting suspicious connections between neighbouring delineations for this purpose.

It was observed that GBOD is more competitive when compared to the classifiers at default hyperparameters. This work focused on a hyperparameter-free approach to GBOD, however, several improvements can be considered. First, one could consider any number of distance metrics to formulate the equation of the multivariate LISA (Anselin, 2019); second, a neighbourhood constraint can be imposed on edges connected between non-neighbouring crowns based on distance (as Delauney triangulation will connect long edges between non-neighbouring crowns for an optimal triangulation) (Shi et al., 2017); alternative to the previous suggestion, the adjacency matrix presented in this study can be substituted for a weights matrix based on inverse distance weighting to significantly down-weight non-neighbouring crowns; finally, it is common practice to perform a sensitivity analysis in the selection of neighbourhood criterion (proximity graph), therefore, hyperparameter tuning of different graph structures may improve performance.

A final step that should be considered is the labelling and correction of poor delineations. Fine shape descriptors can be used to match shapes efficiently to provide

clusters (Zhang and Lu, 2004). False positives can be removed from the observations. Intersection over union and non-maximum suppression can be used to handle over-segmentations (Vargas-Muñoz et al., 2022). Finally, ultimate erosion and dynamic reconstruction can be used to handle under-segmentation (Wu et al., 2022).

### 9.3 Concluding Remarks

Ultimately, this research has highlighted a robust and unique post-processing approach for ITCD algorithms that performs efficiently when faced with poor delineations in new orchards. This finding is crucial in orchard management, enabling better handling of scarce resources such as water and fertiliser and strategically limiting environmentally harmful pesticides. The generalisability of this study enables the use of this approach in other ITCD studies, such as forest inventory, that require precise delineations to assess carbon capture and conduct efficient forest management. This highlights the potential uses and resilience of the proposed pipeline.

# Bibliography

- Abioye, E. A., Abidin, M. S. Z., Mahmud, M. S. A., Buyamin, S., Ishak, M. H. I., Rahman, M. K. I. A., Otuoze, A. O., Onotu, P., and Ramli, M. S. A. (2020). A review on monitoring and advanced control strategies for precision irrigation. *Computers and Electronics in Agriculture*, 173:105441.
- Aggarwal, C. C. (2016). *Outlier Analysis*. Springer Publishing Company, Incorporated, 2nd edition.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Akoglu, L., Tong, H., and Koutra, D. (2015). Graph based anomaly detection and description: a survey. *Data Min. Knowl. Discov.*, 29(3):626–688.
- Alganci, U., Besol, B., and Sertel, E. (2018). Accuracy assessment of different digital surface models. *ISPRS International Journal of Geo-Information*, 7(3).
- Almardeny, Y., Boujnah, N., and Cleary, F. (2022). A novel outlier detection method for multivariate data. *IEEE Transactions on Knowledge and Data Engineering*, 34(9):4052–4062.
- Altman, M. B., Kavanaugh, J. A., Wooten, H. O., Green, O. L., DeWees, T. A., Gay, H., Thorstad, W. L., Li, H., and Mutic, S. (2015). A framework for automated contour quality assurance in radiation therapy including adaptive techniques. *Physics in Medicine & Biology*, 60(13):5199.
- Alves Varella, C. A., Gleriani, J. M., and dos Santos, R. M. (2015). Chapter 9 - precision agriculture and remote sensing. In Santos, F., Borém, A., and Caldas, C., editors, *Sugarcane*, pages 185–203. Academic Press, San Diego.
- Ampatzidis, Y. and Partel, V. (2019). UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*, 11(4).

- Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical Analysis*, 27(2):93–115.
- Anselin, L. (2019). A local indicator of multivariate spatial association: Extending geary’s c. *Geographical Analysis*, 51(2):133–150.
- Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., and Moran, M. S. (2000). Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. In *Proceedings of the 5th International Conference on Precision Agriculture*, page 1–15. American Society of Agronomy, Madison.
- Berg, M. d., Cheong, O., van Kreveld, M., and Overmars, M. (2008). *Delaunay Triangulations*, pages 191–218. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Bivand, R., Pebesma, E., and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Use R! Springer New York.
- Boiarskii, B. and Hasegawa, H. (2019). Comparison of NDVI and ndre indices to detect differences in vegetation and chlorophyll content. *Journal Of Mechanics Of Continua and Mathematical Sciences*, spl1:20–29.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). LOF: identifying density-based local outliers. *SIGMOD Record*, 29(2):93–104.
- Bribiesca, E. (1997). Measuring 2-d shape compactness using the contact perimeter. *Computers & Mathematics with Applications*, 33(11):1–9.
- Buschmann, C. and Nagel, E. (1993). In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*, 14(4):711–722.
- Caruso, G., Zarco-Tejada, P. J., González-Dugo, V., Moriondo, M., Tozzini, L., Palai, G., Rallo, G., Hornero, A., Primicerio, J., and Gucci, R. (2019). High-resolution imagery acquired from an unmanned platform to estimate biophysical and geometrical parameters of olive trees under different irrigation regimes. *PLOS ONE*, 14(1):1–19.
- Chalapathy, R. and Chawla, S. (2019). Deep learning for anomaly detection: A survey.
- Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).
- Chawla, S. and Sun, P. (2006). SLOM: a new measure for local spatial outliers. *Knowledge and Information Systems*, 9(4):412–429.

- Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848.
- Cheng, Z., Qi, L., Cheng, Y., Wu, Y., and Zhang, H. (2020). Interlacing orchard canopy separation and assessment using UAV images. *Remote Sensing*, 12(5).
- Cressie, N. A. C. (1993). *Geostatistics*, chapter 2, pages 27–104. John Wiley & Sons, Ltd.
- Culvenor, D. S. (2002). TIDA: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences*, 28(1):33–44.
- Davis, J. and Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, page 233–240, New York, NY, USA. Association for Computing Machinery.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.
- Deng, M., Yang, X., Shi, Y., Gong, J., Tang, J., and Liu, Y. (2018). A non-parametric statistical test method to detect significant cross-outliers in spatial points. *Transactions in GIS*, 22(6):1462–1483.
- Dimitrios Panagiotidis, Azadeh Abdollahnejad, P. S. and Chiteculo, V. (2017). Determining tree height and crown diameter from high-resolution UAV imagery. *International Journal of Remote Sensing*, 38(8-10):2392–2410.
- Donald G. Leckie, N. W. and Gougeon, F. (2016). Recognition and possible remediation of automated tree delineations with multiple isolations per tree (split cases) on high-resolution imagery. *Canadian Journal of Remote Sensing*, 42(6):656–679.
- Dong, X., Zhang, Z., Yu, R., Tian, Q., and Zhu, X. (2020). Extraction of information about individual trees from high-spatial-resolution UAV-acquired images of an orchard. *Remote Sensing*, 12(1).
- Dražić, S., Ralević, N., and Žunić, J. (2010). Shape elongation from optimal encasing rectangles. *Computers & Mathematics with Applications*, 60(7):2035–2042.
- Ernst, M. and Haesbroeck, G. (2017). Comparison of local outlier detection techniques in spatial multivariate data. *Data Mining and Knowledge Discovery*, 31(2):371–399.
- Filzmoser, P., Ruiz-Gazen, A., and Thomas-Agnan, C. (2014). Identification of local multivariate outliers. *Statistical Papers*, 55(1):29–47.

- Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3):115–141.
- Ghalati, M. K., Nunes, A., Ferreira, H., Serranho, P., and Bernardes, R. (2022). Texture analysis and its applications in biomedical imaging: A survey. *IEEE Reviews in Biomedical Engineering*, 15:222–246.
- Goldstein, M. and Dengel, A. (2012). Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. *KI-2012: poster and demo track*, 1:59–63.
- Gougeon, F. A. (1995). A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian journal of remote sensing*, 21(3):274–284.
- Han, S., Hu, X., Huang, H., Jiang, M., and Zhao, Y. (2022). ADBench: Anomaly detection benchmark. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- Hariri, S., Kind, M. C., and Brunner, R. J. (2021). Extended isolation forest. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1479–1489.
- Hawkins, D. M. (1974). The detection of errors in multivariate data using principal components. *Journal of the American Statistical Association*, 69(346):340–344.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397.
- He, Z., Xu, X., and Deng, S. (2003). Discovering cluster-based local outliers. *Pattern Recognition Letters*, 24(9):1641–1650.
- He, Z. and Yu, W. (2010). Stable feature selection for biomarker discovery. *Computational Biology and Chemistry*, 34(4):215–225.
- Hentschel, M. L. and Page, N. W. (2003). Selection of descriptors for particle shape characterization. *Particle & Particle Systems Characterization*, 20(1):25–38.
- Hodge, V. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126.
- Huang, H., Li, X., and Chen, C. (2018). Individual tree crown detection and delineation from very-high-resolution UAV images based on bias field and marker-controlled watershed segmentation algorithms. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(7):2253–2262.

- Huang, Y., Ren, Z., Li, D., and Liu, X. (2020). Phenotypic techniques and applications in fruit trees: a review. *Plant Methods*, 16(1):107.
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3):295–309.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195–213. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.
- Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963.
- Julzarika, A. and Harintaka (2019). Indonesian DEMNAS: DSM or DTM? In *2019 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS)*, pages 31–36.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York: New York.
- Ke, Y. and Quackenbush, L. J. (2011). A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *International Journal of Remote Sensing*, 32(17):4725–4747.
- Kitagawa, H., Gibbons, P. B., Papadimitriou, S., and Faloutsos, C. (2003). LOCI: Fast Outlier Detection Using the Local Correlation Integral . In *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, page 315, Los Alamitos, CA, USA. IEEE Computer Society.
- Knorr, E. M. and Ng, R. T. (1998). Algorithms for mining distance-based outliers in large datasets. In *Proceedings of the 24rd International Conference on Very Large Data Bases, VLDB '98*, page 392–403, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Kriegel, H.-P., Kröger, P., Schubert, E., and Zimek, A. (2009). Outlier detection in axis-parallel subspaces of high dimensional data. In *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, PAKDD '09*, page 831–838, Berlin, Heidelberg. Springer-Verlag.
- Kriegel, H.-P., Schubert, M., and Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, page 444–452, New York, NY, USA. Association for Computing Machinery.

- Kulkarni, S., Borde, P., Manza, R., and Yannawar, P. (2014). Recognition of handwritten MODI numerals using hu and zernike features. *Computing Research Repository (CoRR)*, abs/1404.1151. Withdrawn.
- Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., and Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. *Cluster Computing*, 22(1):949–961.
- Latecki, L. J., Lazarevic, A., and Pokrajac, D. (2007). Outlier detection with kernel density functions. In Perner, P., editor, *Machine Learning and Data Mining in Pattern Recognition*, pages 61–75, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Laurikkala, J., Juhola, M., and Kentala, E. (2000). Informal identification of outliers in medical data. *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*.
- Leckie, D. G., Walsworth, N., and Gougeon, F. A. (2016). Identifying tree crown delineation shapes and need for remediation on high resolution imagery using an evidence based approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:206–227.
- Li, J., Wu, Y., Zhang, H., and Wang, H. (2023). A novel unsupervised segmentation method of canopy images from UAV based on hybrid attention mechanism. *Electronics*, 12(22).
- Li, Z., Zhao, Y., Botta, N., Ionescu, C., and Hu, X. (2020). COPOD: Copula-Based Outlier Detection. In *2020 IEEE International Conference on Data Mining (ICDM)*, page 1118–1123. IEEE.
- Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., and Chen, G. H. (2022). ECOD: unsupervised outlier detection using empirical cumulative distribution functions. *CoRR*, abs/2201.00382.
- Lisiewicz, M., Kamińska, A., Kraszewski, B., and Stereńczak, K. (2022). Correcting the Results of CHM-Based Individual Tree Detection Algorithms to Improve Their Accuracy and Reliability. *Remote Sensing*, 14(8).
- Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422.
- Lynham, P. (2022). Chapter 2 - from data to information. In Mottram, T., editor, *Digital Agritechnology*, pages 11–47. Academic Press.
- Manolakis, D. G., Lockwood, R. B., and Cooley, T. W. (2016). *Spectral Properties of Materials*, page 117–153. Cambridge University Press.

- Mu, Y., Fujii, Y., Takata, D., Zheng, B., Noshita, K., Honda, K., Ninomiya, S., and Guo, W. (2018). Characterization of peach tree crown by using high-resolution images from an unmanned aerial vehicle. *Horticulture Research*, 5:74.
- Ning, J., Chen, L., and Chen, J. (2018). Relative density-based outlier detection algorithm. In *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, CSAI '18, page 227–231, New York, NY, USA. Association for Computing Machinery.
- Pang, G., Shen, C., Cao, L., and Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2).
- Papaconstadopoulos, P., González, P., Carbaat, C., Simões, R., Beemster, H., Koetsveld, F., Remeijer, P., and Janssen, T. (2021). An anomaly detector as a clinical decision support system for parotid gland delineations. *Physics in Medicine & Biology*, 66(10):105017.
- Parra, L., Deco, G., and Miesbach, S. (1997). Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation*, 8(2):260–269.
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9):503–510.
- Pforte, F., Selbeck, J., and Hensel, O. (2012). Comparison of two different measurement techniques for automated determination of plum tree canopy cover. *Biosystems Engineering*, 113(4):325–333.
- Pinz, A. (1991). A computer vision system for the recognition of trees in aerial photographs. *Multisource Data Integration in Remote Sensing*, 3099:111–124.
- Plant, R. E., Pettygrove, G. S., and Reinert, W. (2000). Precision agriculture can increase profits and limit environmental impacts. *California Agriculture*, 54(4):66–71.
- Pourghahramani, P. and Forssberg, E. (2005). Review of applied particle shape descriptors and produced particle shapes in grinding environments. part i: Particle shape descriptors. *Mineral Processing and Extractive Metallurgy Review*, 26(2):145–166.
- Radočaj, D., Šiljeg, A., Marinović, R., and Jurišić, M. (2023). State of major vegetation indices in precision agriculture studies indexed in web of science: A review. *Agriculture*, 13(3).

- Ramaswamy, S., Rastogi, R., and Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 427–438, New York, NY, USA. Association for Computing Machinery.
- Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2):95–107.
- Rossi, A., Tiezzi, M., Dimitri, G. M., Bianchini, M., Maggini, M., and Scarselli, F. (2018). Inductive–transductive learning with graph neural networks. In Pancioni, L., Schwenker, F., and Trentin, E., editors, *Artificial Neural Networks in Pattern Recognition*, pages 201–212, Cham. Springer International Publishing.
- Roth, S. I., Leiterer, R., Volpi, M., Celio, E., Schaepman, M. E., and Joerg, P. C. (2019). Automated detection of individual clove trees for yield quantification in northeastern Madagascar based on multi-spectral satellite data. *Remote Sensing of Environment*, 221:144–156.
- Rouse, J. W., J., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. In Freden, S. C., Mercanti, E. P., and Becker, M. A., editors, *NASA Special Publication*, volume 351, page 309. NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A.
- Roussos, P. A. and Gasparatos, D. (2009). Apple tree growth and overall fruit quality under organic and conventional orchard management. *Scientia Horticulturae*, 123(2):247–252.
- Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., and Muller, K.-R. (2021). A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5):756–795.
- Santhosh, K. K., Dogra, D. P., and Roy, P. P. (2020). Anomaly detection in road traffic using visual surveillance: A survey. *ACM Computing Surveys*, 53(6).
- Schneider, P. and Xhafa, F. (2022). Chapter 3 - anomaly detection: Concepts and methods. In Schneider, P. and Xhafa, F., editors, *Anomaly Detection and Complex Event Processing over IoT Data Streams*, pages 49–66. Academic Press.
- Shekhar, S. and Chawla, S. (2003). *A tour of spatial databases*. Prentice Hall Upper Saddle River.
- Shekhar, S., Lu, C.-T., and Zhang, P. (2002). Detecting graph-based spatial outliers. *Intelligent Data Analysis*, 6:451–468. 5.
- Shi, Y., Deng, M., Yang, X., and Liu, Q. (2017). A spatial anomaly points and regions detection method using multi-constrained graphs and local density. *Transactions in GIS*, 21(2):376–405.

- Shi, Y., Gong, J., Deng, M., Yang, X., and Xu, F. (2018). A graph-based approach for detecting spatial cross-outliers from two types of spatial point events. *Computers, Environment and Urban Systems*, 72:88–103.
- Shukla, S. and Lalitha, S. (2023). Geographically weighted comedian method for spatial outlier detection. *Japanese Journal of Statistics and Data Science*, 6(1):279–299.
- Shumilo, L. (2020). Automatic anomaly detection methodology for crop classification data using morphological features. In *2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, pages 1–5.
- Shyu, M.-L., Chen, S.-C., Sarinnapakorn, K., and Chang, L. (2003). A novel anomaly detection scheme based on principal component classifier. In *Proceedings of the IEEE foundations and new directions of data mining workshop*, pages 172–179. IEEE Press Piscataway, NJ, USA.
- Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, 38:100306.
- Stafford, J. V. (2000). Implementing precision agriculture in the 21st century. *Journal of Agricultural Engineering Research*, 76(3):267–275.
- Tang, B. and He, H. (2017). A local density-based approach for outlier detection. *Neurocomputing*, 241:171–180.
- Tang, J., Chen, Z., Fu, A. W.-c., and Cheung, D. W. (2002). Enhancing effectiveness of outlier detections for low density patterns. In Chen, M.-S., Yu, P. S., and Liu, B., editors, *Advances in Knowledge Discovery and Data Mining*, pages 535–548, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Teague, M. R. (1980). Image analysis via the general theory of moments\*. *Journal of the Optical Society of America A*, 70(8):920–930.
- Terashima, N. (2002). 12 - computer vision. In Terashima, N., editor, *Intelligent Communication Systems*, pages 149–179. Academic Press, San Diego.
- Tuceryan, M. and Jain, A. K. (1993). *TEXTURE ANALYSIS*, chapter 2.1, pages 235–276. World Scientific Publishing Co Pte Ltd.
- Usha, K. and Singh, B. (2013). Potential applications of remote sensing in horticulture—a review. *Scientia Horticulturae*, 153:71–83.
- Vapnik, V. (2013). *The Nature of Statistical Learning Theory*. Information Science and Statistics. Springer New York.

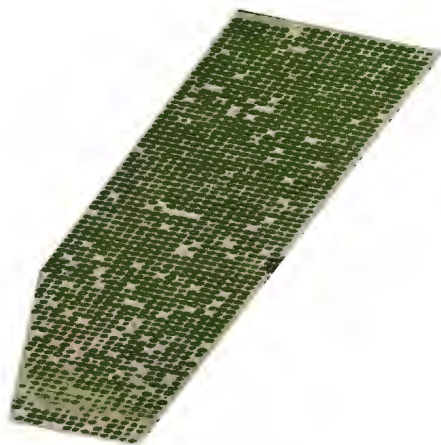
- Vargas-Muñoz, J. E., Lobry, S., Falcão, A. X., and Tuia, D. (2019). Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147:283–293.
- Vargas-Muñoz, J. E., Schibli, D., and Tuia, D. (2022). Towards efficient correction of coconut tree detection errors. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 5065–5068.
- von F. Zernike (1934). Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica*, 1(7):689–704.
- Wang, H., Bah, M. J., and Hammad, M. (2019). Progress in outlier detection techniques: A survey. *IEEE Access*, 7:107964–108000.
- Wang, L., Gong, P., and Biging, G. S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogrammetric Engineering & Remote Sensing*, 70(3):351–357.
- Wu, H., Souedet, N., Jan, C., Clouchoux, C., and Delzescaux, T. (2022). A general deep learning framework for neuron instance segmentation based on Efficient UNet and morphological post-processing. *Computers in Biology and Medicine*, 150:106180.
- Xu, H., Pang, G., Wang, Y., and Wang, Y. (2023). Deep isolation forest for anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12591–12604.
- Ye, D., Wu, L., Li, X., Atoba, T. O., Wu, W., and Weng, H. (2023). A synthetic review of various dimensions of non-destructive plant stress phenotyping. *Plants*, 12(8).
- Young, I. T., Walker, J. E., and Bowie, J. E. (1974). An analysis technique for biological shape. i. *Information and Control*, 25(4):357–370.
- Yousef Erfanifard, Krzysztof Stereńczak, B. K. and Kamińska, A. (2018). Development of a robust canopy height model derived from ALS point clouds for predicting individual crown attributes at the species level. *International Journal of Remote Sensing*, 39(23):9206–9227.
- Zdilla, M. J., Hatfield, S. A., McLean, K. A., Cyrus, L. M., Laslo, J. M., and Lambert, H. W. (2016). Circularity, Solidity, Axes of a Best Fit Ellipse, Aspect Ratio, and Roundness of the Foramen Ovale: A Morphometric Analysis With Neurosurgical Considerations. *Journal of Craniofacial Surgery*, 27(1).
- Zhang, C., Valente, J., Kooistra, L., Guo, L., and Wang, W. (2019a). Opportunities of UAVs in orchard management. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13:673–680.

- Zhang, C., Valente, J., Kooistra, L., Guo, L., and Wang, W. (2021). Orchard management with small unmanned aerial vehicles: a survey of sensing and analysis approaches. *Precision Agriculture*, 22(6):2007–2052.
- Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19.
- Zhang, J. and Tan, T. (2002). Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735–747. Image/Video Communication.
- Zhang, K., Hutter, M., and Jin, H. (2009). A new local distance-based outlier detection approach for scattered real-world data. In Theeramunkong, T., Kijsirikul, B., Cercone, N., and Ho, T.-B., editors, *Advances in Knowledge Discovery and Data Mining*, pages 813–822, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Zhang, Y., Plautz, T. E., Hao, Y., Kinchen, C., and Li, X. A. (2019b). Texture-based, automatic contour validation for online adaptive replanning: A feasibility study on abdominal organs. *Medical Physics*, 46(9):4010–4020.
- Zhao, H., Morgenroth, J., Pearse, G., and Schindler, J. (2023). A systematic review of individual tree crown detection and delineation with convolutional neural networks (CNN). *Current Forestry Reports*, 9(3):149–170.
- Zhao, Y. and Akoglu, L. (2022). Hyperparameter optimization for unsupervised outlier detection.
- Zhao, Y., Nasrullah, Z., and Li, Z. (2019). PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research*, 20(96):1–7.
- Zhen, Z., Quackenbush, L. J., and Zhang, L. (2016). Trends in automatic individual tree crown detection and delineation—evolution of lidar data. *Remote Sensing*, 8(4).

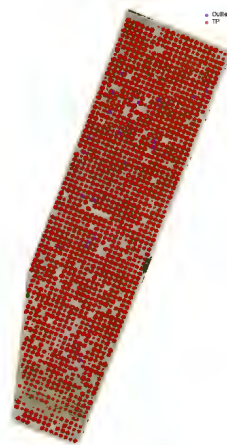
# Appendix A

## Exploratory Data Analysis

### A.1 Visual Analysis



(a) Orchard 1



(b) Orchard 1 reference data



(c) Orchard 2



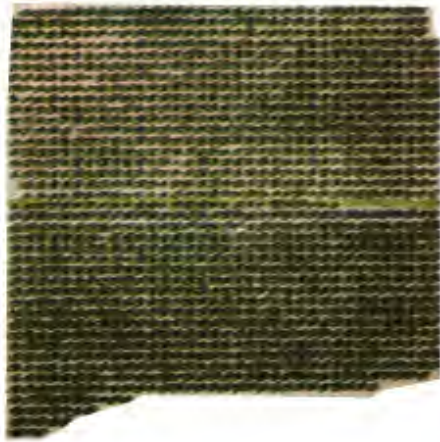
(d) Orchard 2 reference data



(e) Orchard 3



(f) Orchard 3 reference data



(g) Orchard 4



(h) Orchard 4 reference data



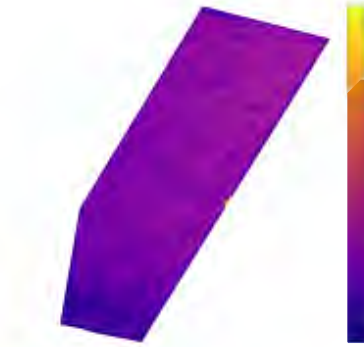
(i) Orchard 5



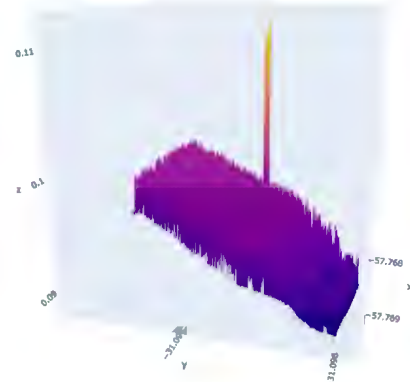
(j) Orchard 5 reference data

Figure A.1: Depicting the orchard and reference data of the the first 5 randomly selected training datasets.

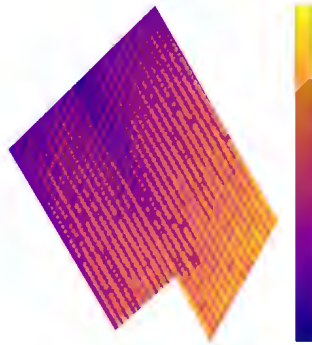
## A.2 DSM Profile Plots



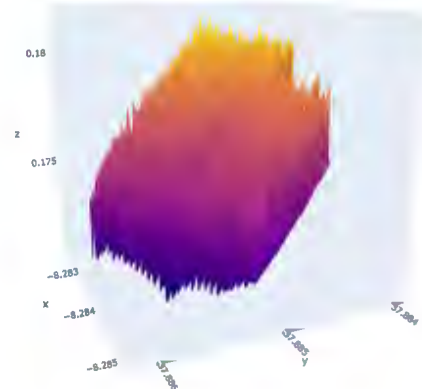
(a) Orchard 1 2-D Representation



(b) Orchard 1 3-D Representation



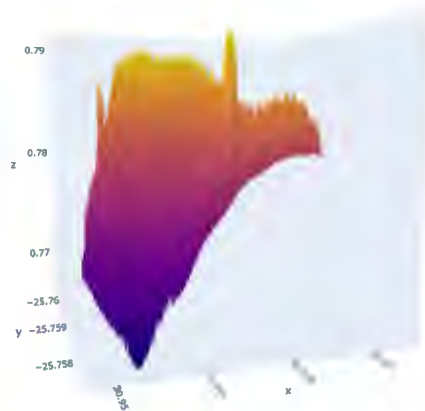
(c) Orchard 2 2-D Representation



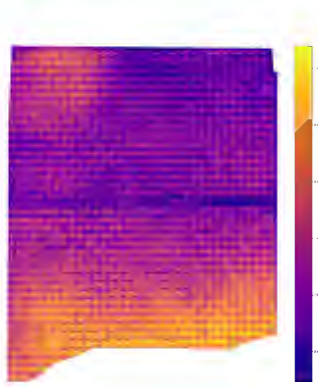
(d) Orchard 2 3-D Representation



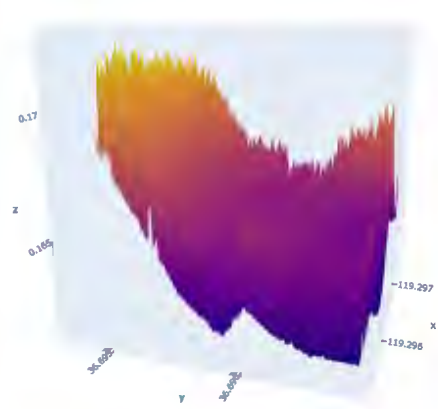
(e) Orchard 3 2-D Representation



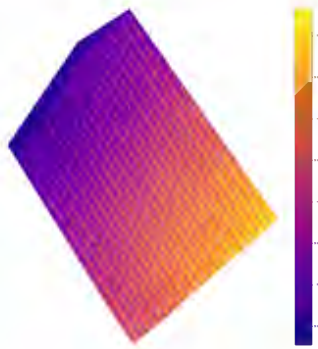
(f) Orchard 3 3-D Representation



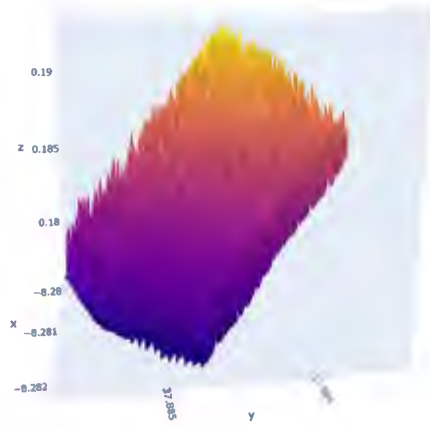
(g) Orchard 4 2-D Representation



(h) Orchard 4 3-D Representation



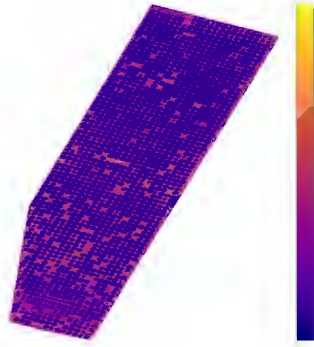
(i) Orchard 5 2-D Representation



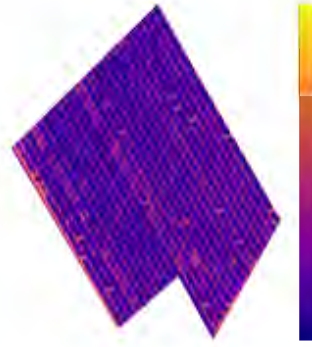
(j) Orchard 5 3-D Representation

Figure A.2: Depicting the 2-D and 3-D DSM data of the the first 5 randomly selected training datasets.

### A.3 Spectral Profile Plots



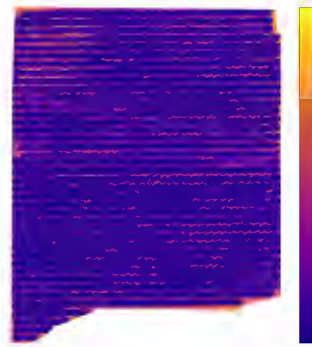
(a) Orchard 1



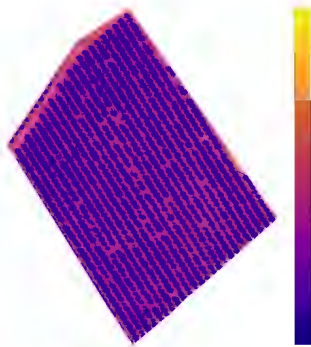
(b) Orchard 2



(c) Orchard 3

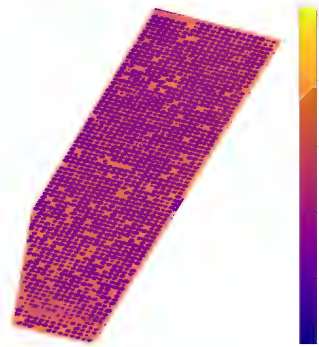


(d) Orchard 4

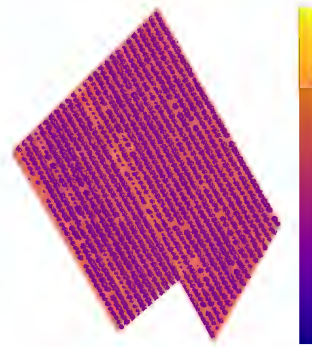


(e) Orchard 5

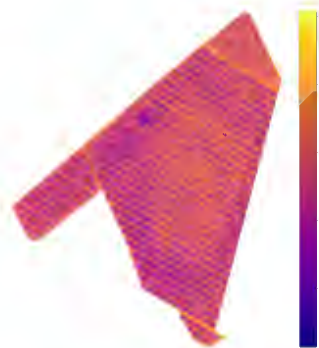
Figure A.3: Profile plot of red aerial reflectance band of the first 5 randomly selected training datasets.



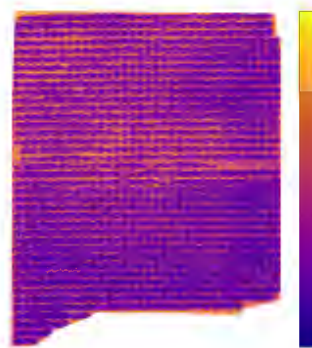
(a) Orchard 1



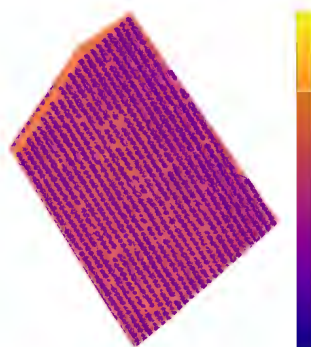
(b) Orchard 2



(c) Orchard 3

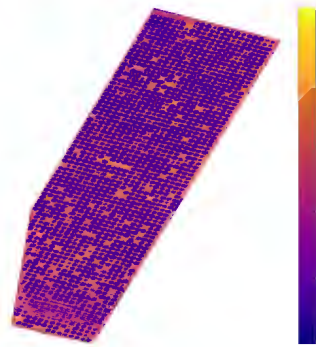


(d) Orchard 4

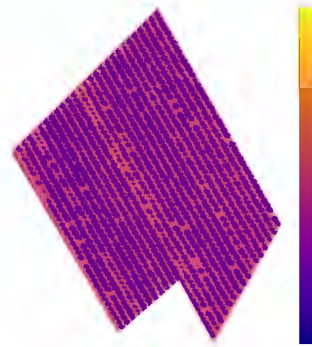


(e) Orchard 5

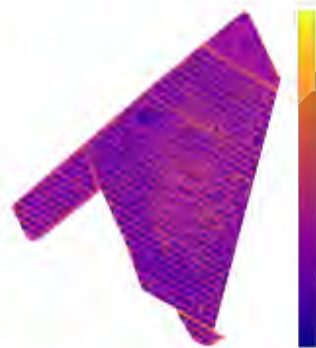
Figure A.4: Profile plot of green aerial reflectance band of the first 5 randomly selected training datasets.



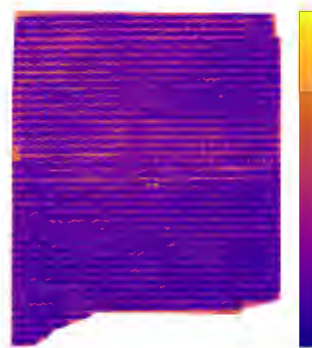
(a) Orchard 1



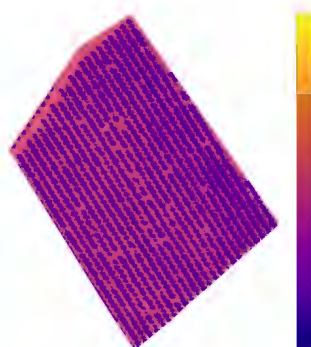
(b) Orchard 2



(c) Orchard 3



(d) Orchard 4

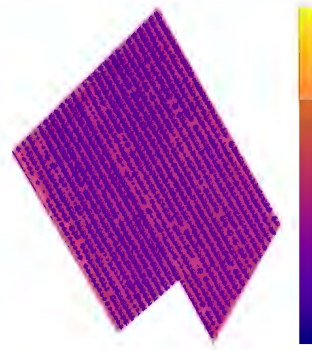


(e) Orchard 5

Figure A.5: Profile plot of blue aerial reflectance band of the first 5 randomly selected training datasets.



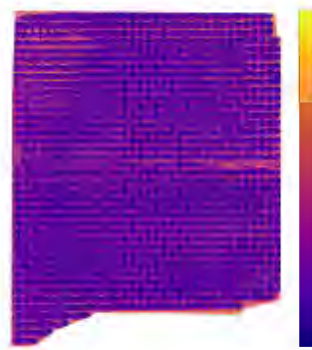
(a) Orchard 1



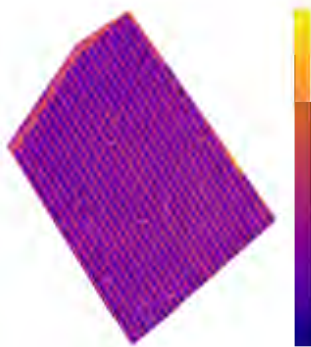
(b) Orchard 2



(c) Orchard 3

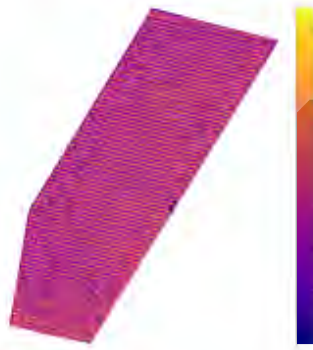


(d) Orchard 4

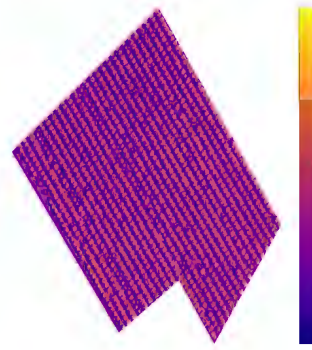


(e) Orchard 5

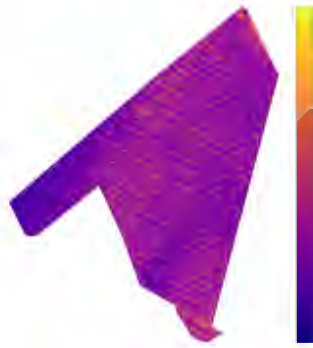
Figure A.6: Profile plot of REG aerial reflectance band of the first 5 randomly selected training datasets.



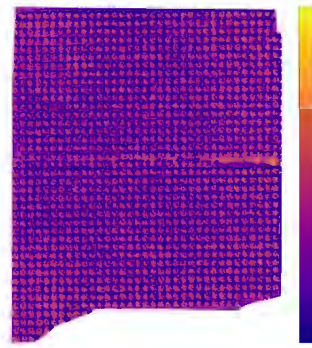
(a) Orchard 1



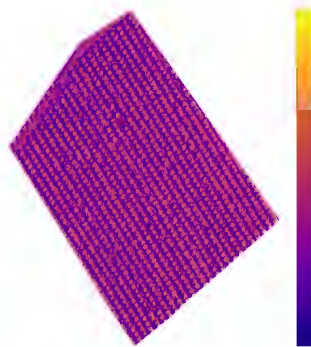
(b) Orchard 2



(c) Orchard 3

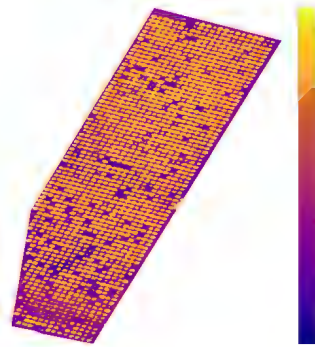


(d) Orchard 4

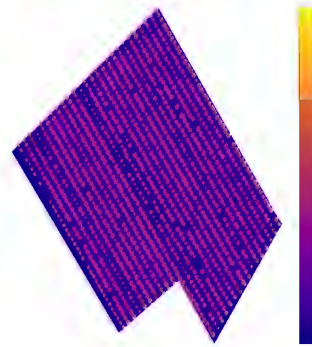


(e) Orchard 5

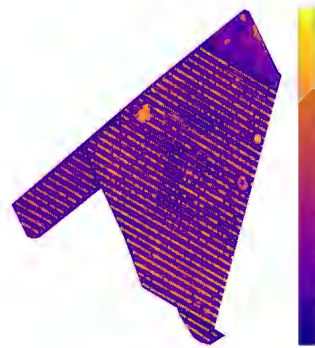
Figure A.7: Profile plot of NIR aerial reflectance band of the first 5 randomly selected training datasets.



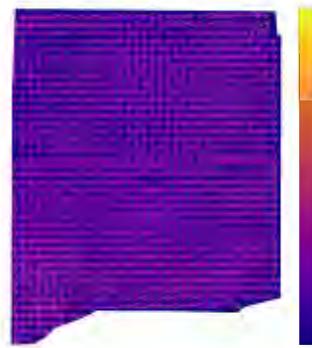
(a) Orchard 1



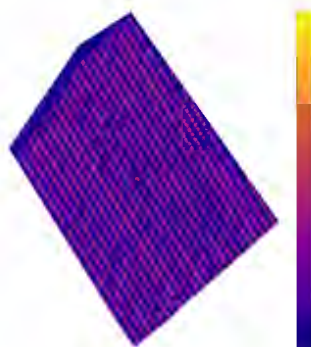
(b) Orchard 2



(c) Orchard 3



(d) Orchard 4



(e) Orchard 5

Figure A.8: Profile plot of EVI vegetative index of the first 5 randomly selected training datasets.



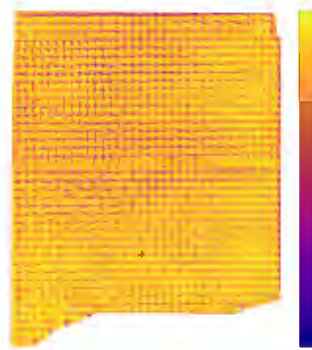
(a) Orchard 1



(b) Orchard 2



(c) Orchard 3

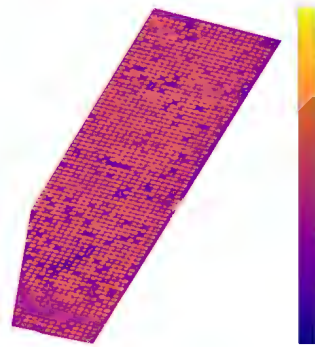


(d) Orchard 4

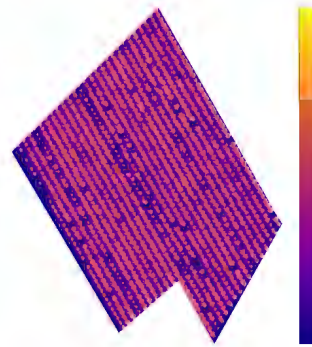


(e) Orchard 5

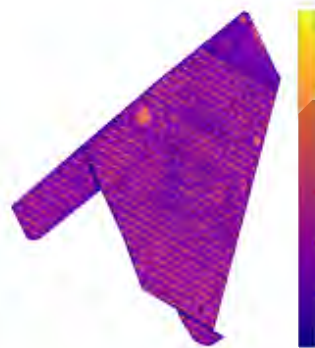
Figure A.9: Profile plot of GNDVI vegetative index of the first 5 randomly selected training datasets.



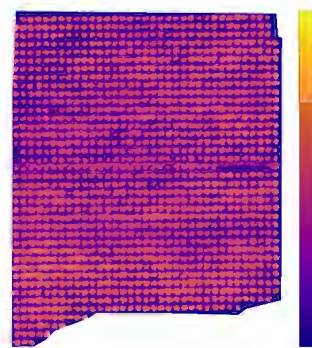
(a) Orchard 1



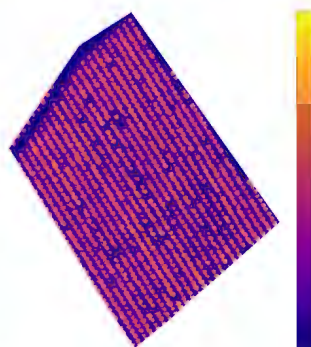
(b) Orchard 2



(c) Orchard 3

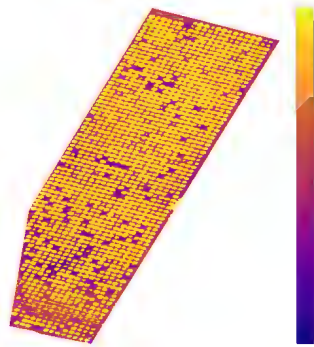


(d) Orchard 4

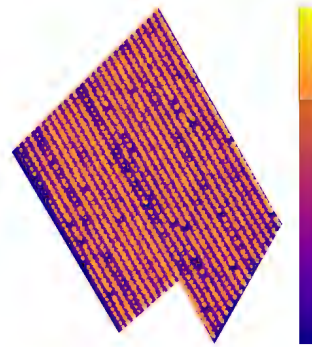


(e) Orchard 5

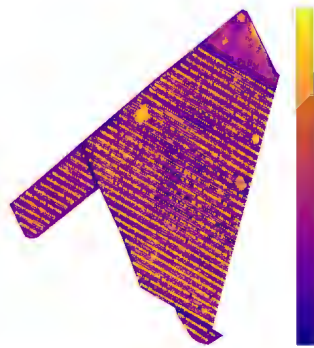
Figure A.10: Profile plot of NDRE vegetative index of the first 5 randomly selected training datasets.



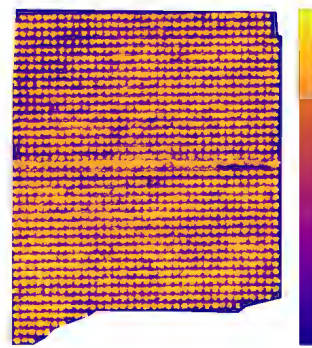
(a) Orchard 1



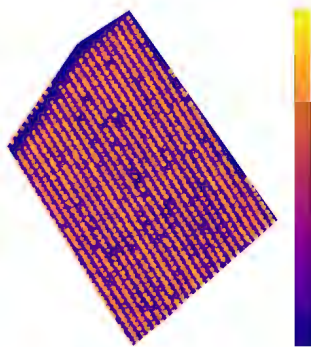
(b) Orchard 2



(c) Orchard 3

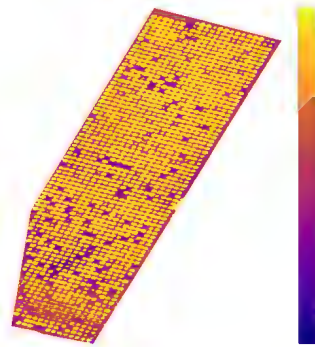


(d) Orchard 4

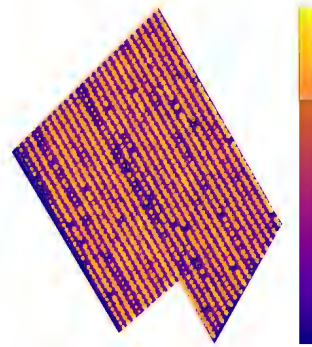


(e) Orchard 5

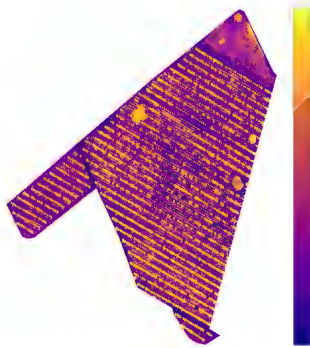
Figure A.11: Profile plot of NDVI vegetative index of the first 5 randomly selected training datasets.



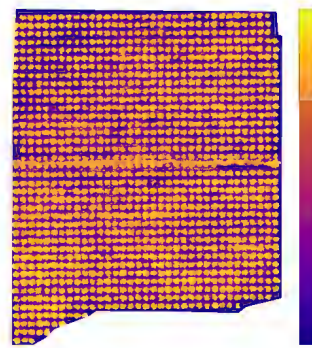
(a) Orchard 1



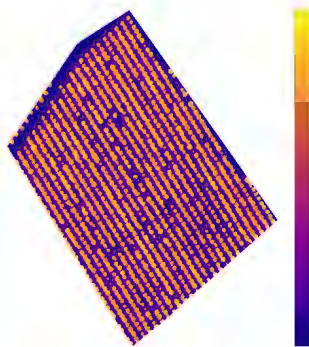
(b) Orchard 2



(c) Orchard 3

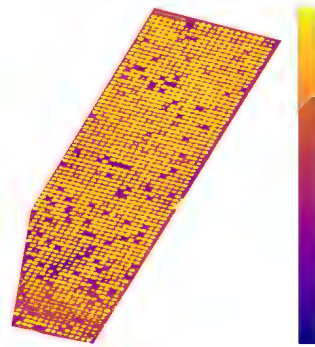


(d) Orchard 4

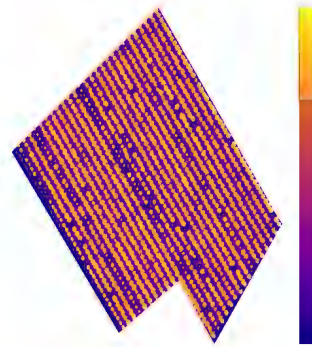


(e) Orchard 5

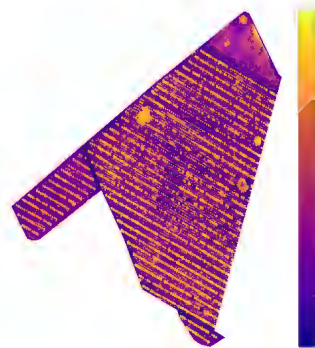
Figure A.12: Profile plot of OSAVI vegetative index of the first 5 randomly selected training datasets.



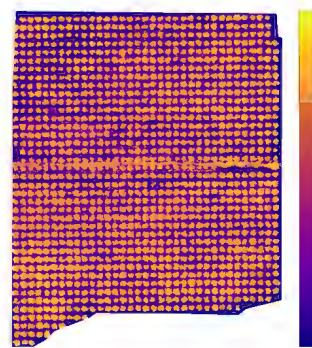
(a) Orchard 1



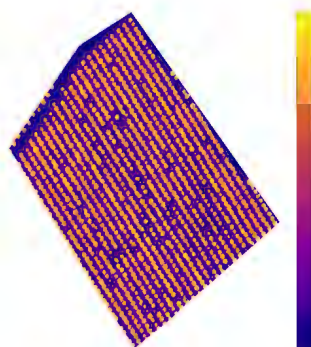
(b) Orchard 2



(c) Orchard 3



(d) Orchard 4



(e) Orchard 5

Figure A.13: Profile plot of SAVI vegetative index of the first 5 randomly selected training datasets.

## A.4 Univariate Analysis

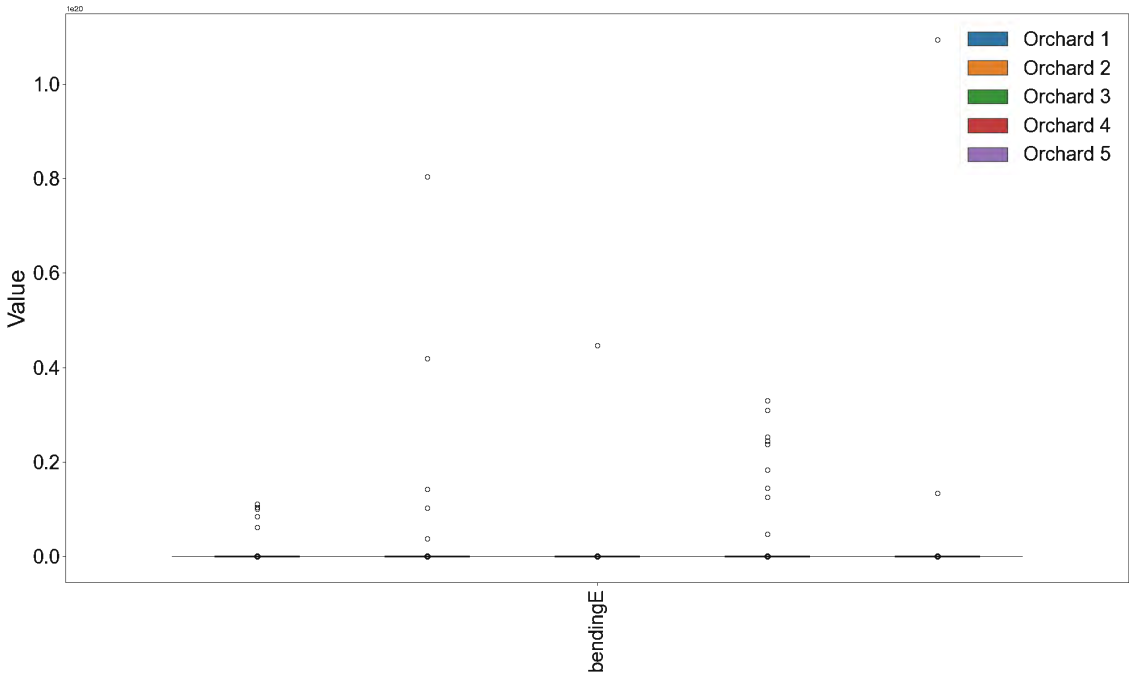
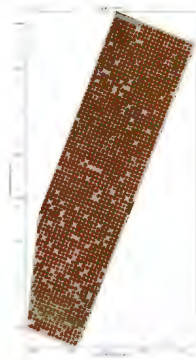
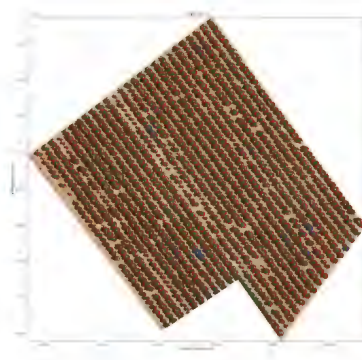


Figure A.14: Bending energy.

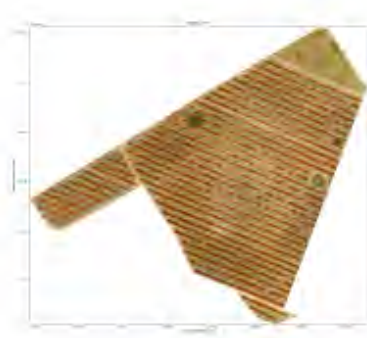
## A.5 Texture Boxplot Outliers



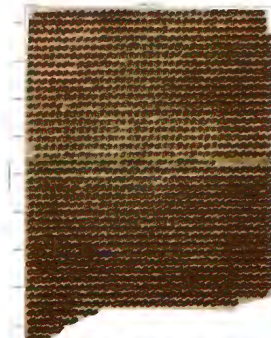
(a) Orchard 1



(b) Orchard 2



(c) Orchard 3

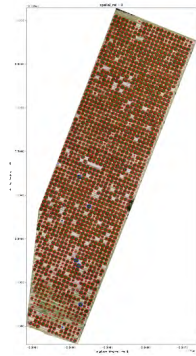


(d) Orchard 4

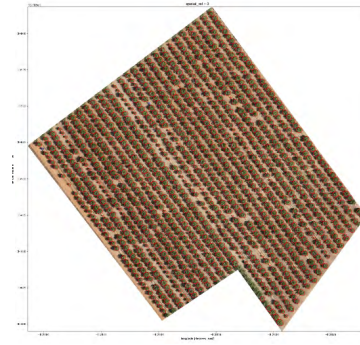


(e) Orchard 5

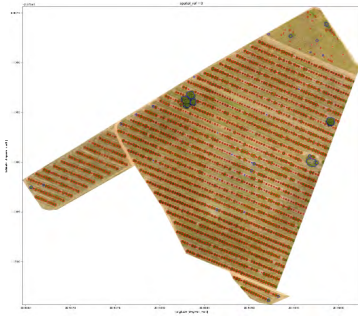
Figure A.15: Illustration of ASM outliers of the first 5 randomly selected training datasets.



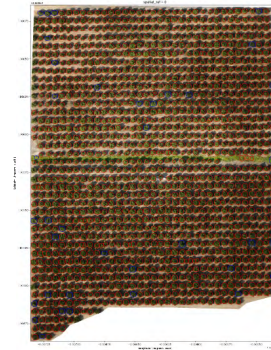
(a) Orchard 1



(b) Orchard 2

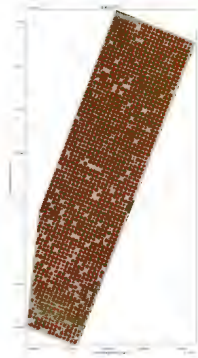


(c) Orchard 3

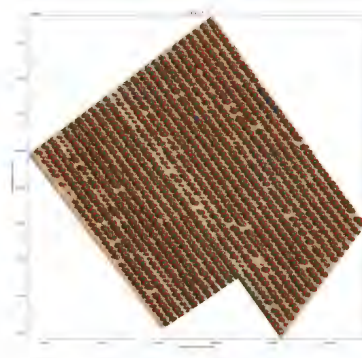


(d) Orchard 4

Figure A.16: Illustration of contrast's left tail outliers of the first 4 randomly selected training datasets.



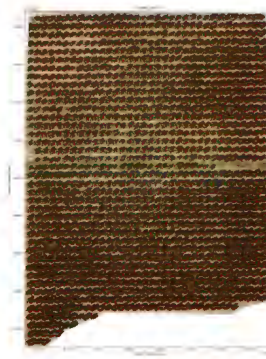
(a) Orchard 1



(b) Orchard 2



(c) Orchard 3

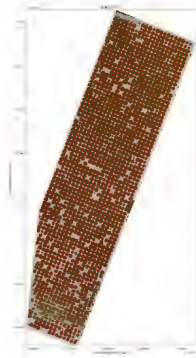


(d) Orchard 4

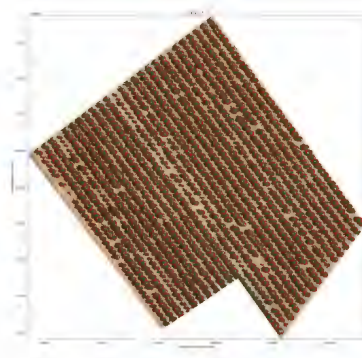


(e) Orchard 5

Figure A.17: Illustration of contrast's right tail outliers of the first 5 randomly selected training datasets.



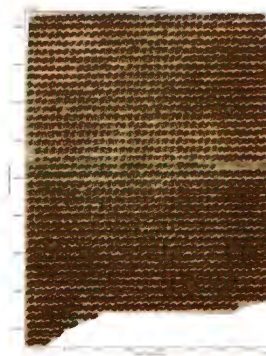
(a) Orchard 1



(b) Orchard 2



(c) Orchard 3



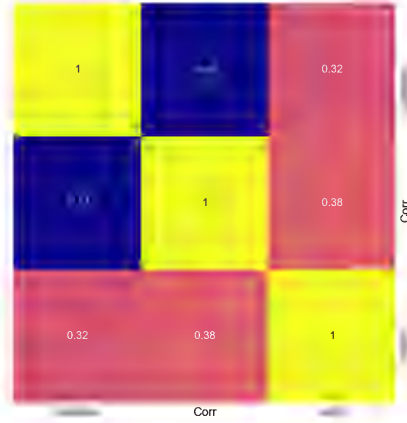
(d) Orchard 4



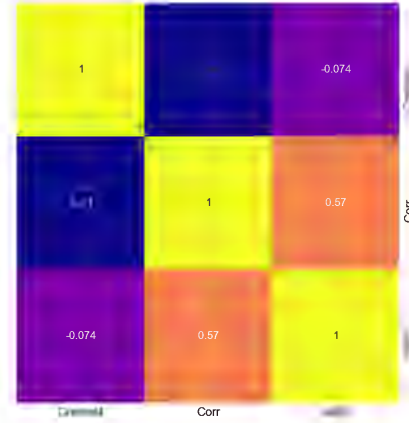
(e) Orchard 5

Figure A.18: Illustration of correlation outliers of the first 5 randomly selected training datasets.

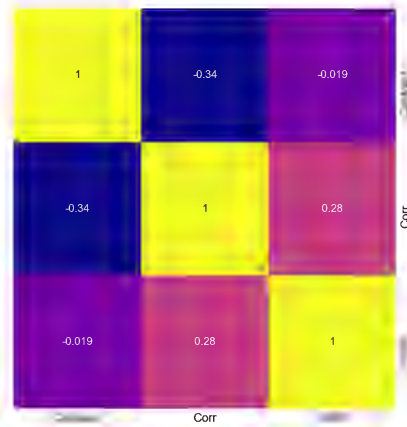
## A.6 Bivariate Analysis



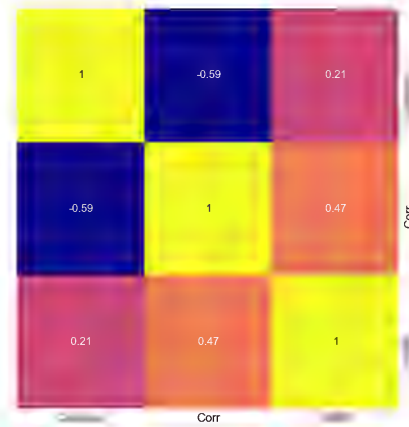
(a) Orchard 1



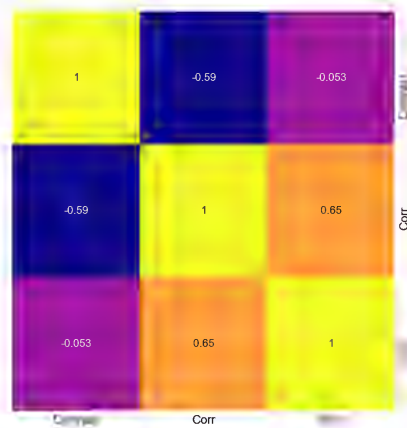
(b) Orchard 2



(c) Orchard 3



(d) Orchard 4



(e) Orchard 5

Figure A.19: Clustermap of texture features of the first 5 randomly selected training datasets.

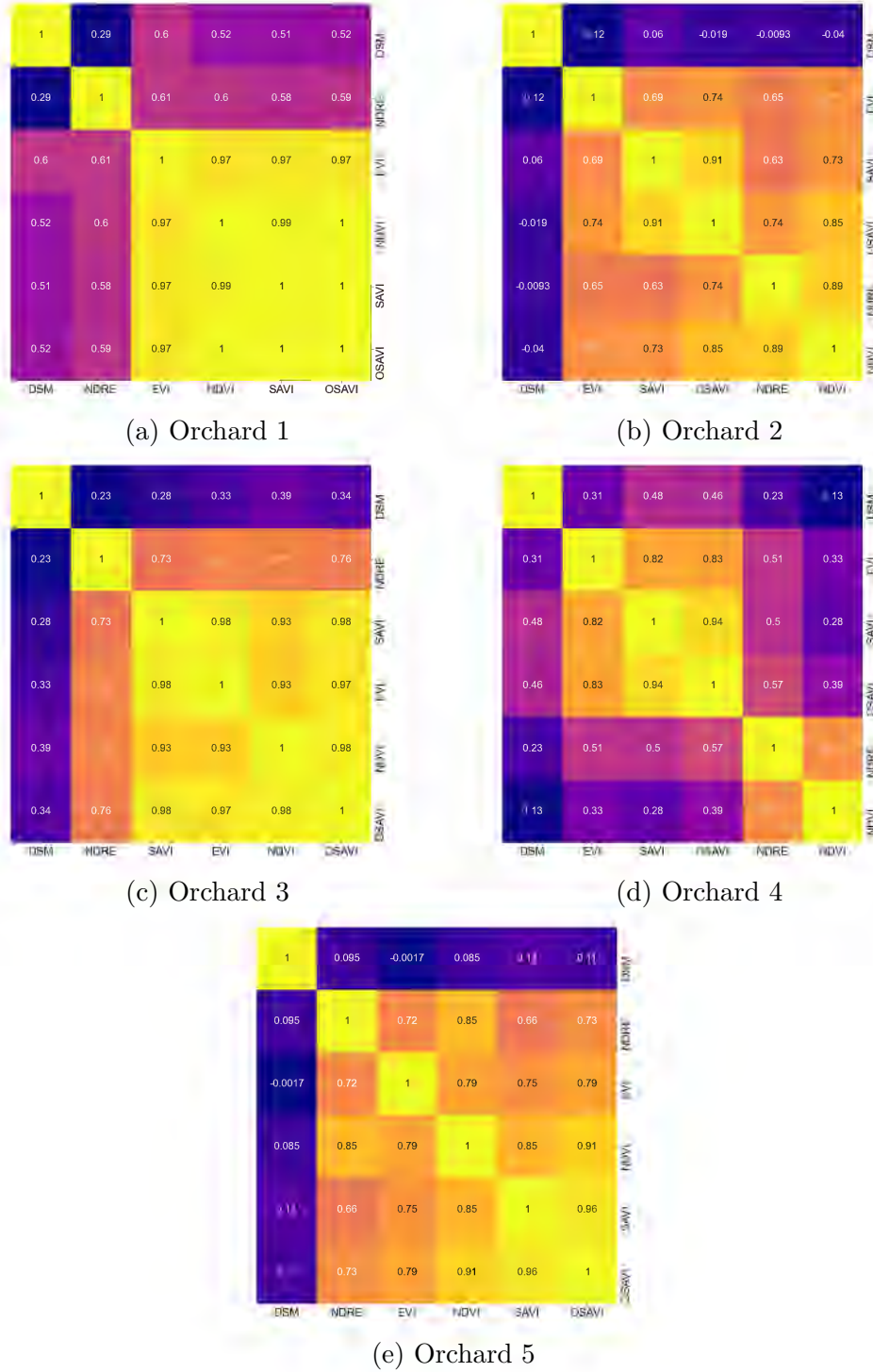
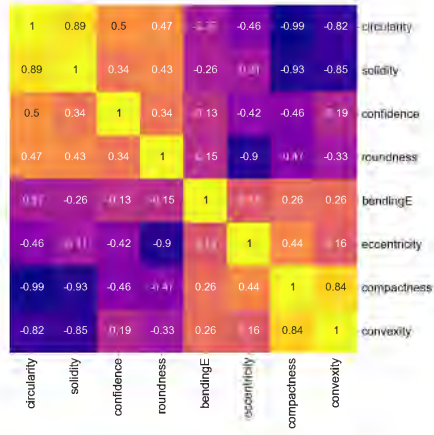
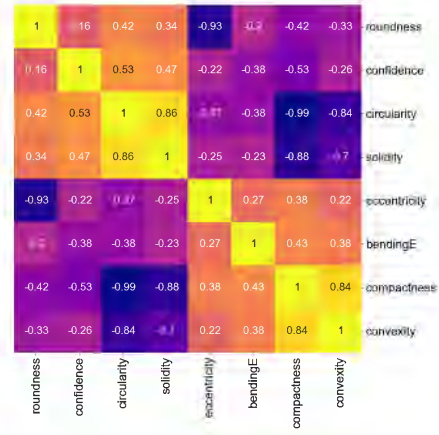


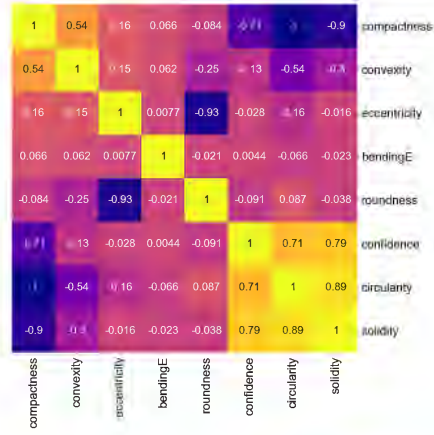
Figure A.20: Clustermap of spectral features of the first 5 randomly selected training datasets.



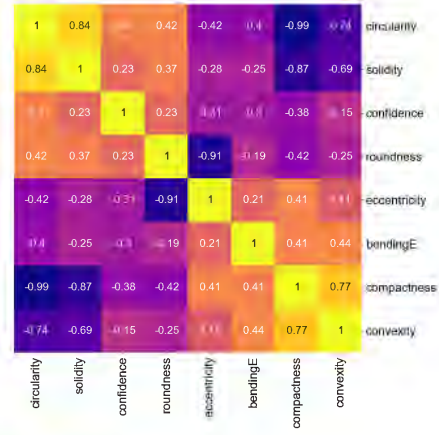
(a) Orchard 1



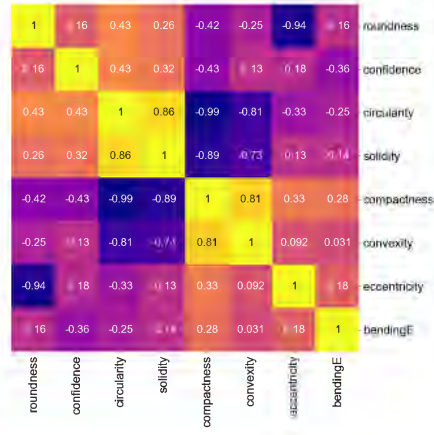
(b) Orchard 2



(c) Orchard 3



(d) Orchard 4



(e) Orchard 5

Figure A.21: Clustermap of shape descriptors of the first 5 randomly selected training datasets.

# Appendix B

## Additional Experimental Results

### B.1 Model Evaluation

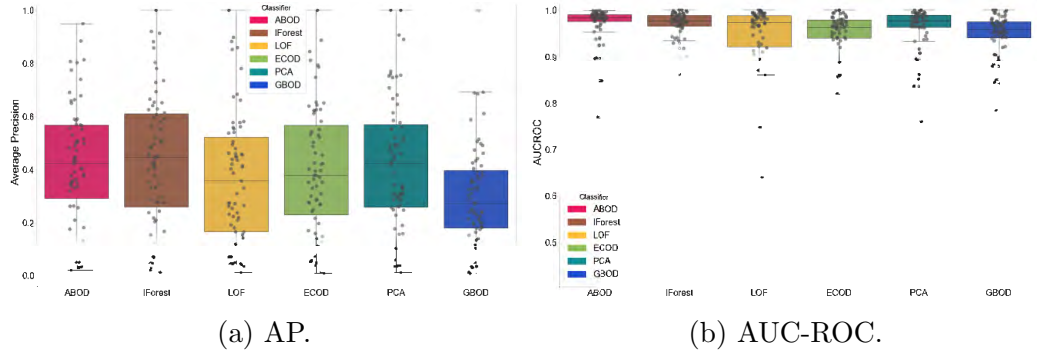


Figure B.1: Summary of performance captured over 70 different orchards.

Table B.1: Average Precision, AP

| Data       | ABOD             | IForest          | LOF              | ECOD             | PCA              | Geary          |
|------------|------------------|------------------|------------------|------------------|------------------|----------------|
| Orchard 1  | <b>0.435 (3)</b> | <b>0.453 (2)</b> | <b>0.481 (1)</b> | 0.381 (4)        | 0.423 (5)        | 0.329 (6)      |
| Orchard 2  | <b>0.021 (1)</b> | <b>0.013 (2)</b> | <b>0.012 (3)</b> | 0.009 (6)        | 0.012 (3)        | 0.008 (5)      |
| Orchard 3  | <b>0.665 (1)</b> | <b>0.616 (3)</b> | 0.071 (6)        | <b>0.624 (2)</b> | 0.48 (4)         | 0.249 (5)      |
| Orchard 4  | <b>0.491 (2)</b> | 0.455 (4)        | <b>0.561 (1)</b> | <b>0.457 (3)</b> | 0.441 (5)        | 0.328 (6)      |
| Orchard 5  | 0.063 (6)        | <b>0.124 (3)</b> | 0.072 (5)        | <b>0.355 (1)</b> | <b>0.179 (2)</b> | 0.087 (4)      |
| Orchard 6  | <b>0.134 (3)</b> | <b>0.17 (1)</b>  | <b>0.145 (2)</b> | 0.129 (4)        | 0.103 (6)        | 0.119 (5)      |
| Orchard 7  | <b>0.45 (2)</b>  | <b>0.667 (1)</b> | <b>0.45 (2)</b>  | 0.278 (6)        | 0.333 (5)        | 0.393 (4)      |
| Orchard 8  | <b>0.093 (1)</b> | <b>0.052 (2)</b> | <b>0.05 (3)</b>  | 0.049 (4)        | 0.039 (5)        | 0.035 (6)      |
| Orchard 9  | 0.804 (3)        | 0.804 (3)        | <b>1.0 (1)</b>   | 0.888 (5)        | <b>1.0 (1)</b>   | <b>1.0 (1)</b> |
| Orchard 10 | <b>0.776 (2)</b> | 0.73 (4)         | <b>0.781 (1)</b> | 0.653 (5)        | <b>0.771 (3)</b> | 0.516 (6)      |

|            |                  |                  |                  |                  |                  |                  |
|------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Orchard 11 | <b>0.568 (1)</b> | <b>0.41 (2)</b>  | <b>0.464 (3)</b> | 0.332 (4)        | 0.346 (5)        | 0.214 (6)        |
| Orchard 12 | <b>0.386 (3)</b> | <b>0.552 (2)</b> | 0.257 (5)        | <b>0.625 (1)</b> | 0.389 (4)        | 0.184 (6)        |
| Orchard 13 | 0.34 (5)         | <b>0.523 (1)</b> | 0.235 (6)        | <b>0.489 (3)</b> | <b>0.495 (2)</b> | 0.508 (4)        |
| Orchard 14 | <b>0.434 (2)</b> | 0.428 (4)        | 0.168 (6)        | 0.427 (5)        | <b>0.489 (1)</b> | <b>0.406 (3)</b> |
| Orchard 15 | <b>0.885 (3)</b> | <b>0.914 (1)</b> | 0.883 (4)        | 0.812 (5)        | <b>0.907 (2)</b> | 0.613 (6)        |
| Orchard 16 | <b>0.501 (2)</b> | 0.379 (4)        | <b>0.508 (1)</b> | 0.303 (5)        | <b>0.453 (3)</b> | 0.24 (6)         |
| Orchard 17 | 0.652 (5)        | <b>0.656 (2)</b> | <b>0.86 (1)</b>  | 0.641 (4)        | <b>0.683 (3)</b> | 0.444 (6)        |
| Orchard 18 | <b>0.353 (2)</b> | 0.246 (4)        | <b>0.422 (1)</b> | 0.15 (6)         | <b>0.305 (3)</b> | 0.218 (5)        |
| Orchard 19 | <b>0.808 (2)</b> | 0.481 (5)        | <b>0.899 (1)</b> | 0.536 (4)        | <b>0.752 (3)</b> | 0.69 (6)         |
| Orchard 20 | 0.339 (5)        | <b>0.628 (1)</b> | 0.035 (6)        | <b>0.553 (3)</b> | <b>0.554 (2)</b> | 0.355 (4)        |
| Orchard 22 | <b>0.36 (2)</b>  | <b>0.418 (1)</b> | <b>0.359 (3)</b> | 0.311 (4)        | 0.293 (5)        | 0.272 (6)        |
| Orchard 23 | <b>0.66 (1)</b>  | <b>0.643 (2)</b> | 0.276 (6)        | 0.544 (4)        | <b>0.669 (3)</b> | 0.59 (5)         |
| Orchard 24 | <b>0.742 (2)</b> | <b>0.781 (1)</b> | 0.087 (6)        | 0.629 (4)        | <b>0.71 (3)</b>  | 0.493 (5)        |
| Orchard 25 | <b>0.051 (3)</b> | <b>0.071 (1)</b> | 0.048 (5)        | <b>0.057 (2)</b> | 0.049 (4)        | 0.026 (6)        |
| Orchard 26 | <b>0.483 (1)</b> | 0.206 (5)        | <b>0.438 (2)</b> | 0.2 (6)          | <b>0.31 (3)</b>  | 0.243 (4)        |
| Orchard 27 | <b>0.95 (1)</b>  | <b>0.921 (2)</b> | 0.222 (6)        | 0.845 (4)        | <b>0.891 (3)</b> | 0.693 (5)        |
| Orchard 28 | <b>0.321 (1)</b> | <b>0.277 (2)</b> | 0.194 (4)        | <b>0.27 (3)</b>  | 0.192 (5)        | 0.05 (6)         |
| Orchard 29 | <b>0.814 (2)</b> | 0.583 (4)        | <b>0.853 (1)</b> | 0.524 (5)        | <b>0.648 (3)</b> | 0.393 (6)        |
| Orchard 30 | <b>0.593 (1)</b> | 0.156 (5)        | <b>0.405 (2)</b> | 0.115 (6)        | <b>0.305 (3)</b> | 0.201 (4)        |
| Orchard 31 | 0.474 (5)        | <b>0.63 (3)</b>  | <b>0.663 (2)</b> | <b>0.787 (1)</b> | 0.616 (4)        | 0.435 (6)        |
| Orchard 32 | <b>0.533 (2)</b> | 0.505 (4)        | <b>0.579 (1)</b> | 0.374 (6)        | <b>0.511 (3)</b> | 0.365 (5)        |
| Orchard 33 | <b>0.259 (3)</b> | <b>0.28 (2)</b>  | 0.276 (4)        | 0.251 (5)        | <b>0.287 (1)</b> | 0.187 (6)        |
| Orchard 34 | <b>0.587 (1)</b> | <b>0.46 (3)</b>  | 0.121 (6)        | 0.399 (4)        | <b>0.493 (2)</b> | 0.356 (5)        |
| Orchard 35 | <b>0.553 (1)</b> | <b>0.447 (3)</b> | 0.24 (6)         | 0.366 (4)        | <b>0.463 (2)</b> | 0.304 (5)        |
| Orchard 36 | 0.031 (5)        | 0.031 (5)        | <b>0.045 (2)</b> | <b>0.073 (1)</b> | <b>0.037 (3)</b> | 0.03 (6)         |
| Orchard 37 | 0.511 (4)        | <b>0.609 (2)</b> | 0.497 (5)        | <b>0.613 (1)</b> | <b>0.523 (3)</b> | 0.328 (6)        |
| Orchard 38 | 0.184 (6)        | <b>0.597 (2)</b> | 0.157 (5)        | <b>0.628 (1)</b> | <b>0.339 (3)</b> | 0.241 (4)        |
| Orchard 39 | <b>0.446 (3)</b> | 0.45 (4)         | <b>0.451 (2)</b> | 0.425 (5)        | <b>0.493 (1)</b> | 0.312 (6)        |
| Orchard 40 | 0.031 (5)        | <b>0.5 (1)</b>   | 0.03 (6)         | <b>0.143 (2)</b> | <b>0.25 (3)</b>  | 0.143 (2)        |
| Orchard 41 | <b>0.266 (3)</b> | <b>0.411 (1)</b> | 0.265 (4)        | <b>0.355 (2)</b> | 0.161 (6)        | 0.177 (5)        |
| Orchard 42 | 0.378 (4)        | <b>0.431 (1)</b> | <b>0.417 (2)</b> | 0.38 (3)         | <b>0.431 (1)</b> | 0.252 (6)        |
| Orchard 43 | 0.407 (4)        | 0.399 (5)        | <b>0.441 (3)</b> | 0.279 (6)        | <b>0.519 (1)</b> | <b>0.464 (2)</b> |
| Orchard 44 | 0.559 (4)        | <b>0.654 (1)</b> | 0.556 (5)        | <b>0.645 (2)</b> | <b>0.579 (3)</b> | 0.338 (6)        |
| Orchard 45 | <b>0.68 (1)</b>  | <b>0.568 (3)</b> | <b>0.676 (2)</b> | 0.465 (4)        | 0.563 (5)        | 0.427 (6)        |
| Orchard 46 | <b>0.356 (2)</b> | <b>0.332 (3)</b> | <b>0.389 (1)</b> | 0.244 (4)        | 0.269 (5)        | 0.226 (6)        |
| Orchard 47 | <b>0.378 (1)</b> | <b>0.22 (3)</b>  | 0.169 (5)        | 0.219 (4)        | <b>0.231 (2)</b> | 0.155 (6)        |
| Orchard 48 | 0.502 (4)        | <b>0.594 (1)</b> | <b>0.579 (3)</b> | 0.505 (5)        | <b>0.587 (2)</b> | 0.38 (6)         |
| Orchard 50 | <b>0.052 (1)</b> | 0.021 (5)        | <b>0.043 (2)</b> | 0.014 (6)        | <b>0.034 (3)</b> | 0.035 (4)        |
| Orchard 51 | <b>0.692 (3)</b> | <b>0.737 (2)</b> | 0.686 (5)        | 0.656 (6)        | <b>0.754 (1)</b> | 0.688 (4)        |
| Orchard 52 | <b>0.062 (3)</b> | 0.048 (5)        | 0.045 (6)        | 0.053 (4)        | <b>0.111 (2)</b> | <b>0.167 (1)</b> |
| Orchard 53 | 0.212 (5)        | <b>0.281 (3)</b> | 0.226 (4)        | <b>0.287 (2)</b> | <b>0.304 (1)</b> | 0.202 (6)        |
| Orchard 54 | 0.178 (6)        | <b>0.488 (2)</b> | 0.183 (5)        | <b>0.532 (1)</b> | <b>0.467 (3)</b> | 0.294 (4)        |
| Orchard 55 | <b>0.348 (1)</b> | <b>0.335 (3)</b> | 0.267 (4)        | <b>0.34 (2)</b>  | 0.259 (5)        | 0.204 (6)        |

|                |                  |                  |                  |                  |                  |                  |
|----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Orchard 56     | 0.572 (4)        | <b>0.882 (1)</b> | 0.542 (5)        | <b>0.75 (3)</b>  | <b>0.761 (2)</b> | 0.476 (6)        |
| Orchard 57     | <b>0.265 (2)</b> | 0.069 (5)        | <b>0.583 (1)</b> | 0.038 (6)        | 0.098 (4)        | <b>0.104 (3)</b> |
| Orchard 58     | <b>0.592 (3)</b> | <b>0.695 (2)</b> | 0.539 (4)        | 0.584 (5)        | <b>0.749 (1)</b> | 0.378 (6)        |
| Orchard 59     | 0.333 (4)        | <b>1.0 (1)</b>   | 0.333 (4)        | <b>1.0 (1)</b>   | <b>1.0 (1)</b>   | 0.5 (6)          |
| Orchard 60     | <b>0.393 (2)</b> | 0.071 (5)        | <b>0.412 (1)</b> | 0.033 (6)        | <b>0.159 (3)</b> | 0.137 (4)        |
| Orchard 61     | <b>0.406 (2)</b> | 0.24 (4)         | <b>0.457 (1)</b> | 0.162 (5)        | <b>0.312 (3)</b> | 0.119 (6)        |
| Orchard 62     | <b>0.352 (3)</b> | <b>0.354 (2)</b> | 0.274 (5)        | <b>0.397 (1)</b> | 0.29 (4)         | 0.234 (6)        |
| Orchard 63     | <b>0.034 (3)</b> | <b>0.042 (1)</b> | 0.027 (4)        | <b>0.041 (2)</b> | 0.023 (5)        | 0.015 (6)        |
| Orchard 64     | <b>0.57 (1)</b>  | 0.512 (4)        | <b>0.564 (2)</b> | 0.404 (5)        | <b>0.554 (3)</b> | 0.367 (6)        |
| Orchard 65     | 0.08 (6)         | <b>0.212 (3)</b> | 0.066 (5)        | <b>0.422 (1)</b> | <b>0.312 (2)</b> | 0.28 (4)         |
| Orchard 66     | <b>0.366 (1)</b> | <b>0.292 (3)</b> | <b>0.312 (2)</b> | 0.272 (4)        | 0.251 (5)        | 0.197 (6)        |
| Orchard 67     | <b>0.424 (2)</b> | <b>0.299 (3)</b> | <b>0.46 (1)</b>  | 0.173 (6)        | 0.263 (4)        | 0.181 (5)        |
| Orchard 68     | <b>0.506 (1)</b> | <b>0.35 (2)</b>  | 0.212 (5)        | 0.291 (4)        | <b>0.327 (3)</b> | 0.243 (6)        |
| Orchard 69     | 0.517 (4)        | <b>0.859 (1)</b> | 0.462 (5)        | <b>0.794 (2)</b> | <b>0.757 (3)</b> | 0.417 (6)        |
| <i>Average</i> | <b>0.422 (3)</b> | <b>0.437 (1)</b> | 0.365 (5)        | 0.397 (4)        | <b>0.423 (2)</b> | 0.302 (6)        |
| <i>p-value</i> |                  |                  |                  |                  |                  | -                |

Table B.2: Area Under the Curve of the Receiver Operating Curve, AUC-ROC

| Data       | ABOD             | IForest          | LOF              | ECOD             | PCA              | Geary            |
|------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Orchard 1  | <b>0.988 (1)</b> | 0.984 (3)        | <b>0.988 (1)</b> | 0.970 (5)        | <b>0.986 (2)</b> | 0.976 (4)        |
| Orchard 2  | <b>0.939 (1)</b> | <b>0.902 (2)</b> | <b>0.897 (3)</b> | 0.840 (6)        | 0.895 (4)        | 0.842 (5)        |
| Orchard 3  | <b>0.991 (1)</b> | <b>0.984 (2)</b> | 0.641 (6)        | <b>0.981 (3)</b> | 0.977 (4)        | 0.960 (5)        |
| Orchard 4  | <b>0.980 (2)</b> | <b>0.971 (3)</b> | <b>0.984 (1)</b> | 0.942 (6)        | 0.966 (4)        | 0.944 (5)        |
| Orchard 5  | <b>0.849 (1)</b> | <b>0.817 (3)</b> | 0.749 (5)        | <b>0.818 (2)</b> | 0.838 (4)        | 0.680 (6)        |
| Orchard 6  | <b>0.900 (1)</b> | <b>0.864 (3)</b> | 0.862 (4)        | 0.861 (5)        | 0.853 (6)        | <b>0.895 (2)</b> |
| Orchard 7  | <b>0.994 (1)</b> | <b>0.994 (1)</b> | <b>0.994 (1)</b> | 0.985 (6)        | 0.990 (4)        | 0.990 (4)        |
| Orchard 8  | <b>0.982 (1)</b> | <b>0.962 (3)</b> | <b>0.964 (2)</b> | 0.952 (4)        | 0.939 (5)        | 0.885 (6)        |
| Orchard 9  | 0.999 (5)        | 0.999 (5)        | <b>1.000 (1)</b> | <b>1.000 (1)</b> | <b>1.000 (1)</b> | <b>1.000 (1)</b> |
| Orchard 10 | <b>0.995 (1)</b> | 0.994 (4)        | <b>0.995 (1)</b> | 0.987 (5)        | <b>0.995 (1)</b> | 0.987 (5)        |
| Orchard 11 | <b>0.983 (1)</b> | <b>0.973 (2)</b> | <b>0.973 (2)</b> | 0.965 (4)        | 0.966 (5)        | 0.941 (6)        |
| Orchard 12 | <b>0.991 (3)</b> | <b>0.995 (1)</b> | 0.978 (5)        | <b>0.995 (1)</b> | <b>0.991 (3)</b> | 0.975 (6)        |
| Orchard 13 | 0.972 (4)        | <b>0.982 (1)</b> | 0.969 (5)        | <b>0.974 (3)</b> | <b>0.982 (1)</b> | 0.968 (6)        |
| Orchard 14 | <b>0.994 (1)</b> | <b>0.988 (3)</b> | 0.925 (6)        | 0.979 (5)        | <b>0.990 (2)</b> | 0.979 (4)        |
| Orchard 15 | <b>0.993 (1)</b> | 0.991 (4)        | <b>0.993 (1)</b> | 0.976 (5)        | <b>0.992 (3)</b> | 0.969 (6)        |
| Orchard 16 | <b>0.993 (2)</b> | 0.989 (4)        | <b>0.994 (1)</b> | 0.969 (6)        | <b>0.993 (3)</b> | 0.977 (5)        |
| Orchard 17 | <b>0.991 (2)</b> | 0.980 (4)        | <b>0.994 (1)</b> | 0.975 (5)        | <b>0.974 (3)</b> | 0.922 (6)        |
| Orchard 18 | <b>0.982 (2)</b> | 0.966 (4)        | <b>0.984 (1)</b> | 0.938 (6)        | <b>0.970 (3)</b> | 0.944 (5)        |
| Orchard 19 | <b>0.999 (1)</b> | 0.994 (4)        | <b>0.999 (1)</b> | 0.991 (6)        | <b>0.998 (3)</b> | 0.994 (4)        |
| Orchard 20 | 0.991 (4)        | <b>0.995 (1)</b> | 0.579 (6)        | <b>0.993 (3)</b> | <b>0.995 (1)</b> | 0.987 (5)        |
| Orchard 22 | <b>0.977 (1)</b> | <b>0.964 (3)</b> | <b>0.977 (1)</b> | 0.949 (4)        | 0.889 (6)        | 0.924 (5)        |
| Orchard 23 | <b>0.985 (1)</b> | <b>0.980 (2)</b> | 0.932 (6)        | 0.945 (5)        | <b>0.965 (3)</b> | 0.973 (4)        |

|            |                  |                  |                  |                  |                  |                  |
|------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Orchard 24 | <b>0.992 (1)</b> | <b>0.987 (2)</b> | 0.433 (6)        | 0.963 (5)        | <b>0.980 (3)</b> | 0.972 (4)        |
| Orchard 25 | <b>0.899 (2)</b> | <b>0.920 (1)</b> | 0.889 (4)        | 0.889 (4)        | <b>0.899 (2)</b> | 0.786 (6)        |
| Orchard 26 | <b>0.986 (1)</b> | 0.942 (4)        | <b>0.987 (1)</b> | 0.858 (6)        | <b>0.953 (3)</b> | 0.946 (5)        |
| Orchard 27 | <b>0.997 (1)</b> | <b>0.991 (2)</b> | 0.499 (6)        | 0.976 (4)        | <b>0.984 (3)</b> | 0.948 (5)        |
| Orchard 28 | <b>0.988 (1)</b> | <b>0.967 (2)</b> | <b>0.967 (2)</b> | 0.954 (5)        | 0.964 (4)        | 0.880 (6)        |
| Orchard 29 | <b>0.995 (1)</b> | 0.971 (4)        | <b>0.995 (1)</b> | 0.957 (5)        | <b>0.987 (3)</b> | 0.957 (5)        |
| Orchard 30 | <b>0.991 (1)</b> | 0.973 (4)        | 0.972 (5)        | 0.956 (6)        | <b>0.987 (2)</b> | <b>0.976 (3)</b> |
| Orchard 31 | 0.985 (5)        | <b>0.993 (2)</b> | <b>0.993 (2)</b> | <b>0.994 (1)</b> | <b>0.993 (2)</b> | 0.983 (6)        |
| Orchard 32 | <b>0.983 (3)</b> | 0.982 (4)        | <b>0.985 (1)</b> | 0.964 (5)        | <b>0.985 (1)</b> | 0.967 (6)        |
| Orchard 33 | 0.955 (4)        | <b>0.956 (2)</b> | <b>0.956 (2)</b> | 0.943 (5)        | <b>0.966 (1)</b> | 0.952 (6)        |
| Orchard 34 | <b>0.988 (1)</b> | <b>0.977 (2)</b> | 0.666 (6)        | 0.964 (4)        | <b>0.978 (3)</b> | 0.953 (5)        |
| Orchard 35 | <b>0.984 (1)</b> | <b>0.976 (2)</b> | 0.915 (6)        | 0.951 (5)        | <b>0.976 (2)</b> | 0.955 (4)        |
| Orchard 36 | <b>0.860 (2)</b> | <b>0.835 (3)</b> | <b>0.912 (1)</b> | 0.791 (4)        | 0.761 (6)        | 0.766 (5)        |
| Orchard 37 | <b>0.990 (1)</b> | <b>0.990 (1)</b> | <b>0.988 (3)</b> | 0.981 (5)        | 0.986 (4)        | 0.967 (6)        |
| Orchard 38 | 0.984 (4)        | <b>0.993 (1)</b> | 0.981 (6)        | <b>0.986 (3)</b> | <b>0.992 (2)</b> | 0.981 (5)        |
| Orchard 39 | <b>0.992 (1)</b> | 0.988 (4)        | <b>0.991 (2)</b> | 0.981 (5)        | <b>0.989 (3)</b> | 0.979 (6)        |
| Orchard 40 | 0.968 (6)        | <b>0.999 (1)</b> | 0.967 (7)        | <b>0.994 (3)</b> | <b>0.997 (2)</b> | <b>0.994 (3)</b> |
| Orchard 41 | 0.771 (4)        | <b>0.783 (3)</b> | 0.716 (5)        | 0.710 (6)        | <b>0.837 (2)</b> | <b>0.955 (1)</b> |
| Orchard 42 | 0.851 (4)        | <b>0.862 (3)</b> | <b>0.871 (1)</b> | 0.822 (6)        | <b>0.863 (2)</b> | 0.846 (5)        |
| Orchard 43 | <b>0.991 (3)</b> | 0.990 (4)        | <b>0.993 (2)</b> | 0.969 (6)        | <b>0.994 (1)</b> | <b>0.991 (3)</b> |
| Orchard 44 | <b>0.965 (2)</b> | <b>0.968 (1)</b> | <b>0.965 (2)</b> | 0.961 (4)        | 0.964 (5)        | 0.907 (6)        |
| Orchard 45 | <b>0.989 (1)</b> | <b>0.975 (3)</b> | <b>0.990 (1)</b> | 0.941 (6)        | <b>0.975 (3)</b> | 0.969 (5)        |
| Orchard 46 | <b>0.986 (2)</b> | <b>0.983 (3)</b> | <b>0.988 (1)</b> | 0.964 (6)        | 0.981 (4)        | 0.973 (5)        |
| Orchard 47 | <b>0.983 (1)</b> | <b>0.967 (2)</b> | 0.942 (5)        | 0.943 (4)        | <b>0.966 (3)</b> | 0.943 (4)        |
| Orchard 48 | 0.977 (4)        | <b>0.984 (1)</b> | <b>0.982 (3)</b> | 0.973 (5)        | <b>0.984 (1)</b> | 0.953 (6)        |
| Orchard 50 | <b>0.942 (2)</b> | <b>0.931 (3)</b> | 0.660 (6)        | 0.861 (5)        | 0.919 (4)        | <b>0.964 (1)</b> |
| Orchard 51 | 0.976 (4)        | <b>0.977 (3)</b> | <b>0.980 (1)</b> | 0.968 (6)        | <b>0.978 (2)</b> | <b>0.977 (3)</b> |
| Orchard 52 | <b>0.987 (3)</b> | 0.983 (5)        | 0.982 (6)        | 0.985 (4)        | <b>0.993 (2)</b> | <b>0.996 (1)</b> |
| Orchard 53 | 0.926 (4)        | <b>0.939 (2)</b> | <b>0.928 (3)</b> | 0.930 (6)        | <b>0.946 (1)</b> | 0.879 (5)        |
| Orchard 54 | 0.965 (4)        | <b>0.975 (1)</b> | 0.921 (6)        | <b>0.971 (3)</b> | <b>0.974 (2)</b> | 0.968 (5)        |
| Orchard 55 | <b>0.982 (1)</b> | <b>0.979 (2)</b> | 0.968 (5)        | 0.970 (4)        | <b>0.978 (3)</b> | 0.966 (6)        |
| Orchard 56 | <b>0.990 (3)</b> | <b>0.996 (1)</b> | 0.989 (4)        | 0.980 (6)        | <b>0.994 (2)</b> | 0.976 (5)        |
| Orchard 57 | <b>0.984 (2)</b> | 0.944 (4)        | <b>0.989 (1)</b> | 0.874 (6)        | <b>0.963 (3)</b> | 0.943 (5)        |
| Orchard 58 | 0.999 (3)        | <b>1.000 (1)</b> | 0.999 (3)        | 0.999 (3)        | <b>1.000 (1)</b> | 0.999 (3)        |
| Orchard 59 | 0.998 (5)        | <b>1.000 (1)</b> | 0.998 (5)        | <b>1.000 (1)</b> | <b>1.000 (1)</b> | <b>1.000 (1)</b> |
| Orchard 60 | <b>0.984 (1)</b> | <b>0.936 (3)</b> | <b>0.971 (2)</b> | 0.833 (6)        | 0.935 (4)        | 0.901 (5)        |
| Orchard 61 | <b>0.989 (2)</b> | 0.974 (4)        | <b>0.991 (1)</b> | 0.936 (5)        | <b>0.983 (3)</b> | 0.936 (5)        |
| Orchard 62 | <b>0.979 (1)</b> | <b>0.977 (2)</b> | 0.974 (4)        | <b>0.976 (3)</b> | 0.974 (4)        | 0.955 (6)        |
| Orchard 63 | <b>0.960 (3)</b> | <b>0.967 (1)</b> | 0.948 (4)        | <b>0.967 (1)</b> | 0.938 (5)        | 0.906 (6)        |
| Orchard 64 | <b>0.985 (2)</b> | 0.982 (4)        | <b>0.987 (1)</b> | 0.962 (5)        | <b>0.985 (2)</b> | 0.960 (6)        |
| Orchard 65 | <b>0.927 (2)</b> | <b>0.941 (1)</b> | 0.908 (4)        | <b>0.922 (3)</b> | 0.910 (5)        | 0.852 (6)        |
| Orchard 66 | <b>0.978 (1)</b> | 0.970 (4)        | <b>0.972 (2)</b> | 0.964 (6)        | <b>0.971 (3)</b> | 0.953 (5)        |
| Orchard 67 | <b>0.984 (1)</b> | <b>0.975 (3)</b> | <b>0.981 (2)</b> | 0.940 (6)        | 0.966 (4)        | 0.952 (5)        |

|                |                  |                  |           |                  |                  |           |
|----------------|------------------|------------------|-----------|------------------|------------------|-----------|
| Orchard 68     | <b>0.983 (1)</b> | <b>0.983 (1)</b> | 0.585 (6) | 0.974 (4)        | <b>0.976 (3)</b> | 0.974 (5) |
| Orchard 69     | <b>0.984 (3)</b> | <b>0.989 (1)</b> | 0.975 (5) | <b>0.984 (3)</b> | <b>0.987 (2)</b> | 0.976 (6) |
| <i>Average</i> | <b>0.970 (1)</b> | <b>0.964 (2)</b> | 0.919 (6) | 0.946 (4)        | <b>0.962 (3)</b> | 0.943 (5) |
| <i>p-value</i> |                  |                  |           |                  |                  | -         |

Table B.3: Time, seconds

| Data       | ABOD             | IForest          | LOF              | ECOD             | PCA              | Geary      |
|------------|------------------|------------------|------------------|------------------|------------------|------------|
| Orchard 0  | 1.830 (5)        | 0.337 (4)        | <b>0.048 (3)</b> | <b>0.093 (2)</b> | <b>0.027 (1)</b> | 2.658 (6)  |
| Orchard 1  | 43.609 (6)       | 0.120 (4)        | <b>0.046 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 3.258 (5)  |
| Orchard 2  | 8.628 (6)        | 0.158 (4)        | <b>0.083 (3)</b> | <b>0.012 (2)</b> | <b>0.000 (1)</b> | 4.233 (5)  |
| Orchard 3  | 37.946 (6)       | 0.384 (4)        | <b>0.063 (3)</b> | <b>0.004 (2)</b> | <b>0.000 (1)</b> | 2.600 (5)  |
| Orchard 4  | 0.288 (5)        | 0.160 (4)        | <b>0.068 (3)</b> | <b>0.007 (2)</b> | <b>0.005 (1)</b> | 2.143 (6)  |
| Orchard 5  | 0.491 (5)        | 0.594 (6)        | <b>0.030 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 1.788 (4)  |
| Orchard 6  | 8.508 (6)        | 0.696 (5)        | <b>0.007 (3)</b> | <b>0.004 (2)</b> | <b>0.000 (1)</b> | 0.598 (4)  |
| Orchard 7  | 7.131 (6)        | 0.489 (4)        | <b>0.081 (3)</b> | <b>0.005 (1)</b> | <b>0.005 (2)</b> | 2.867 (5)  |
| Orchard 8  | <b>0.164 (4)</b> | 0.306 (5)        | <b>0.037 (3)</b> | <b>0.000 (1)</b> | 0.002 (2)        | 2.140 (6)  |
| Orchard 9  | 9.559 (6)        | 0.143 (4)        | <b>0.048 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 2.292 (5)  |
| Orchard 10 | 3.736 (6)        | 1.031 (5)        | <b>0.010 (3)</b> | <b>0.002 (1)</b> | <b>0.003 (2)</b> | 0.802 (4)  |
| Orchard 11 | 18.591 (6)       | 0.906 (4)        | <b>0.098 (3)</b> | <b>0.007 (2)</b> | <b>0.000 (1)</b> | 4.332 (5)  |
| Orchard 12 | 21.216 (6)       | 0.108 (4)        | <b>0.026 (3)</b> | <b>0.004 (2)</b> | <b>0.000 (1)</b> | 1.580 (5)  |
| Orchard 13 | 43.505 (6)       | 1.128 (4)        | <b>0.424 (3)</b> | <b>0.016 (2)</b> | <b>0.005 (1)</b> | 10.784 (5) |
| Orchard 14 | 31.617 (6)       | 1.191 (4)        | <b>0.068 (3)</b> | <b>0.004 (2)</b> | <b>0.000 (1)</b> | 2.450 (5)  |
| Orchard 15 | 88.146 (6)       | <b>0.502 (3)</b> | 0.347 (4)        | <b>0.008 (1)</b> | <b>0.004 (2)</b> | 8.150 (5)  |
| Orchard 16 | 8.034 (6)        | 0.771 (4)        | <b>0.023 (3)</b> | <b>0.004 (2)</b> | <b>0.002 (1)</b> | 1.674 (5)  |
| Orchard 17 | 25.356 (6)       | 0.367 (4)        | <b>0.054 (3)</b> | <b>0.011 (2)</b> | <b>0.000 (1)</b> | 1.993 (5)  |
| Orchard 18 | 24.978 (6)       | 2.060 (4)        | <b>0.047 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 1.792 (5)  |
| Orchard 19 | 83.813 (6)       | <b>1.456 (3)</b> | 1.067 (4)        | <b>0.020 (2)</b> | <b>0.007 (1)</b> | 22.924 (5) |
| Orchard 20 | 7.363 (6)        | 0.322 (4)        | <b>0.000 (1)</b> | <b>0.002 (3)</b> | <b>0.000 (1)</b> | 0.576 (5)  |
| Orchard 21 | 7.234 (5)        | <b>0.169 (3)</b> | 0.226 (4)        | <b>0.000 (1)</b> | <b>0.007 (2)</b> | 5.426 (6)  |
| Orchard 22 | 86.500 (6)       | 1.204 (4)        | <b>0.165 (3)</b> | <b>0.006 (2)</b> | <b>0.002 (1)</b> | 6.689 (5)  |
| Orchard 23 | 7.503 (6)        | 1.675 (5)        | <b>0.221 (3)</b> | <b>0.002 (1)</b> | <b>0.002 (1)</b> | 0.582 (4)  |
| Orchard 24 | 27.774 (6)       | 1.272 (4)        | <b>0.039 (3)</b> | <b>0.004 (2)</b> | <b>0.000 (1)</b> | 2.019 (5)  |
| Orchard 25 | 95.341 (6)       | <b>0.163 (3)</b> | 0.447 (4)        | <b>0.009 (2)</b> | <b>0.000 (1)</b> | 7.751 (5)  |
| Orchard 26 | 5.810 (6)        | 0.518 (4)        | <b>0.016 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 1.104 (5)  |
| Orchard 27 | 10.650 (6)       | 0.977 (4)        | <b>0.016 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 0.915 (5)  |
| Orchard 28 | <b>0.174 (3)</b> | 0.701 (5)        | <b>0.016 (2)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 1.269 (6)  |
| Orchard 29 | 31.127 (6)       | 0.142 (4)        | <b>0.048 (3)</b> | <b>0.012 (2)</b> | <b>0.005 (1)</b> | 2.181 (5)  |
| Orchard 30 | 1.770 (5)        | 0.231 (4)        | <b>0.016 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 1.367 (6)  |
| Orchard 31 | 26.115 (6)       | 0.132 (4)        | <b>0.054 (3)</b> | <b>0.000 (1)</b> | <b>0.002 (2)</b> | 2.327 (5)  |
| Orchard 32 | <b>1.142 (4)</b> | 0.722 (5)        | <b>0.048 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 3.428 (6)  |
| Orchard 33 | 9.913 (6)        | 1.125 (4)        | <b>0.032 (3)</b> | <b>0.000 (1)</b> | <b>0.000 (1)</b> | 2.291 (5)  |

|                |                  |                  |                   |                   |                   |            |
|----------------|------------------|------------------|-------------------|-------------------|-------------------|------------|
| Orchard 34     | <b>0.359 (4)</b> | <b>0.162 (3)</b> | <b>0.087 (2)</b>  | 0.002 (1)         | 0.002 (1)         | 5.328 (6)  |
| Orchard 35     | 31.967 (6)       | 0.409 (4)        | <b>0.127 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 4.677 (5)  |
| Orchard 36     | 12.100 (6)       | 0.123 (4)        | <b>0.030 (3)</b>  | <b>0.004 (2)</b>  | <b>0.001 (1)</b>  | 3.297 (5)  |
| Orchard 37     | 74.524 (6)       | <b>0.143 (3)</b> | 0.344 (4)         | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 5.815 (5)  |
| Orchard 38     | 12.409 (6)       | 0.108 (4)        | <b>0.020 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.747 (5)  |
| Orchard 39     | 9.211 (6)        | 0.112 (4)        | <b>0.016 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 0.713 (5)  |
| Orchard 40     | <b>0.254 (4)</b> | 0.445 (5)        | <b>0.036 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.917 (6)  |
| Orchard 41     | 3.944 (6)        | 0.443 (5)        | <b>0.013 (3)</b>  | <b>0.005 (2)</b>  | <b>0.002 (1)</b>  | 0.831 (4)  |
| Orchard 42     | 0.412 (5)        | 0.239 (4)        | <b>0.016 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.366 (6)  |
| Orchard 43     | 10.042 (6)       | 0.912 (5)        | <b>0.038 (3)</b>  | <b>0.000 (1)</b>  | <b>0.004 (2)</b>  | 2.359 (4)  |
| Orchard 44     | 8.564 (6)        | 0.412 (4)        | <b>0.158 (3)</b>  | <b>0.016 (2)</b>  | <b>0.000 (1)</b>  | 4.926 (5)  |
| Orchard 45     | <b>0.238 (3)</b> | 0.844 (5)        | <b>0.048 (2)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 3.122 (6)  |
| Orchard 46     | <b>0.058 (3)</b> | 0.111 (4)        | <b>0.032 (2)</b>  | <b>0.000 (1)</b>  | 0.016 (5)         | 1.545 (6)  |
| Orchard 47     | 1.571 (5)        | 1.080 (4)        | <b>0.064 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 2.634 (6)  |
| Orchard 48     | <b>0.124 (3)</b> | 0.296 (4)        | <b>0.069 (2)</b>  | <b>0.000 (1)</b>  | 0.016 (5)         | 2.951 (6)  |
| Orchard 49     | 0.650 (5)        | 0.201 (4)        | <b>0.048 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 2.076 (6)  |
| Orchard 50     | <b>0.032 (3)</b> | 0.476 (5)        | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 0.411 (4)  |
| Orchard 51     | 2.956 (5)        | <b>0.159 (3)</b> | 0.348 (4)         | <b>0.016 (2)</b>  | <b>0.014 (1)</b>  | 8.832 (6)  |
| Orchard 52     | 16.076 (6)       | 0.890 (4)        | <b>0.079 (3)</b>  | <b>0.000 (1)</b>  | <b>0.016 (2)</b>  | 3.708 (5)  |
| Orchard 53     | <b>0.713 (4)</b> | 1.644 (5)        | <b>0.016 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.977 (6)  |
| Orchard 54     | 0.480 (5)        | 0.154 (4)        | <b>0.067 (3)</b>  | <b>0.000 (1)</b>  | <b>0.002 (2)</b>  | 3.516 (6)  |
| Orchard 55     | 155.844 (6)      | <b>0.571 (3)</b> | 0.592 (4)         | <b>0.012 (2)</b>  | <b>0.004 (1)</b>  | 12.844 (5) |
| Orchard 56     | 10.178 (6)       | 2.309 (5)        | <b>0.044 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 2.263 (4)  |
| Orchard 57     | 14.710 (6)       | 0.853 (4)        | <b>0.060 (3)</b>  | <b>0.008 (2)</b>  | <b>0.000 (1)</b>  | 3.286 (5)  |
| Orchard 58     | 4.196 (6)        | 0.241 (4)        | <b>0.030 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.496 (5)  |
| Orchard 59     | 0.876 (5)        | 0.115 (4)        | <b>0.011 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 0.548 (6)  |
| Orchard 60     | 24.644 (6)       | 0.933 (4)        | <b>0.055 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.797 (5)  |
| Orchard 61     | 24.437 (6)       | 0.445 (4)        | <b>0.044 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 1.899 (5)  |
| Orchard 62     | 8.789 (6)        | 0.451 (5)        | <b>0.009 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 0.673 (4)  |
| Orchard 63     | 5.253 (6)        | 1.355 (4)        | <b>0.146 (3)</b>  | <b>0.014 (2)</b>  | <b>0.000 (1)</b>  | 3.900 (5)  |
| Orchard 64     | 10.899 (6)       | 0.803 (4)        | <b>0.049 (3)</b>  | <b>0.004 (2)</b>  | <b>0.000 (1)</b>  | 2.499 (5)  |
| Orchard 65     | <b>0.468 (3)</b> | 0.369 (4)        | <b>0.236 (2)</b>  | 0.011 (5)         | <b>0.002 (1)</b>  | 8.523 (6)  |
| Orchard 66     | 18.254 (6)       | 0.866 (4)        | <b>0.095 (3)</b>  | <b>0.000 (1)</b>  | <b>0.000 (1)</b>  | 4.104 (5)  |
| <i>Average</i> | 18.2348 (6)      | 0.6047 (4)       | <b>0.1142 (3)</b> | <b>0.0048 (2)</b> | <b>0.0029 (1)</b> | 3.1266 (5) |

## B.2 Out-of-sample

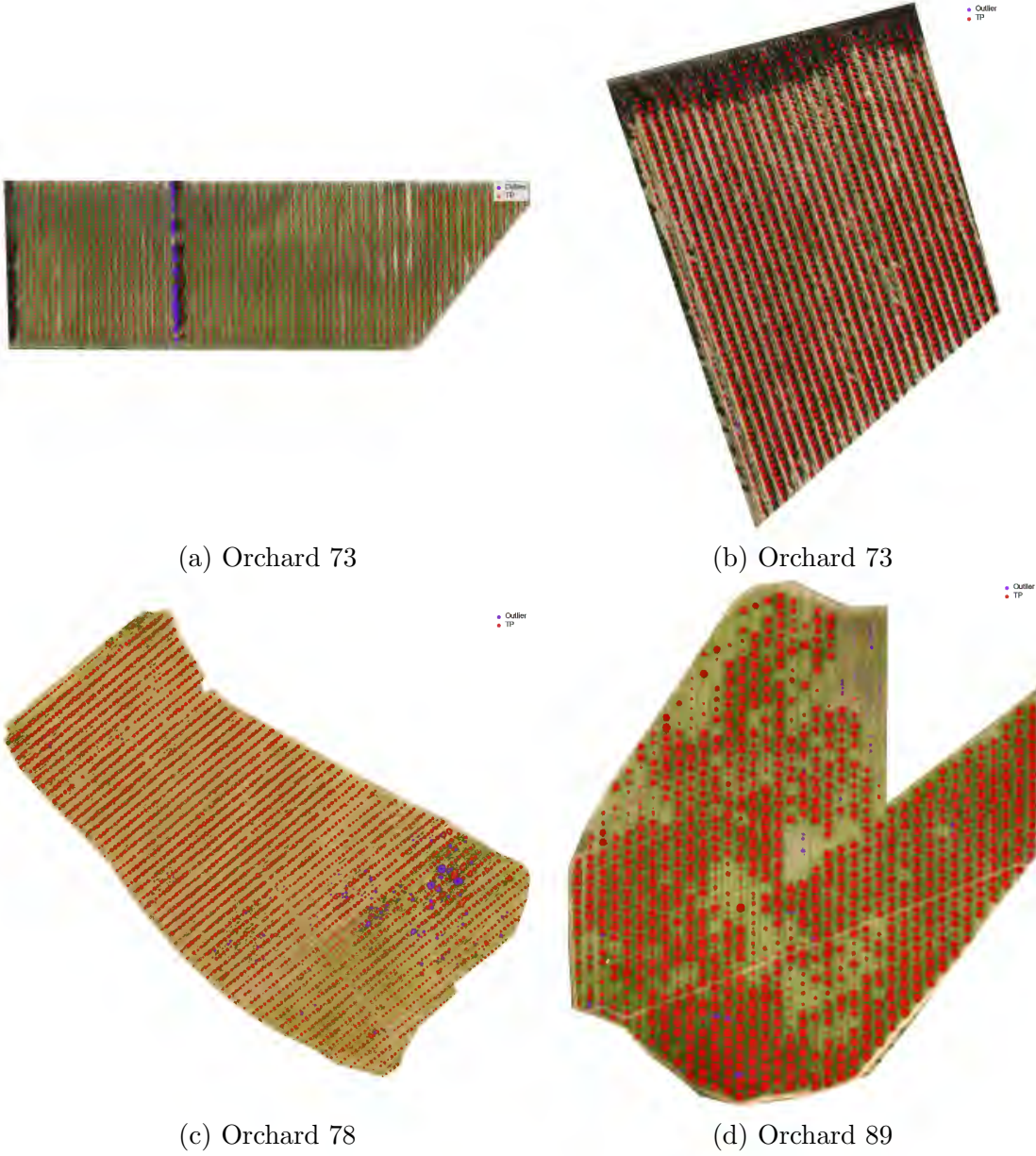


Figure B.2: Reference data of orchards observed OOS.

## B.3 A Look at GBOD

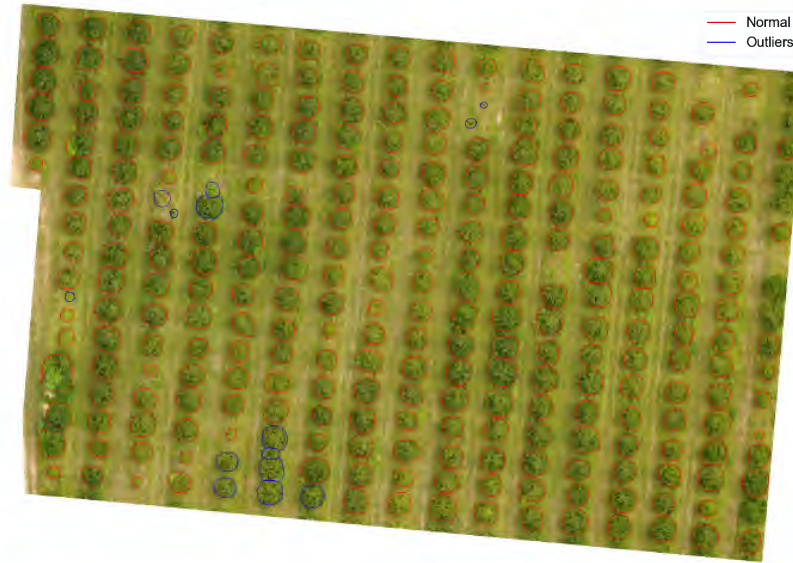


Figure B.3: Orchard 62, depicting anomalous delineations captured by the Geary-based method