

## Article

# Comparison of Selected Dimensionality Reduction Methods for Detection of Root-Knot Nematode Infestations in Potato Tubers Using Hyperspectral Imaging

Janez Lapajne, Matej Knapič and Uroš Žibrat \* Plant Protection Department, Agricultural Institute of Slovenia, 1000 Ljubljana, Slovenia;  
janez.lapajne@kis.si (J.L.); matej.knapic@kis.si (M.K.)

\* Correspondence: uros.zibrat@kis.si; Tel.: +386-1-280-5131

**Abstract:** Hyperspectral imaging is a popular tool used for non-invasive plant disease detection. Data acquired with it usually consist of many correlated features; hence most of the acquired information is redundant. Dimensionality reduction methods are used to transform the data sets from high-dimensional, to low-dimensional (in this study to one or a few features). We have chosen six dimensionality reduction methods (partial least squares, linear discriminant analysis, principal component analysis, RandomForest, ReliefF, and Extreme gradient boosting) and tested their efficacy on a hyperspectral data set of potato tubers. The extracted or selected features were pipelined to support vector machine classifier and evaluated. Tubers were divided into two groups, healthy and infested with *Meloidogyne luci*. The results show that all dimensionality reduction methods enabled successful identification of inoculated tubers. The best and most consistent results were obtained using linear discriminant analysis, with 100% accuracy in both potato tuber inside and outside images. Classification success was generally higher in the outside data set, than in the inside. Nevertheless, accuracy was in all cases above 0.6.

**Keywords:** hyperspectral imaging; dimensionality reduction; LDA; PLS; PCA; RandomForest; ReliefF; XGB; *Meloidogyne*; *Solanum tuberosum*



**Citation:** Lapajne, J.; Knapič, M.; Žibrat, U. Comparison of Selected Dimensionality Reduction Methods for Detection of Root-Knot Nematode Infestations in Potato Tubers Using Hyperspectral Imaging. *Sensors* **2022**, *22*, 367. <https://doi.org/10.3390/s22010367>

Academic Editors: Jiyul Chang and Sigfredo Fuentes

Received: 30 November 2021

Accepted: 27 December 2021

Published: 4 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Quarantine pests are of major importance for agriculture and the food industry, and are being officially monitored and controlled [1]. Among these, root-knot nematodes (RKN) of the genus *Meloidogyne* present the most destructive group of plant-parasitic nematodes. They can infest a broad range of host plants, and are alone responsible for approximately 5% of global crop losses. These are soil-borne parasites, where they infest the host plants' root system and cause non-specific symptoms on above-ground parts of plants. Furthermore, they can cause latent (asymptomatic) infestations in potato tubers [2], which pose an additional threat in seeding material, as they could be spread over larger areas quite quickly. The parasite *M. luci* has been originally described by Carneiro et al. from samples from Brazil, Chile, and Iran [3], and has since been found several times in Europe as well [4]. Even though *M. luci* belongs to the group of tropical RKNs, it can survive winter in fields under temperate and Mediterranean climates [5]. It is therefore considered an emerging pest in Europe and was included in the alert list of harmful organisms in 2017 [6].

Since RKNs cause non-specific symptoms, laboratory diagnoses are required for accurate identification. Traditionally RKN species are identified morphometrically, and by analysing dehydrogenase and esterase isozyme phenotypes [7]. These methods require the isolation of mature females from plant tissue, making them unsuitable for a large number of samples. First visible symptoms are presented as reduced plant growth. Infections start in small areas of the crop, but can over the years become full field infestations if not appropriately handled. Particularly latent infestations of potato seed tubers have the

potential to facilitate this process and lead to infestations over larger areas in just one or two seasons. These characteristic of RKN infestations show a clear need for detection of infestations in early stages [8].

Precision agriculture helps reduce the spread of diseases, and includes well-established practices to mitigate losses [9]. The plant immune system reacts to stressors by changing their biophysical and biochemical makeup, which in turn affects their spectral properties [10]. Currently the most common remote sensing method, used for plant diseases detection, is hyperspectral imaging (HSI) [1]. Nowadays, HSI is used in various applications, such as biotechnology, agriculture, environmental monitoring, and chemistry [11]. In HSI, reflected light is captured and data stored in several spectral bands, with a high spectral resolution (bandwidths are typically around 4–5 nm). Consequently, a large number of spectral bands are acquired at each capture, for the entire spectrum the sensors record (typically between 400 and 2500 nm).

Hyperspectral imaging has been used extensively for assessing plant root and tuber quality. The published research can generally be divided into three interlinked groups, according to the investigated properties: (1) physical properties (e.g., colour and texture), (2) chemical constituents (e.g., proteins and polysaccharides), and (3) pest and disease detection. Research into the latter is mostly focused on early detection of infections and infestations in above-ground parts of plants [12]. Biotic and abiotic stressors can cause changes in spectral signatures, by triggering various defense mechanisms, such as production of specific metabolites, induction of hypersensitive reactions, and changes in plant tissues [7]. The most extensive use of remote sensing of potato tubers has been for quality assessment, e.g., detection of defects [13] and bruises [14], and chemometric analyses, such as sugar [15], cellulose and starch content [16]. On the other hand, only a handful of studies deal with pest and disease detection in potato tubers. For example, Dacal-Nieto et al. and Huang et al. used hyperspectral imaging and support vector machines to detect hollow heart disease [17,18], and Zhou et al. used partial least squares as a pre-processing step in linear discriminant analysis to detect blackheart [19]. A partial least squares discriminant analysis approach was used by Garhwal et al. to detect zebra chip disease [20], and Al Riza et al. used a combination of genetic algorithms for feature selection and partial least squares to detect common scab [21]. But results aren't always so clear-cut, as Zhao et al. found that infrared and thermal imaging did not distinguish between healthy and *Liberibacter solanacearum* infected tubers in storage [22].

Hyperspectral data shows a high level of collinearity between spectral bands, leading to high redundancy and decreases the signal-to-noise ratio. Furthermore, the data sets are of high dimensionality, which increases the difficulty of knowledge discovery and pattern recognition. Dimensionality reduction is therefore an obligatory and crucial step in HSI data pre-processing [10].

With dimensionality reduction (DR) methods we retain the descriptive power of the data, but reduce the number of dimensions. This process removes some patterns in the data, but the features of interest remain. The large number of spectral bands can cause reduced discriminating ability of the HSI features. This problem is especially severe when the available training set consists of a small amount of samples (referred as the curse of dimensionality). Dimensionality reduction is therefore a crucial step that transforms the data to lower dimensional space, while preserving relevant information [23]. Generally, DR methods can be grouped into two sets: (1) feature extraction, and (2) feature selection. Feature extraction methods transform the whole feature space to a lower dimensional one, while feature selection approach picks out the most significant features from the whole feature space. Furthermore, contrary to feature extraction, feature selection preserves physical characteristics of the original feature space. But, feature selection information is lost, as features are removed from further analysis. Furthermore, feature extraction is less prone to overfitting and often results in better classification accuracy [23,24]. However, there is no standard approach which would yield the best possible result for any specific dataset [25].

In recent years researchers have shown big interest in developing new methods and tools for processing of hyperspectral data. Trends show that many authors decide to choose feature selection over feature extraction DR algorithms. However conventional feature extraction algorithms are still present, due to a better performance in some cases. Moghimi et al. tested the performance of NaCl treated wheat with different feature selection methods [11]. They aggregated all with ensemble method to increase robustness and accuracy. The transformed data were classified using quadratic discriminants analysis (QDA) and validated using 5-fold cross-validation. Similarly, AlSuwaidi et al. also used feature selection method for crop disease detection [26]. As classification method they used support vector machines (SVM). Moghadam et al. showed that SVM is one of the best classification algorithms used for plant disease detection due to its generalization ability [1]. Their feature extraction method is used based on probabilistic topic modelling. Collected features with reduced dimensionality were pipelined to Latent Dirichlet Allocation model for plant leaves disease detection. On the other hand, Jin et al. directly used convolutional neural networks (CNN) for classification, without any dimensionality reduction [27]. Feature extraction problems for HSI are oftentimes solved by using convolutional neural networks [28,29]. However, CNNs usually need more training data in comparison with conventional methods [28,30], limiting their applicability.

This study was motivated by the need to find a dimensionality reduction method for detection of *M. luci*-infested potato tubers using hyperspectral imaging. The DR method sets were twofold: (1) DR methods can achieve good classification accuracy even with should enable accurate identification in combination with support vector machines (SVM), with only one extracted feature, or a very limited number thereof. Our hypothesis only one feature, and (2) data from the outside of tubers will achieve better classification success. Since only a limited amount of information is available about spectral and chemical effects of nematode infestations on potato tubers, we tested two tuber processing methods. We selected six dimensionality reduction methods, three from each group: Partial least squares (PLS), Linear discriminant analysis (LDA), Principal component analysis (PCA), RandomForest (RF), ReliefF (RFF), and Extreme gradient boosting (XGB). Of these, LDA provided the best results, as it achieved the highest classification accuracy in both external and internal images of potato tubers. All DR methods achieved better success with data from outside of tubers, except for LDA, where the results were equal.

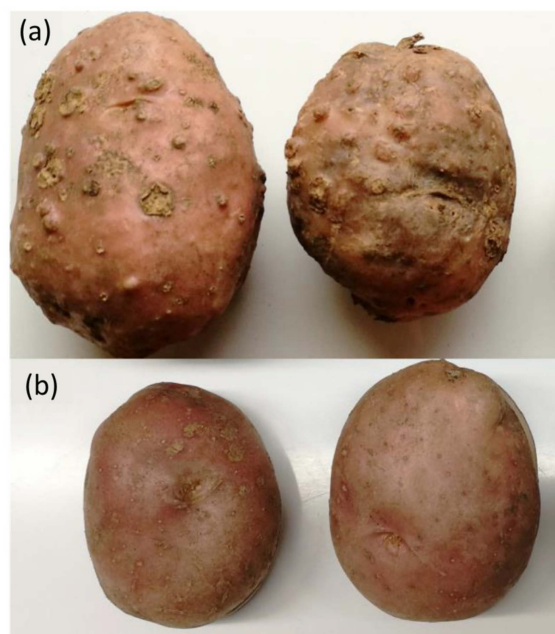
## 2. Materials and Methods

### 2.1. Tuber Cultivation and Preparation

The tubers were obtained from an experiment on potato (*Solanum tuberosum* cv., variety Desiree) infestation with *M. luci*, which was established from June to September 2018 in a glasshouse at the Agricultural Institute of Slovenia. A total of 20 day-old plants were transplanted to 13 cm-diameter pots ( $V = 1$  L) and supported with 1 m plastic-coated stakes. The substrate of 10 randomly selected plants was inoculated with *M. luci* at the beginning of the experiment [31]. Roots of tomato plants, infested by *M. luci* (i.e., egg-masses were visible on the root surface; the parasites were from the collection at the Agricultural institute of Slovenia) were cut into pieces and mixed. A subset of infested roots was then weighed and nematode eggs were collected in suspension, in accordance with Hussey & Barker [32]. The number of eggs was determined visually under a stereomicroscope (Nikon SMZ800). Infested roots were introduced to the substrate to a final concentration of  $250 \times 10^3$  eggs/plant. The presence of *M. luci* was confirmed with isoenzyme analysis [4]. The microplot experiment was completed at the end of the growing season at 97 days after inoculation.

Potato tubers were harvested at the end of the growing season in 2018 and stored in boxes in a dark storage room with ventilation and a temperature of  $18 \pm 2$  °C for the incubation period, until further processing. Tubers from infested pots were visually checked for signs of infection (surface galls) (Figure 1). Visibly decaying tubers were excluded from

further analysis. The tubers were divided into two groups, inoculated and healthy, of 5 tubers each.



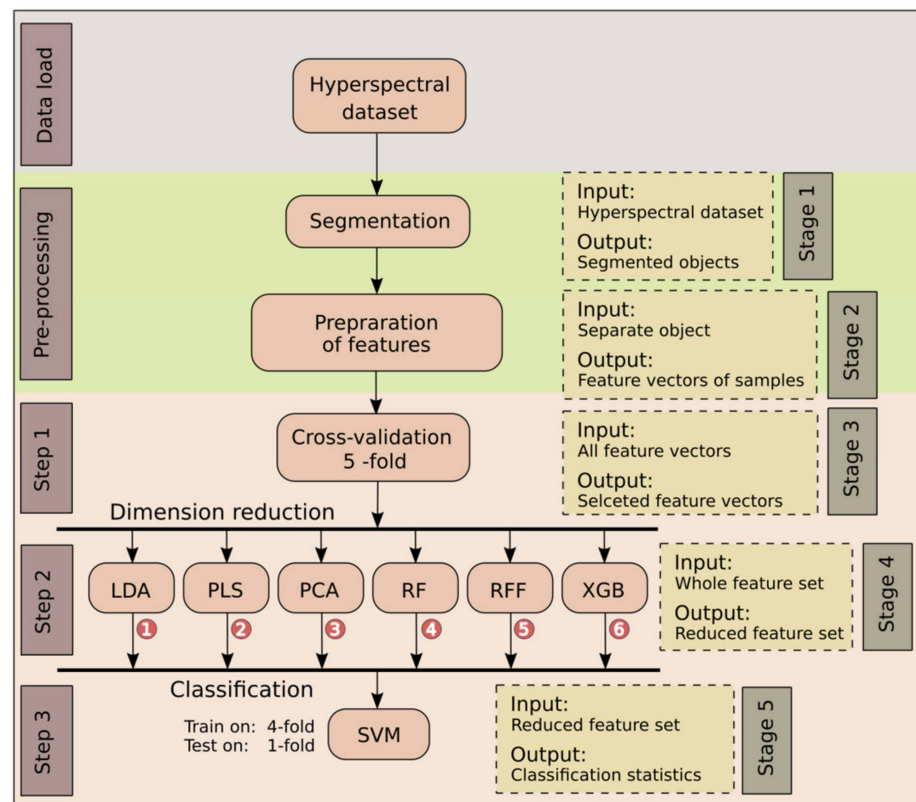
**Figure 1.** Inoculated and healthy potato tuber. (a) inoculated tubers, and (b) healthy tubers. Note the galls on the surface of the infested tubers. The difference in size is not necessarily symptomatic.

### 2.2. Hyperspectral Image Acquisition

For hyperspectral imaging tubers were sliced in half and placed on a black background. This way both the outside and inside of tubers could be imaged simultaneously. Reflectance spectra in 448 bands in the VNIR (visible to near infrared) and SWIR (short-wave infrared) regions were acquired using two Norsk Elektro Optikk AS (Oslo, Norway) pushbroom hypspec cameras, VNIR-1600 (400–988 nm, 160 bands, bandwidth 3.6 nm, spatial resolution at 1 m distance 0.1 mm) and SWIR-384 (950–2500 nm, 288 bands, bandwidth 5.4 nm, spatial resolution 0.25 mm), mounted vertically above the samples at a distance of 1 m. The samples were illuminated with two calibrated halogen lamps with homogeneous light intensity between 400 and 2500 nm, placed above the samples next to the cameras. The lamps were switched on 15 min before image acquisition to stabilize the light source's temperature drift and establish spatial uniformity of illumination [33]. A calibrated diffuse grey reference plate with 20% reflectance (SphereOptics, Herrsching, Germany) was included in each image and used to calculate reflectance. The signal-to-noise ratio was increased by scanning each line three times and calculating the average. Hyperspectral images were radiometrically calibrated to radiance units ( $W\ sr^{-1}m^{-2}$ ).

### 2.3. Pre-Processing and Analysis

The data analysis workflow consisted of five stages (Figure 2). First, radiometrically corrected images were loaded into working memory. Second, images were segmented and these segments were then used to calculate reflectance values and mean spectra of each sample. The segmented image of each tuber was then divided into six equal parts. Reflectance values and mean spectra for each of these sub-segments were extracted. Then, data was split into training (4 potatoes) and test sets (1 potato). This process was repeated 5 times. In the next stage, we applied dimensionality reduction algorithms to extract the most relevant features. In addition to DR methods, we also included a data set without any dimensionality reduction (labelled "None"). In the last stage, the chosen features were tested using support vector machine classification.



**Figure 2.** Analysis flowchart for proposed procedure. Data load: Image acquisition. Pre-processing: Image segmentation and feature preparation techniques. Step 1: Separation of training and validation dataset with cross-validation. Step 2: Dimensionality reduction with LDA, PCA, RF, RFF, XGB or PLS. Step 3: Classification with the use of SVM and 5-fold cross-validation.

### 2.3.1. Segmentation

Image segmentation was performed using spectral information divergence (SID). It uses a divergence measure to match HSI image pixels to reference pixels [34]. In HSI each pixel consists of multiple values which form a discrete signal. For  $j$ -th pixel we can write:

$$\mathbf{x}^{(j)} = (x_1, x_2, \dots, x_D)^T \quad (1)$$

where  $x_i$  represents the value of spectral band  $B_i$  acquired at wavelength  $\lambda_i$ . Index  $D$  represents a number of spectral bands. Probabilistic measure  $p_i$  can be calculated for each element  $x_i$ . Probabilities for all elements are then written into vector of probabilities  $\mathbf{p}$ :

$$\begin{aligned} p_i &= p(x_i) = \frac{x_i}{\sum_{i=1}^D x_i} \\ \mathbf{p}(\mathbf{x}^{(j)}) &= (p_1, p_2, \dots, p_D)^T \end{aligned} \quad (2)$$

Relative entropy can be calculated between  $\mathbf{p}$  and  $\mathbf{q}$  probability vectors with Kullback–Leibler information measure:

$$KL(\mathbf{p}(\mathbf{x}^{(j)}) || \mathbf{q}(\mathbf{r}^{(C_k)})) = \sum_{i=1}^D p_i \cdot \log \frac{p_i}{q_i} \quad (3)$$

where  $\mathbf{q}(\mathbf{r}^{(C_k)})$  represents probability measure for reference vector  $\mathbf{r}$  of  $k$ -th segmentation class. In our case, possible segmentation classes are included in a set:  $C_k \in \{\text{potato tuber, background, reference panel}\}$ . Reference vectors for each segmentation class separately are constructed from manually selected area of pixels. An area for segmentation

class  $C_k$  is defined as:  $S_k = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})$  where  $N$  represents number of pixel vectors included in area corresponding to segmentation class  $k$ . Reference values can be calculated from pixels for each spectral band:

$$r_i = \frac{1}{N} \sum_{j=1}^N x_{ij} \quad i = (1, 2, \dots, D) \quad (4)$$

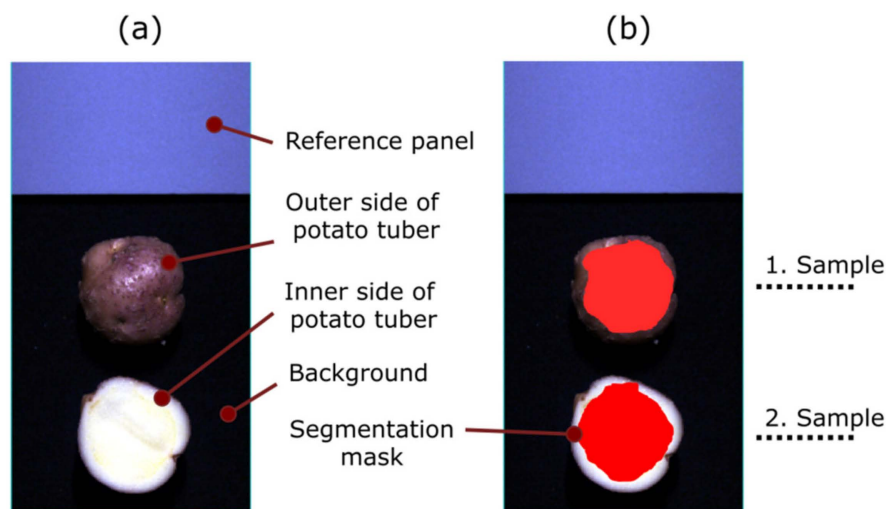
Reference vector is then defined as:

$$\mathbf{r}^{(C_k)} = (r_1, r_2, \dots, r_D)^T \quad (5)$$

Probability vector  $\mathbf{q}(\mathbf{r}^{(C_k)}) = (q_1, q_2, \dots, q_D)$  can then be calculated by (3) for all segmentation classes. When reference vectors are known, SID values can be calculated [35]:

$$SID(\mathbf{p}(\mathbf{x}^{(j)}), \mathbf{q}(\mathbf{r}^{(C_k)})) = KL(\mathbf{p}(\mathbf{x}^{(j)}) \parallel \mathbf{q}(\mathbf{r}^{(C_k)})) + KL(\mathbf{q}(\mathbf{r}^{(C_k)}) \parallel \mathbf{p}(\mathbf{x}^{(j)})) \quad (6)$$

Equation (6) assigns divergence value to each pixel constructing the HSI image. Pixels from an HSI image are classified to segmentation class  $C_k$  with smallest divergence value. The greater the similarity of pixel to reference signal, the smaller the value of divergence. With the use of additional thresholding it is possible to fully separate predefined segmentation classes. Thresholding values were chosen with trial-and-error approach. Segmentation masks for each class can then be built based on calculated divergence values (Figure 3).



**Figure 3.** Hyperspectral image of tubers prior to segmentation. (a) With labelled segmentation class which construct the image. (b) Same image with applied segmentation masks of potato tubers and labelled consecutive sample number.

### 2.3.2. Preparation of Features

Feature vectors can be created from segmentation masks. Pixels that correspond to the same segmentation class  $C_k$  define observing object  $\mathbf{o}^{(z)} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(j)}, \dots, \mathbf{x}^{(N)})^T$ , where  $N$  represents number of connected pixels located within the segmentation mask. Feature vector  $\mathbf{O}$  is calculated from defined objects with arithmetic mean for  $D$  spectral bands. For  $z$ -th object it can be calculated using following equations:

$$\begin{aligned} O_i^{(z)} &= \frac{1}{N} \sum_{j=1}^N x_{ij}^{(z)} \quad i = (1, 2, \dots, D) \\ \mathbf{O}^{(z)} &= (O_1, O_2, \dots, O_D)^T \end{aligned} \quad (7)$$

Spectral reflectance was calculated for tuber data, using a 50% grey reference panel. For each HSI image we calculate feature vector  $\mathbf{O}^{(0)}$ , which represents the reference panel. Reflectance values of reference panel were provided by the manufacturer (SphereOptics, Germany). We assigned those values to vector  $\mathbf{R}^{(0)}$ . For arbitrary object  $z$ , in the same image where  $\mathbf{O}^{(0)}$  was calculated, converted feature vector  $\mathbf{R}^{(z)}$  can be calculated as:

$$\begin{aligned} R_i^{(z)} &= R_i^{(0)} / O_i^{(0)} \cdot O_i^{(z)} \quad i = (1, 2, \dots, D) \\ \mathbf{R}^{(z)} &= (R_1, R_2, \dots, R_D)^T \end{aligned} \quad (8)$$

An important note to add is that object of reference panel  $\mathbf{o}^{(0)}$  is filtered before usage in Equation (8). The reason is to remove outliers for more accurate calculation of feature vector  $\mathbf{O}^{(0)}$ . Outliers are filtered with median absolute deviation (MAD). Values outside  $\pm 2\text{MAD}$  were removed before further calculation. MAD can be calculated by Equation (9), for each spectral band:  $i = (1, 2, \dots, D)$ . In equation  $\mathbf{x}^{(i)}$  represents vector of all pixels at  $i$ -th spectral band. Value labeled as  $\tilde{x}^{(i)}$  represents median value of this vector. With  $\mathbf{I}$  row vector of ones is labelled.

$$\begin{aligned} \mathbf{x}^{(i)} &= \left\{ o_{ij}^{(0)} \mid i \in D, 1 \leq j \leq N \right\}_{1 \times N} \\ \tilde{x}^{(i)} &= \text{median}(\mathbf{x}^{(i)}) \\ \text{MAD}^{(i)} &= \text{median} \left( \left| \mathbf{x}^{(i)} - \tilde{x}^{(i)} \mathbf{I} \right| \right) \end{aligned} \quad (9)$$

The feature vector  $\mathbf{R}^{(z)}$  was then smoothed using a Savitzky–Golay filter to emphasize small spectral variations the same way as in Schafer [36]. Savitzky–Golay filter is based on local least squares polynomial approximation. It was shown that it reduces noise while maintaining shape and important information in a feature vector [26]. In this study the filter window length was 15, polynomial order was 2 and second order derivatives were used. Parameters were chosen with regard to the highest exhibition of performance evaluated with classification accuracy.

### Dimensionality Reduction

In this paper three feature extraction methods (Principal component analysis, Linear discriminant analysis, and Partial least squares), and three feature selection methods (RandomForest, Extreme Gradient Boosting, and ReliefF) are taken into consideration. Main reason for pre-processing the data with the use of stated algorithms is to reduce the number of dimensions in initial space. From all algorithms we extracted only the most prominent features.

### Principal Component Analysis

Principal component analysis is an unsupervised linear transformation technique used in machine learning applications and multivariate statistics. It is widely used across different fields, most prominently for feature extraction, dimensionality reduction and visualization. It helps identify patterns in data based on the correlation between features. PCA aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one [37]. This is achieved based on a covariance matrix formulation of centered and normalized data. Axes of original coordinate system are transformed so that newly created axes describe maximal covariance of the data. Each axis is described by an eigenvector, whose variance corresponds to its eigenvalue. Eigenvectors and eigenvalues of covariance matrix  $\Sigma$  can be calculated using Singular value decomposition (SVD), which can be written as:

$$\Sigma = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} = \mathbf{P} \Lambda \mathbf{P}^T = \sum_{j=1}^m \lambda_j \mathbf{p}_j \mathbf{p}_j^T \quad (10)$$

In Equation (10) data instances are included in matrix  $X \in \mathbb{R}^{n \times m}$ , where  $n$  represents number of data instances and  $m$  number of attributes in each instance. Matrix  $P \in \mathbb{R}^{m \times m}$  represents  $m$  orthogonal basis vectors  $p_j, j = 1, \dots, m$  and  $\Lambda$  diagonal matrices composed of eigenvalues  $\lambda_j, j = 1, \dots, m$ . To each eigenvalue  $\lambda_j$  belongs particular basis vector  $p_j$ . Matrix of eigenvectors  $P$  is organized so that column vectors are sorted by decreasing magnitude of eigenvalues  $\lambda_1 < \lambda_2 < \dots < \lambda_m$ . In other words, eigenvectors are sorted by decreasing amount of information they provide.

Since HIS data contains many correlated features (i.e., spectral bands), the data set can be fully described by using only a subset of eigenvectors of covariance matrix  $\Sigma$ . The general assumption is that part of the information can be explained with  $k$  eigenvectors, which we call principal components. Various criteria can be used to determine the number of principal components, e.g., percentage of explained variance in the data. Mathematically we can write:

$$\frac{\sum_{j=1}^s \lambda_j}{\sum_{j=1}^m \lambda_j} \geq \epsilon \quad (11)$$

where  $\epsilon$  represents predefined threshold. Usually it is set to 0.95 to keep 95% of initial variance of the data. Another option is to directly choose desired number of principal component. The covariance matrix can then be approximated by neglecting eigenvectors with small corresponding eigenvalues. In other words, we neglect  $p_j$ , where  $j = s + 1, \dots, m$ . Approximated covariance matrix can then be calculated as:

$$\Sigma_s = \sum_{j=1}^s \lambda_j p_j p_j^T \quad (12)$$

where vectors  $p_j, j = 1, \dots, s$  define principal directions in which the data extends and is weighted by corresponding eigenvalues.

### Linear Discriminant Analysis

Linear discriminant analysis is a robust classification method, but can also be used for dimension reduction and data visualization. Unlike PCA, which tries to maximize variance, it is a supervised machine learning method that computes decision boundaries which enhance the separation between multiple classes used.

It tries to separate different classes by maximizing distances between projected means and minimizing projected variance. Both optimization problems are incorporated in one single criterion function which can be, for binary classification, written as:

$$\max_w J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \quad (13)$$

where  $(m_1 - m_2)^2$  represents the difference in means between the two classes and  $s_1^2 + s_2^2$  the total scatter (standard deviations) of the two classes. The goal of LDA is to find the vector  $w$  that maximizes criterion function  $J(w)$ .

In LDA it is assumed that all  $K$  classes have equal covariance. Following this assumption we can obtain the following discriminant function for  $k$ -th class:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \quad (14)$$

Which predicts the class with the highest value of  $\delta_k(x)$  given an input  $x \in \mathbb{R}^{p \times 1}$ . In Equation (14)  $\Sigma \in \mathbb{R}^{p \times p}$  represents common covariance matrix,  $\mu_k \in \mathbb{R}^{p \times 1}$  the mean of inputs for class  $k$  and  $\pi_k$  prior distribution of class  $k$ . Symbol  $p$  represents number of attributes in each data instance.

Features are transformed so that classes are as separate as possible from each other and that features within a class are as close as possible. Transformed dimensions are ranked



based on the separation ability. Maximal number of components must be at least one fewer than the number of classes used for classification. Therefore, since we performed binary classification in this study, only the first and only linear discriminant was used [38].

### Partial Least Squares

Partial least squares is a technique that transforms the initial dataset to a reduced set of uncorrelated features using a technique similar to principal component analysis. It extracts features that describe maximum correlation with target variables; i.e., they provide the greatest predictive ability. This method is especially useful when features in initial dataset are highly collinear [39].

The underlying core equations of PLS could be written as:

$$X = TP^T + E \quad (15)$$

$$Y = UQ^T + F \quad (16)$$

where  $X \in \mathbb{R}^{n \times m}$  is the matrix of independent variables (with hyperspectral data spectral bands) and  $Y \in \mathbb{R}^{n \times p}$  is the matrix of dependant variables (these can be measured variables or dummy coded nominal variables). Symbol  $n$  represents number of data instances,  $m$  number of attributes in each independent variable and  $p$  number of attributes in each dependent variable. Matrices  $T$ ,  $U \in \mathbb{R}^{n \times l}$  respectively represent projections (scores) of  $X$  and  $Y$ . Matrices  $P \in \mathbb{R}^{m \times l}$  and  $Q \in \mathbb{R}^{p \times l}$  respectively represent orthogonal loadings matrices of  $X$  and  $Y$ . Symbol  $l$  represents a user-defined number of latent factors used in for regression. Model is optimized in such a way that the first score in  $X$  has maximum covariance with the first score in  $Y$ . Therefore, we can predict the first score in  $Y$  from the first score in  $X$ .

Partial least squares has some advantages over basic ordinary least square (OLS) solution. It is able to dispose correlated variables and model their shared and underlying information. In contrast to many machine learning methods, it can directly model multiple dependent variables at the same time. Several variants of PLS exist; we used Partial least squares discriminant analysis, which is an extension of PLS regression, the foundation for other variants.

### ReliefF

ReliefF is an extension of the basic Relief algorithm, and is a generally well-performing attribute selector. It can provide a combined view of relevance and conditional dependencies between attributes. The algorithm prescribes a separate weight ( $w_j$ ) to each attribute, where higher values correspond to more important attributes. The basic idea of the algorithm is that it penalizes attributes which provide different result of the same class in comparison with its nearest neighbours [40].

At the beginning, ReliefF sets all attribute weights  $w_j^1$  to zero, these are then iteratively adapted. Then, it selects a random observation  $x_r$  and  $k$ -nearest observations for each class. All the weights are updated for each nearest neighbours  $x_q$  by equations:

$$w_j^i = w_j^{i-1} - \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq} \quad (17)$$

$$w_j^i = w_j^{i-1} + \frac{p_{yq}}{1 - p_{yr}} \cdot \frac{\Delta_j(x_r, x_q)}{m} \cdot d_{rq} \quad (18)$$

where  $w_j^i$  represents the weight of the  $j$ -th attribute at iteration  $i$ ,  $m$  is the total number of iterations,  $p_{yq}$  and  $p_{yr}$  are prior probabilities of classes where  $x_q$  and  $x_r$  respectively belong, and  $d_{rq}$  is the distance function, which is subject to scaling. Symbol  $\Delta_j(x_r, x_q)$  represents

the difference between prediction values of observations  $x_r$  and  $x_q$  for  $j$ -th attribute  $F_j$ . For continuous attributes it is calculated as:

$$\Delta_j(x_r, x_q) = \frac{|x_{rj} - x_{qj}|}{\max(F_j) - \min(F_j)} \quad (19)$$

### RandomForest

Random forest is an ensemble technique that combines multiple de-correlated decision trees. Decision trees are fitted on various randomly chosen subsets of a given dataset. Overall performance of the model is increased by aggregating predictions from all trees and performing a majority vote for each class in classification problems.

In the training phase of Random forest a technique called bootstrap aggregation or bagging is used.

Given training data instances in a matrix  $X \in \mathbb{R}^{n \times m}$  and corresponding labels  $y \in \mathbb{R}^{n \times 1}$  (where  $n$  represents number of data instances and  $m$  number of attributes in each instance), bagging repeatedly selects random data instances with replacement and fits  $B$  decision trees  $f$  to these instances. Unseen data instances  $x'$  are predicted by averaging all predictions made by individual decision trees:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (20)$$

where  $\hat{y}$  represents approximated predicted output. Bagging decreases variance without increase of bias. This leads to more accurate performance even if each individual decision tree is highly sensitive to noise. Furthermore, Random forest also includes feature bagging, i.e. selection of a random subset of the attributes in the training set. A small number of attributes may have a very strong prediction power for the response variable. Consequently, these attributes would be selected many times causing decision trees to become correlated. We used the Gini index as split criterion and for assessing variable importance. For each Random forest 100 trees were constructed [41].

### Extreme Gradient Boosting

Gradient boosting is one of the most powerful and flexible machine learning methods, which can be applied to various machine learning problems. It refers to a class of ensemble methods used for predictive modelling problems. Similarly to Random forests, it is constructed from decision tree models (weak learners). Unlike Random forest, weak learners are added one at a time to correct errors produced by prior decision trees. This type of error correction is called boosting, where models are iteratively trained with gradient descent optimization of any differentiable loss function. For instance, a squared error may be used for regression problems and logarithmic loss for classification problems. New decision trees are trained on error residuals produced by initial learner. Intuitively, newly trained models are influenced more by misclassified observations or by areas where they are performing poorly. The contribution from all decision trees are aggregated to make the final prediction [42].

Simplified optimization could be mathematically written as follows. First model is initialized with a constant value with minimization of loss function  $L(x_i, y_i, \theta)$ :

$$\hat{f}_{(0)}(X) = \operatorname{argmin}_{\theta} \sum_{i=1}^n L(x_i, y_i, \theta) \quad (21)$$

where  $X \in \mathbb{R}^{n \times m}$  is a matrix of input data instances with corresponding labels  $y \in \mathbb{R}^{n \times 1}$ . Symbols  $n$  and  $m$  represent number of data instances and number of attributes in each instance, respectively. Based on the weak learner from the previous iteration, gradients and

hessians are calculated and then a new weak learner is fitted using optimization problems. At the end of iteration, the model transfer function is updated as:

$$\hat{f}_{(m)}(\mathbf{X}) = \hat{f}_{(m-1)}(\mathbf{X}) + \hat{f}_{(m)}(\mathbf{X}) \quad (22)$$

where  $m = (1, 2, \dots, M)$ , where  $M$  is the total number of weak learners. Unseen data instances  $\mathbf{x}'$  are then predicted by summation of all predictions made by individual decision trees:

$$\hat{y} = \hat{f}(\mathbf{x}') = \hat{f}_{(M)}(\mathbf{x}') = \sum_{m=0}^M \hat{f}_{(m)}(\mathbf{x}') \quad (23)$$

#### 2.4. Support Vector Machines

Extracted or selected features from dimensionality reduction were pipelined to support vector machine classifier. Classification models were therefore built on reduced data sets, consisting of only the most prominent feature. In this study performance is tested on radial-basis kernel function for data transformation [43]. Hyperparameter tuning (gamma and C) was performed using a grid search, whereupon combinations yielding the best accuracy were retained.

Trained SVM classifier was evaluated using mean accuracy. It was iteratively trained and tested 5 times for each DR algorithm. Accuracies from all iterations were then averaged. For this reason an objective criteria is devised for comparison between all DR methods:

$$\begin{aligned} \bar{c}^{(m)} &= \frac{1}{F} \frac{1}{P} \sum_{f=1}^F \sum_{p=1}^P \Gamma(y_p, \hat{y}_p^{(m)}) \\ \Gamma(y_p, \hat{y}_p) &= \begin{cases} 1; & \text{if } y_p = \hat{y}_p^{(m)} \\ 0; & \text{else} \end{cases} \\ y_p, \hat{y}_p^{(m)} &\in \{\text{Healthy, Inoculated}\} \end{aligned} \quad (24)$$

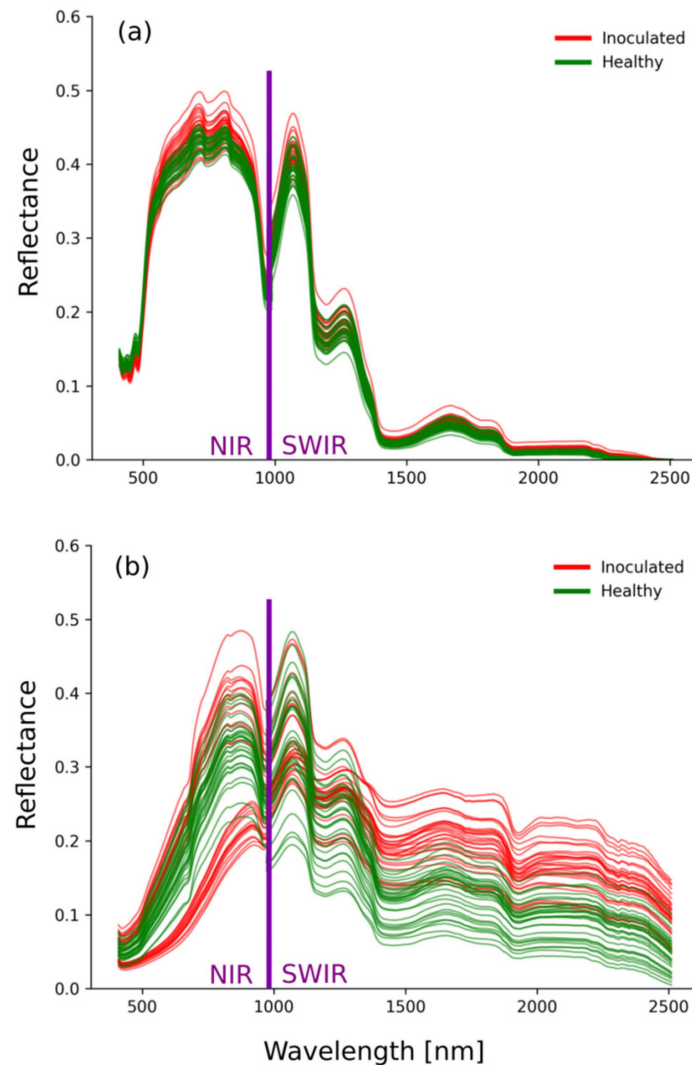
In Equation (24) mean accuracy is labelled as  $\bar{c}^{(m)}$ . Superscript represents  $m$ -th DR method belonging to a set  $M_m \in \{\text{PCA, LDA, PLS, RF, RFF, XGB}\}$ . Mean accuracy is calculated from comparison between predicted  $\hat{y}_p$  and known  $y_p$  labels of potato tuber, which can be either healthy or inoculated. It is calculated on test feature vectors for  $P$  predictions and  $F$  folds. In our case  $F = 5$  and  $P = 12$ . Precision, recall and F1-score were calculated using equations in [44]. All analyses were performed in Python, using libraries scikit-learn [45] and XGBoost [46].

### 3. Results and Discussion

The first extracted features, or limited set of selected features, proved to be sufficient for accurate detection of infested potato tubers. Spectral differences between inoculated and healthy tubers were more pronounced in images of their outside (Figure 4). Spectral signatures of the outside of tubers show a larger variability in infested tubers, than in healthy ones. These differences are more pronounced in the SWIR region, where inoculated tubers uniformly exhibited higher reflectance than healthy tubers. In contrast, data from inside tubers shows comparatively little variability, regardless of inoculation status. The high variability in outside images could be a consequence of tuber surface characteristics. Healthy tubers are comparatively smooth, while galls cover the surface of inoculated tubers. This leads to a more varied viewing geometry, which was accounted for in pre-processing of the images.

We used the first two principal components for data visualization in a generated feature space. Feature vectors were separated into 5 cross-validation folds, and PCA performed on each fold of the training data, and applied to both train and test sets. The generated features from all folds were pooled to generate scatter plots of the first two PCA components (Figure 5). The first two components explain more than 80% of the variance in the data (93% for outside, and 84% for inside tubers). These scatter plots show

a better distinction between healthy and inoculated tubers for data from the outside of tubers. Yet any linear separability does not appear to be present, at least not in the first two PCA components.

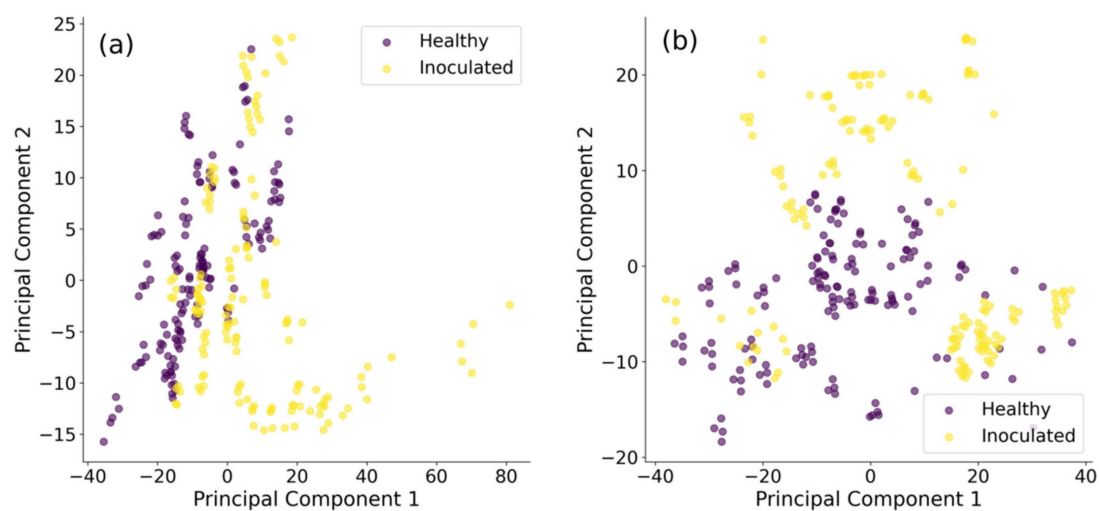


**Figure 4.** Reflectance feature vectors for (a) inside and (b) outside of potato tuber. Green colour refers to healthy and red to inoculated specimens. Separation between VNIR and SWIR cameras is marked with violet colour.

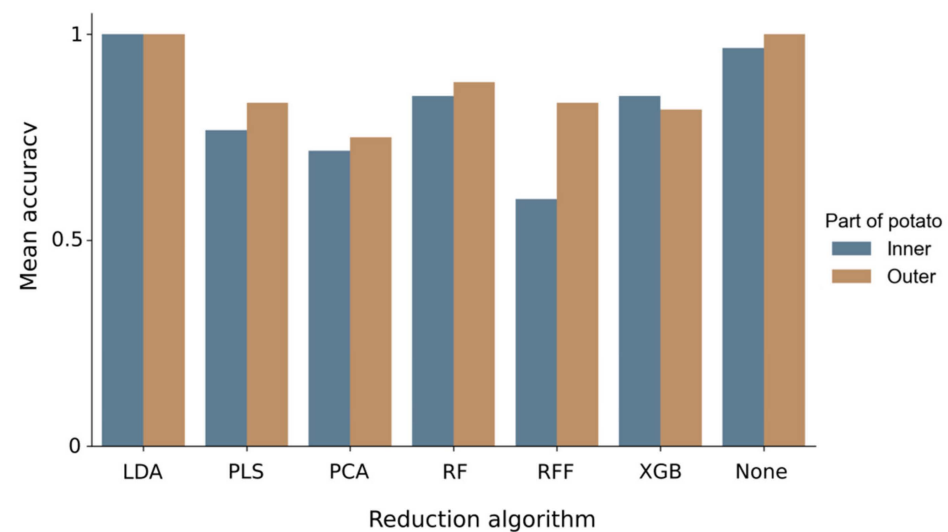
Females of root-knot nematodes reside within a few millimeters below tuber skin, in the vascular ring, where they form comparatively large egg-sacs [47]. When these grow enough, they form galls on tuber surface. Even though some evidence exists that RKNs change the chemistry of the entire tuber, i.e., also the starchy insides [2], we expected classification success to be higher in outside data. This hypothesis was confirmed for all DR methods (Figure 6). All methods achieved a mean accuracy of at least 0.6 (Tables 1 and 2). LDA showed the most consistent results, as it achieved a mean accuracy of 1.0 in both inside and outside tuber data. Overall RF came second, with XGB yielding very similar accuracies. PLS came next, followed by PCA, and lastly ReliefF achieved the worst results. Data was also analysed without any DR. Even though this method was capable of achieving good accuracy, a mean of 0.9 from both data sets, it is computationally much more burdensome. With SVMs solving the quadratic problem involves inverting the kernel matrix, with a complexity of up to  $n^3$ , where  $n$  is the feature space [43]. RandomForest, XGB and ReliefF suffer from the same problem, as they only select features, i.e. they generate a subset of the

original feature space. Furthermore, even though all three feature selection methods are robust, they can suffer from overfitting and should be optimized accordingly [48]. In this regard feature extraction methods are beneficial, since they generate a new feature space, with lower dimensionality.

Increasing the number of features has an expected effect, of increasing classification accuracies (Figure 7). The most profound effect is observable in PCA and RFF, while RF shows the smallest change. Unlike the other five methods, only one feature gets extracted by LDA in binary classification. Interestingly, PCA on inside data decreases accuracy with the first three components. Accuracy then increases with more features, but still remains below PLS and LDA accuracies. With more features extracted, only PLS and XGB achieve a 100% accuracy in both inside and outside tuber data. Our results indicate that even with an extreme reduction, to just one feature, identification accuracies are still acceptable to excellent (mean accuracy between 0.8 and 1.0).



**Figure 5.** Scatter plot of first two principal components for (a) inside and (b) outside of potato tubers. Yellow colour refers to healthy and purple colour to inoculated specimens.



**Figure 6.** Mean accuracy for several dimensional reduction algorithms. Accuracy of classifier trained on data from: outer side of potato (orange columns), inner side of potato (blue columns).

**Table 1.** Comparison of classification results with only one feature for selected DR methods for outside tuber data.

Outer Side of Potato Tuber					
Method	Class	Precision	Recall	F1-Score	Accuracy
LDA	Healthy	1.00	1.00	1.00	1.00
	Inoculated	1.00	1.00	1.00	
PLS	Healthy	0.88	0.77	0.82	0.83
	Inoculated	0.79	0.90	0.84	
PCA	Healthy	0.86	0.60	0.71	0.75
	Inoculated	0.69	0.90	0.78	
RF	Healthy	0.93	0.83	0.88	0.88
	Inoculated	0.85	0.93	0.89	
RFF	Healthy	0.88	0.77	0.82	0.83
	Inoculated	0.79	0.90	0.84	
XGB	Healthy	0.81	0.83	0.82	0.82
	Inoculated	0.83	0.80	0.81	
None	Healthy	1.00	1.00	1.00	1.00
	Inoculated	1.00	1.00	1.00	

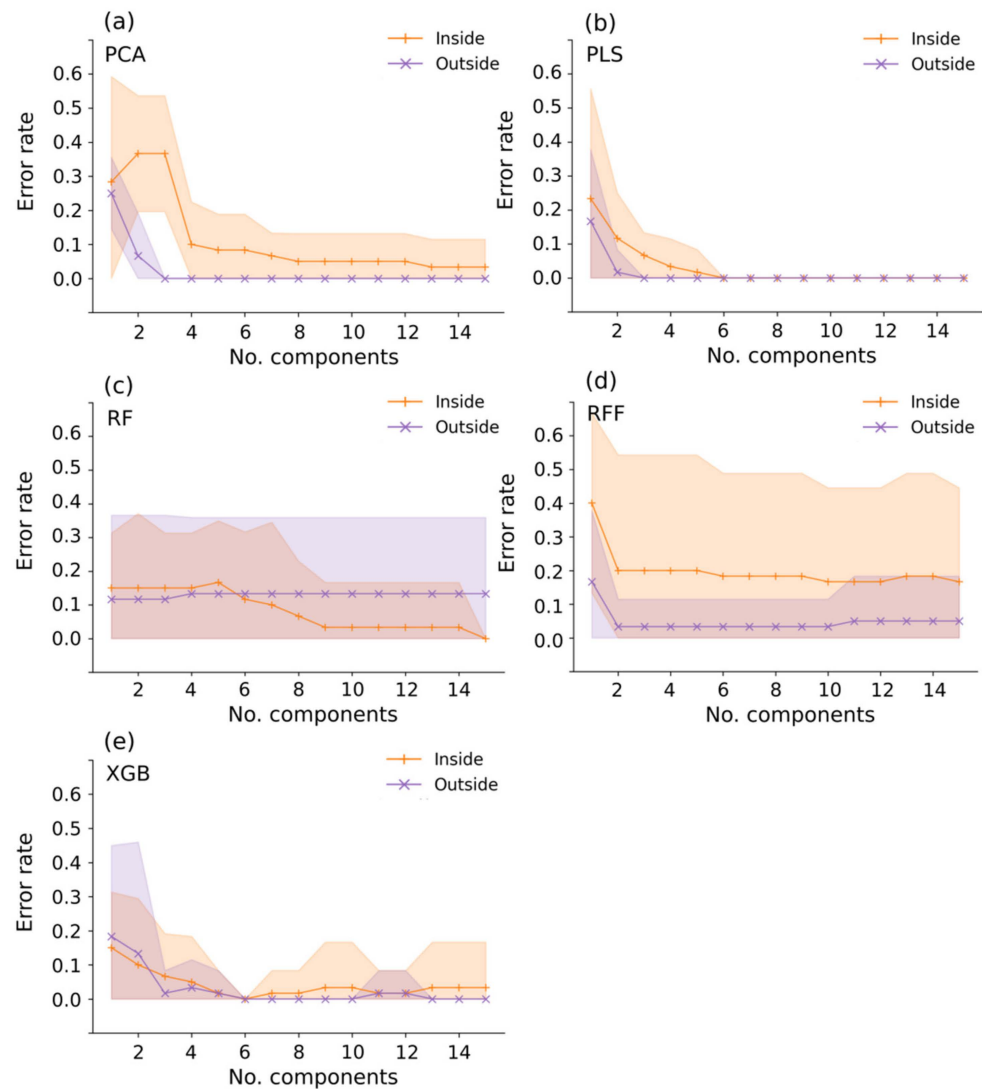
**Table 2.** Comparison of classification results with only one feature for selected DR methods for inside tuber data.

Outer Side of Potato Tuber					
Method	Class	Precision	Recall	F1-Score	Accuracy
LDA	Healthy	1.00	1.00	1.00	1.00
	Inoculated	1.00	1.00	1.00	
PLS	Healthy	0.75	0.80	0.77	0.77
	Inoculated	0.79	0.73	0.76	
PCA	Healthy	0.70	0.77	0.73	0.72
	Inoculated	0.74	0.67	0.70	
RF	Healthy	0.89	0.80	0.84	0.85
	Inoculated	0.82	0.90	0.86	
RFF	Healthy	0.64	0.47	0.54	0.60
	Inoculated	0.58	0.73	0.65	
XGB	Healthy	0.89	0.80	0.84	0.85
	Inoculated	0.82	0.90	0.86	
None	Healthy	0.94	1.00	0.97	0.97
	Inoculated	1.00	0.93	0.97	

Compared to no DR, dimensionality reduction using PLS and PCA on outside data reduced detection accuracy (Table 2). On the other hand, in a data set with less pronounced patterns, such as tuber insides, classification accuracy was increased by using PLS. One of the purposes of dimensionality reduction is to generate or retain only those features, which are informative for the problem under study. This way the signal-to-noise ratio can be improved, leading to better model performance.

Dimensionality reduction algorithms use different metrics to assess feature importance. For example, principal components analysis uses a correlation matrix between generated components and original features. Correlations above or below 0.7 or  $-0.7$ , respectively, are considered as relevant. In PLS correlations can also be considered, but a more accurate assessment of feature importance is possible using variable importance in projection analysis (VIP) [49]. VIP coefficients reflect the relative importance of each variable for each variate in the prediction model. Variable importance in LDA was calculated as LDA scalings, i.e., the eigenvectors of the components. Important to note here is that LDA is a discriminant analysis method and as such it maximizes the between-group variance. The eigenvectors of the comparison matrix of between and within group's sum of squares and cross-products describe how much the original variables contribute to the new component(s). Gini im-

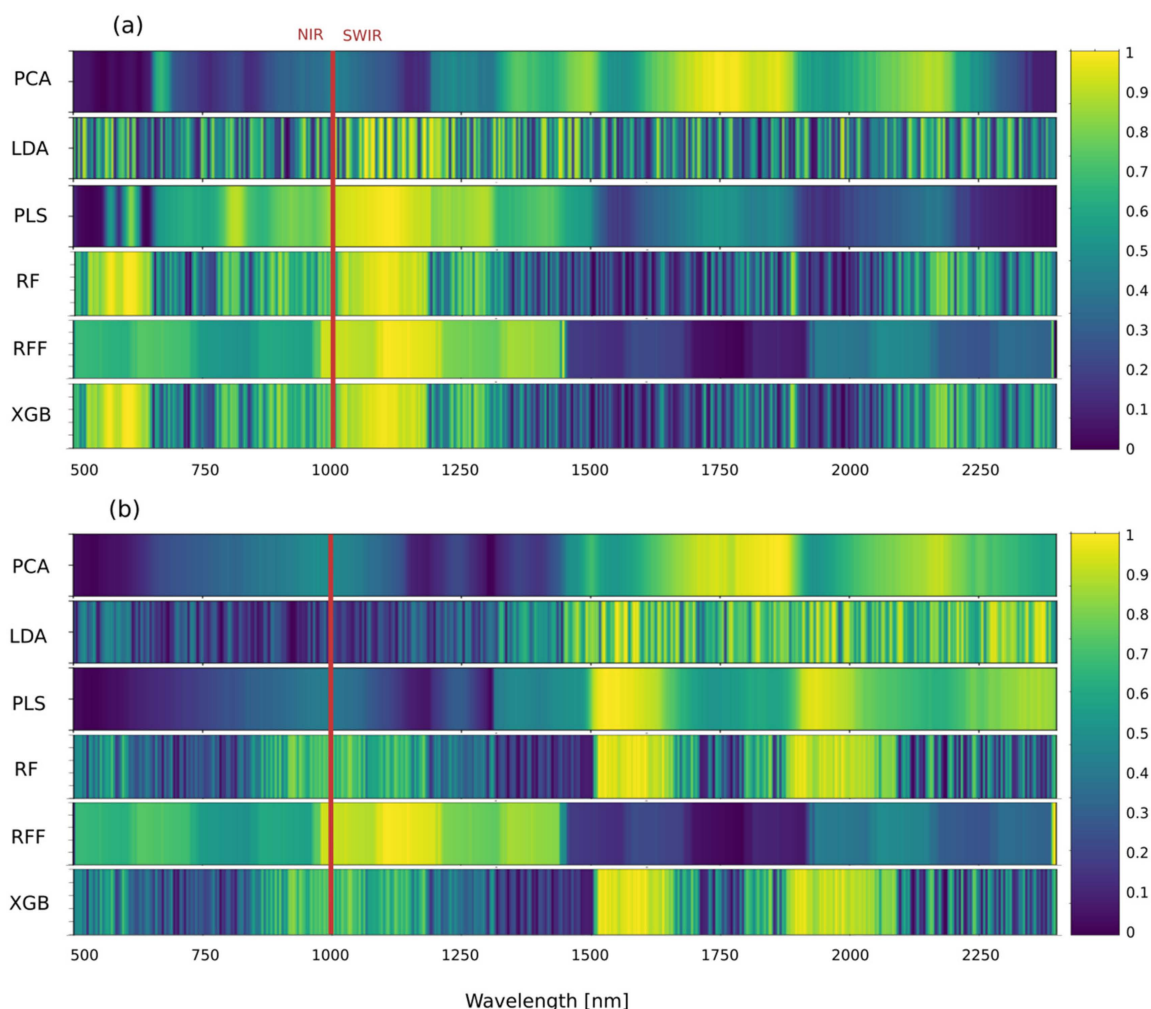
portance, used with RandomForests and Extreme gradient boosting, provides a relative ranking of the original features, and is a by-product of the training of the classifier [50]. Lastly, ReliefF assigns feature relevance depending on the difference between this feature and two neighbours of the same and opposite classes [51]. Each of these methods provides their own metric of variable importance. In order to directly compare all methods, we normalized their values to a range of 0 to 1.



**Figure 7.** Influence of increasing number of features on classification error rates for both inside and outside tuber data. Since LDA generates only one feature in binary classification it wasn't included in this figure. (a) Principal component analysis, (b) Partial least squares, (c) Random forest, (d) ReliefF, and (e) Extreme gradient boosting.

RandomForest and XGB identified several relevant wavelengths, distributed comparatively evenly throughout the spectrum, in both data sets. Similarly, LDA also identified a large number of relevant wavelengths, but unlike RF, these were not evenly distributed in the outside data set. In this set the importance of variable shifted towards the SWIR region. Only ReliefF showed a different pattern, compared to the other five methods. Interestingly, while PLS, LDA, RF, and XGB show a similar grouping of relevant wavelengths in the same SWIR regions (1500–1600 nm, 1850–2000 nm, and 2300–2450 nm), PCA found relevant regions between these groups (1600–1800 nm, and 2100–2200 nm). On the other hand, ReliefF found relevant wavelengths in the range 1000–1400 nm, i.e., in shorter wavelengths than the other methods. With inside data, only PCA found relevant regions in the SWIR

part of the spectrum, above 1500 nm (1600–1850 nm), while the remaining methods found a strong grouping of relevant wavelengths closer to the VNIR region, between 1000 and 1200 nm (Figure 8). This spectral region is linked to various hydrocarbons, both aliphatic and aromatic [52]. The region identified by PCA is also linked to different hydrocarbons (aliphatic, aromatic, and methyl), but also alcohols, amines and proteins. In the outside data set, regions linked to water, polysaccharides, aromatic amines (1850–2000 nm), and lipids and glucose (2300–2450 nm) were identified as relevant. In order to fully test the accuracy of the different variable importance measures, employed by the dimensionality reduction methods, more detailed chemometric analyses of potato tubers are needed.



**Figure 8.** Importance of individual wavelengths, as determined by the different DR methods. (a) inside, and (b) outside of potato tuber. Brighter colours represent higher importance. Separation between VNIR and SWIR cameras is marked with red color.

Differences on the inside of tubers are most likely of a chemical nature, while the outside is influenced by both differences in chemistry as well as surface texture. While surface texture effects can be reduced using normalization and Savitzky–Golay derivatives, they can still affect the classifications. The infested tubers used in this study had visible symptoms (bulges on the outside, approximately 3–5 mm in diameter), covering at most 50% of the surface. In our case, they added another distinguishing dimension, leading to better classification accuracy with data from tuber surfaces. Nevertheless, the feasibility of hyperspectral imaging for detection of latent (i.e., without visible symptoms) infestations with nematodes in potatoes has been shown by Žibrat et al. [2]. In that case surface texture didn't influence spectral signatures; therefore the observed differences are exclusively



due to differences in chemistry between healthy and infested tubers. From a practical application standpoint, even if tubers in production spot checks would have to be halved and imaged, the throughput of such a method would be much higher than with molecular analyses (e.g., real-time PCR). So the benefit of using hyperspectral imaging for detecting infestations, even with processed tubers, is evident.

Molecular spectra are the result of motions of atomic nuclei. They can rotate, vibrate, wag and move together or apart along a straight line (this type of movement is called stretching). Vibrations follow a functional description, i.e. the type of vibration determines the frequency at which it absorbs energy. The amplitude of absorption is determined by absorptivity and the number of molecules in the beam path of a particular sensor. Changes in spectral responses follow Beer's Law, which states that the absorbance is equal to product of absorptivity of a molecule and the concentration of molecules. The above means that the light absorbed by plant tissue depends on the chemical composition of that tissue, the concentration of individual molecule species, and their interactions [52]. We've identified several groups of molecules, which account for the differences between healthy and infested potato tubers (such as aliphatic and aromatic hydrocarbons). The two spectral regions, VNIR and SWIR, are generally linked to different characteristics of plants, but there is some overlap. In the VNIR region we mostly get information about pigments and structure, e.g., morphological structure of plant leaves. In addition, in wavelengths above 700 nm there is also information related to hydrocarbons (mostly aliphatic) and alcohols, these are generally the third and fourth overtones of the C-H stretch. The SWIR region carries information on plant biophysical properties (e.g., hydrocarbons and proteins). Changes in chemistry can therefore be measured in both. Since we do not know how exactly nematode infestations change the chemistry of potato tubers, we decided to use both systems. The wider spectral range of the combined system enables us to better search for the effects of nematode infestations. With the currently available information we can only speculate which exact compounds account for the observed differences between inoculated and healthy potato tubers.

We identified several relevant wavelengths from each DR method. The latter use different methods for deciding, which wavelength is relevant, so we would recommend to look for overlaps between the methods. Those wavelengths which were identified as relevant by different methods are good candidates for a multispectral sensor. Different bandpass filters are commercially available, with different spectral ranges and bandwidths. So it might not be necessarily needed to develop new filters. Using these filters we would effectively get a multispectral data set. New classification models would then have to be developed using this reduced data. Using this procedure we could assess the importance of each new wide band and determine which ones significantly increase classification success. Potentially this would mean that by reducing a hyperspectral dataset to a few-band (~5 bands) multispectral one we would still get acceptable classification accuracy. Production of such a dedicated multispectral system would be much cheaper, than a hyperspectral one, with similar classification success.

#### 4. Conclusions

In this study we have shown that discrimination between healthy and inoculated potatoes with quarantine pests is possible based on hyperspectral image analysis. We successfully reduced the initial hyper-dimensional feature space to one-dimensional (or few-dimensional) with the use of dimensionality reduction algorithms, and still obtained high classification accuracies. These results suggest that a comparatively low-cost imaging system utilizing band-pass filters could be designed for the specific purpose of identifying tubers infested by root-knot nematodes. But the study was performed on a very small data set and further analyses are needed to fully test this idea.

**Author Contributions:** Conceptualization, J.L., U.Ž., M.K.; methodology, J.L. and U.Ž.; formal analysis, J.L. and U.Ž.; writing—original draft preparation, J.L. and U.Ž.; writing—review and editing, U.Ž. and M.K.; project administration, U.Ž.; funding acquisition, U.Ž. and M.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by funds from the Slovenian Research Agency (ARRS) (MR 38128, MR 54720, P4-0072). Some of the research was conducted using equipment financed by EU-FP7 project CropSustaIn, grant agreement FP7-REGPOT-CT2012-316205.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their use in ongoing research projects.

**Acknowledgments:** We would like to thank Nik Susič, Saša Širca, and Barbara Gerič Stare for providing the plant material and performing nematode species identification.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Moghadam, P.; Ward, D.; Goan, E.; Jayawardena, S.; Sikka, P.; Hernandez, E. Plant Disease Detection Using Hyperspectral Imaging. In Proceedings of the 2017 International Conference on Digital Image Computing: Techniques and Applications, Sydney, Australia, 29 November–1 December 2017; pp. 1–8.
- Žibrat, U.; Gerič Stare, B.; Knapič, M.; Susič, N.; Lapajne, J.; Širca, S. Detection of root-knot nematode *Meloidogyne luci* infestation of potato tubers using hyperspectral remote sensing and real-time PCR molecular methods. *Remote Sens.* **2021**, *13*, 1996. [CrossRef]
- Carneiro, R.M.D.G.; Correa, V.R.; Almeida, M.R.A.; Gomes, A.C.M.M.; Deimi, A.M.; Castagnone-Sereno, P.; Karsen, G. *Meloidogyne luci* n. sp. (Nematoda: Meloidogynidae), a root-knot nematode parasitising different crops in Brazil, Chile and Iran. *Nematology* **2014**, *16*, 289–301. [CrossRef]
- Gerič Stare, B.; Strajnar, P.; Susič, N.; Urek, G.; Širca, S. Reported populations of *Meloidogyne ethiopica* in Europe identified as *Meloidogyne luci*. *Plant Dis.* **2017**, *101*, 1627–1632. [CrossRef]
- Strajnar, P.; Širca, S.; Knapič, M.; Urek, G. Effect of Slovenian climatic conditions on the development and survival of the root-knot nematode *Meloidogyne ethiopica*. *Eur. J. Plant Pathol.* **2011**, *129*, 81–88. [CrossRef]
- EPPO Alert List: Addition of *Meloidogyne luci* together with *M. ethiopica*; EPPO Reporting Service No. 11-2017, Num. Article: 2017/218. Available online: <https://gd.eppo.int/reporting/article-6186> (accessed on 10 September 2021).
- Cunha, G.T.; Visôlto, L.I.; Lopes, E.A.; Oliveira, C.M.G.; God, P.I.V.G. Diagnostic methods for identification of root-knot nematodes species from Brazil. *Crop. Prot.* **2018**, *48*, 2. [CrossRef]
- Žibrat, U.; Širca, S.; Susič, N.; Knapič, M.; Gerič Stare, B.; Urek, G. Noninvasive Detection of Plant Parasitic Nematodes Using Hyperspectral and Other Remote Sensing Systems. In *Hyperspectral Remote Sensing*; Pandey, P.C., Srivastava, P.K., Balzter, H., Bhattacharya, B., Petropoulos, G.P., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 357–375.
- Zhang, W.; Qibing, Z.; Min, H.; Ya, G.; Jianwei, Q. Detection and Classification of Potato Defects Using Multispectral Imaging System Based on Single Shot Method. *Food Anal. Methods* **2019**, *12*, 2920–2929. [CrossRef]
- Thomas, S.; Kuska, M.T.; Bohnenkamp, D.; Brugger, A.; Alisaac, E.; Wahabzada, M.; Behmann, J.; Mahlein, A.-K. Benefits of Hyperspectral Imaging for Plant Disease Detection and Plant Protection: A Technical Perspective. *J. Plant Dis Prot.* **2018**, *125*, 5–20. [CrossRef]
- Moghimi, A.; Yang, C.; Marchetto, P.M. Ensemble Feature Selection for Plant Phenotyping: A Journey from Hyperspectral to Multispectral Imaging. *IEEE Access.* **2018**, *6*, 56870–56884. [CrossRef]
- Federico, M.; Riccardo, S.; Salvatore, D.; Stefano, P.; Giuseppe, S. Advanced methods of plant disease detection: A review. *Agron. Sustain. Dev.* **2015**, *35*, 1–25. [CrossRef]
- Ji, Y.; Sun, L.; Li, Y.; Ye, D. Detection of bruised potatoes using hyperspectral imaging technique based on discrete wavelet transform. *Infrared Phys. Technol.* **2019**, *103*, 103054. [CrossRef]
- Ji, Y.; Sun, L.; Li, Y.; Li, J.; Liu, S.; Xie, X.X.Y. Non-destructive classification of defective potatoes based on hyperspectral imaging and support vector machine. *Infrared Phys. Technol.* **2019**, *99*, 71–79. [CrossRef]
- Ayvaz, H.; Santos, A.M.; Moyseenko, J.; Kleinhenz, M.; Rodriguez-Saona, L.E. Application of a portable infrared instrument for simultaneous analysis of sugars, asparagine and glutamine levels in raw potato tubers. *Plant Foods Hum. Nutr.* **2015**, *70*, 215–220. [CrossRef]
- Su, W.H.; Sun, D.W. Chemical imaging for measuring the time series variations of tuber dry matter and starch concentration. *Comput. Electron. Agric.* **2017**, *140*, 361–373. [CrossRef]
- Dacal-Nieto, A.; Formella, A.; Carrión, P.; Vazquez-Fernandez, E.; Fernández-Delgado, M. Non-Destructive Detection of Hollow Heart in Potatoes Using Hyperspectral Imaging. In *Computer Analysis of Images and Patterns, Proceedings of the 14th International Conference, CCAIP 2011, Seville, Spain, 29–31 August 2011*; Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 180–187. [CrossRef]
- Huang, T.; Li, X.Y.; Xu, M.L.; Jin, R.; Ku, J.; Xu, S.M. Non-destructive detection research for hollow heart of potato based on semi-transmission hyperspectral imaging and SVM. *Spectrosc. Spect. Anal.* **2015**, *35*, 198–202. [CrossRef]

19. Zhou, Z.; Zeng, S.; Li, X.; Zheng, J. Nondestructive detection of blackheart in potato by visible/near infrared transmittance spectroscopy. *J. Spectrosc.* **2015**, *2015*, 786709. [[CrossRef](#)]
20. Garhwal, A.S.; Pullanagari, R.R.; Li, M.; Reis, M.M.; Archer, R. Hyperspectral imaging for identification of Zebra Chip disease in potatoes. *Biosyst. Eng.* **2020**, *197*, 306–317. [[CrossRef](#)]
21. Riza, D.F.A.; Suzuki, T.; Ogawa, Y.; Kondo, N. Diffuse reflectance characteristic of potato surface for external defects discrimination. *Postharvest Biol. Technol.* **2017**, *133*, 12–19. [[CrossRef](#)]
22. Zhao, Z.; Prager, S.M.; Cruzado, R.K.; Liang, X.; Cooper, W.R.; Hu, G. Characterizing Zebra Chip symptom severity and identifying spectral signatures associated with ‘Candidatus *Liberibacter solanacearum*’-infected potato tubers. *Am. J. Potato Res.* **2018**, *95*, 584–596. [[CrossRef](#)]
23. Sawant, S.S.; Manoharan, P. A Survey of Band Selection Techniques for Hyperspectral Image Classification. *J. Spectr. Imaging* **2020**, *9*, a5. [[CrossRef](#)]
24. Ballabio, D.; Consonni, V. Classification tools in chemistry. Part 1: Linear models. *Anal. Methods* **2013**, *5*, 3790–3798. [[CrossRef](#)]
25. Chong, I.G.; Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112. [[CrossRef](#)]
26. AlSuwaidi, A.; Grieve, B.; Yin, H. Feature-Ensemble-Based Novelty Detection for Analyzing Plant Hyperspectral Datasets. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1041–1055. [[CrossRef](#)]
27. Jin, X.; Jie, L.; Wang, S.; Qi, H.J.; Li, S.W. Classifying Wheat Hyperspectral Pixels of Healthy Heads and Fusarium Head Blight Disease Using a Deep Neural Network in the Wild Field. *Remote Sens.* **2018**, *10*, 395. [[CrossRef](#)]
28. Liu, B.; Yu, X.; Zhang, P.; Yu, A.; Fu, Q.; Wei, X. Supervised Deep Feature Extraction for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1909–1921. [[CrossRef](#)]
29. Chen, Y.; Zhu, L.; Ghamisi, P.; Jia, X.; Li, G.; Tang, L. Hyperspectral Images Classification with Gabor Filtering and Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2355–2359. [[CrossRef](#)]
30. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A Review of Neural Networks in Plant Disease Detection Using Hyperspectral Data. *Inf. Process. Agric.* **2018**, *5*, 354–371. [[CrossRef](#)]
31. Susič, N.; Žibrat, U.; Sinkovič, L.; Vončina, A.; Razingar, J.; Knapič, M.; Sedlar, A.; Širca, S.; Gerič, S.B. From genome to field—Observation of the multimodal nematocidal and plant growth-promoting effects of *Bacillus firmus* I-1582 on tomatoes using hyperspectral remote sensing. *Plants* **2020**, *9*, 592. [[CrossRef](#)]
32. Hussey, R.S.; Barker, K.R. Comparison of methods of collecting inocula for *Meloidogyne* spp., including a new technique. *Plant Dis. Rep.* **1973**, *57*, 1025–1028.
33. Piqueras, S.; Burger, J.; Tauler, R.; Juan, A. Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution. *Chemometr. Intell. Lab. Syst.* **2012**, *117*, 169–182. [[CrossRef](#)]
34. Chang, C.-I. An Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis. *IEEE Trans. Inf. Theory* **2000**, *46*, 1927–1932. [[CrossRef](#)]
35. Zhang, E.; Zhang, X.; Yang, S.; Wang, S. Improving Hyperspectral Image Classification Using Spectral Information Divergence. *IEEE Geosci. Remote. Sens. Lett.* **2014**, *11*, 249–253. [[CrossRef](#)]
36. Schafer, R.W. What Is a Savitzky-Golay Filter?[lecture Notes]. *IEEE Signal Process. Mag.* **2011**, *28*, 111–117. Available online: <https://inst.eecs.berkeley.edu/~ee123/sp14/docs/SGFilter.pdf> (accessed on 10 September 2021). [[CrossRef](#)]
37. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
38. Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassaniien, A.E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Commun.* **2017**, *30*, 169–190. [[CrossRef](#)]
39. Garthwaite, P.H. An Interpretation of Partial Least Squares. *J. Am. Stat. Assoc.* **1994**, *89*, 122–127. [[CrossRef](#)]
40. Kononenko, I.; Simec, E.; Robnik-Sikonja, M. Overcoming the myopia of inductive learning algorithms with ReliefF. *Appl. Intell.* **1997**, *7*, 39–55. Available online: <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.4740> (accessed on 10 September 2021). [[CrossRef](#)]
41. Breiman, L. Random forests. *J. Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
42. Mason, L.; Baxter, J.; Bartlett, P.L.; Frean, M. Boosting algorithms as gradient descent. In Proceedings of the NIPS’99: 12th International Conference on Neural Information Processing Systems, Cambridge, MA, USA, 29 November–4 December 1999; pp. 512–518. [[CrossRef](#)]
43. Noble, W.S. What Is a Support Vector Machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [[CrossRef](#)]
44. Fawcett, T. An Introduction to Roc Analysis. *Pattern Recognit. Lett.* **2005**, *35*, 299–309. [[CrossRef](#)]
45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
46. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
47. Viaene, N.; Mahieu, T.; Peña, E. Distribution of *Meloidogyne chitwoodi* in potato tubers and comparison of extraction methods. *Nematology* **2007**, *9*, 143–150. [[CrossRef](#)]
48. Bordes, A.; Ertekin, S.; Weston, J.; Bottou, L. Fast kernel classifiers with online and active learning. *J Mach Learn.* **2005**, *6*, 1579–1619. [[CrossRef](#)]

49. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2008; pp. 43–99. ISBN 0-387-95284-5.
50. Menze, B.H.; Kelm, B.M.; Masuch, R.; Himmelreich, U.; Bachert, P.; Petrich, W.; Hamprecht, F.R. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinform.* **2009**, *10*, 213. [[CrossRef](#)]
51. Kononenko, I. Estimating attributes: Analysis and extensions of Relief. In *Machine Learning: ECML-94 1994*; Bergadano, F., De Raedt, L., Eds.; Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence); Springer: Berlin/Heidelberg, Germany, 1994; Volume 784, pp. 171–182. [[CrossRef](#)]
52. Workman, J., Jr.; Weyer, L. *Practical Guide and Spectral Atlas for Interpretative Near-Infrared Spectroscopy*, 2nd ed.; CRC Press: New York, NY, USA, 2012; pp. 231–249, ISBN 9781439875254.